Data Analytics: Adaptive Network-based Fuzzy Inference System for prediction of computer science graduates' employability

تحليل البيانات: التنبؤ بمستقبل، توظيف خريجي علوم الحاسوب باستخدام نظام الاستدلال الضبابي

by

SAADA KHADRAGY

A thesis submitted in fulfilment

of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE

at

The British University in Dubai

April 2020

Data Analytics: Adaptive Network-based Fuzzy Inference System for prediction of computer science graduates' employability

تحليل البيانات، التنبؤ بمستقبل توظيف خريجي علوم الحاسوب باستخدام الاستدلال الضبابي

By Saada Khadragy

A thesis submitted to the Faculty of IT and Engineering in fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY in Computer Science

At the British University in Dubai

April 2020

Thesis Supervisor

Professor Sherief Abdallah

Approved for award:

________________________	________________________
Name	Name
Designation	Designation
________________________	________________________
Name	Name
Designation	Designation

Date: ____________

DECLARATION

I warrant that the content of this research is the direct result of my own work and that any use made in it of published or unpublished copyright material falls within the limits permitted by international copyright conventions.

I understand that a copy of my research will be deposited in the University Library for permanent retention.

I hereby agree that the material mentioned above for which I am author and copyright holder may be copied and distributed by The British University in Dubai for the purposes of research, private study or education and that The British University in Dubai may recover from purchasers the costs incurred in such copying and distribution, where appropriate.

I understand that The British University in Dubai may make a digital copy available in the institutional repository.

I understand that I may apply to the University to retain the right to withhold or to restrict access to my thesis for a period which shall not normally exceed four calendar years from the congregation at which the degree is conferred, the length of the period to be specified in the application, together with the precise reasons for making that application.

Saada Khadragy

Abstract

The increased amount of data generated in the world of today in all fields is considered to be an indicator for future predictions. In recent decades, in any field and as a result of developments in information technology, a huge amount of data has been provided from the educational field, by which students' Employability Prediction has become a main concern for higher education institutions.

The question of employability has become a critical consideration not only for graduates but for the educational institutions themselves. This research study compares a number of classifiers to determine the effective classifier that accurately and efficiently categorizes CS and IT graduates into employed, unemployed, or other, and predict the future employability of CS and IT students in Jordan.

For this purpose, an Adaptive Network Fuzzy Inference System (ANFIS) is applied in this research study. The data of 1095 CS and IT graduates was obtained from three universities in Jordan. This data was collected through a set of tracer studies that were carried out by these universities. ANFIS, Decision Tree, SVM, MLP, and Naïve Bayes classifiers were applied in order to find the classifier with the highest accuracy and efficiency. The final outcomes showed that ANFIS has the highest accuracy, with a percentage of 94% accuracy for its predictions.

A set of recommendations is presented by the researcher according to the most effective factors that influence the CS and IT employment market in the Middle East. The researcher suggests for the ministries of higher education to focus on developing the CS and IT students' programming skills and communication skills, which emerged as essential for increasing CS and IT students' employment prospects. affecting the employment market for CS and IT.

Key words*:* Data Mining, Employability, Neuro Fuzzy Approach, Adaptive Network based Fuzzy Inference System, Classification

ملخص

نظراً للكم الهائل من البيانات الذي ينتج في جميع مجالات الحياة و من بينها مجال التعليم. هدفت الدراسة الحالية إلى تحديد أفضل أساليب تنقيب البيانات لإستخدامها للكشف عن العوامل المؤثرة في توظيف خريجي الجامعات في تخصص علم الحاسوب وتقنية المعلومات، و التنبؤ عن مستقبل توظيف خريجي هذه التخصصات في الشرق الأوسط حيث باتت هذه المشكلة تؤرق مؤسسات التعليم العالي التي تقدم هذه البرامج، والخريجين أنفسهم على حد سواء ؛ وذلك بسبب ارتفاع نسب البطالة بين هؤلاء الخريجين .

ولتحقيق هدف الدراسة استخدمت الباحثة نظام الاستدلال الضبابي (ANFIS) ، الذي طبقته على عينة من خريجي علم الحاسوب وتقنية المعلومات (??M?) من ثلاث جامعات أردنية، حيث تم جمع البيانالت من خلال الدراسات التتبعية التي قامت بها تلك الجامعات.

وبمقارنة الطريقة التي استخدمتها الباحثة بأربع طرق أخرى تستخدم في التنقيب في البيانات بغرض التنبؤ مثل: تقنية شجرة القرار، تعدد الطبقات ، أظهرت الدراسة ازدياد مستوى الدقة بالتنبؤ بازدياد عدد المتغيرات المطبقة في هذه الطريقة ، حيث وصلت نسبة الدقة الى ??% .

وبناء على نتائج هذه الدراسة اقترحت الباحثة مجموعة من التوصيات إلى القائمين على إعداد ومطوري البرامج الأكاديمية في وزارة التربية والتعليم ، والتعليم العالي منها : رفد برامج علم الحاسوب وتقنية المعلومات بمساقات من شأنها تطوير المهارات الدقيقة في البرمجيات .

الكلمات و العبارات الأساسية: تنقيب البيانات، القدرة على التوظيف، نظام الاستدلال الضبابي، التصنيف

DEDICATION

I dedicate this thesis to my whole family and friends, as an appreciation of your constant love and support; your inspiration and motivational words have carried me throughout this process.

Mother, Father, my sister Safaa and My brother Ahmed your encouragement and belief in my competences made the dream a reality…Thank you.

My husband, thank you for your patience and understanding on all my frustrations and stressful deadlines…

To my dearest children Mohamed, Mahmoud and Haneen; Thank you for every time you understood why I was not there. This work is our family achievement. During this journey, Mohamed has been my personal proof-reader throughout the writing journey. Mahmoud has been my positive energy with his beautiful smile and Haneen has been my very best friend and never stopped motivating me to keep going. Your strength is what made me be where I am today. I owe you all the greatest degree of gratitude.

I am blessed to have two wonderful friends who were my source of power in many of my weak days. Life has gifted me two sisters from different mothers and fathers who stayed next to me and be part of my happiness. Thank you Sadia Shahen and Amal Al Jiboury for being in my life.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my Director of Studies, Prof. Sherief Abdallah for his marvellous support, patience, and bright motivation. It was a great honour to work on this thesis under his supervision.

I also want to thank my second supervisor Prof. Khaled Shaalan for his encouragement and constructive advice which have been incredibly valuable and appreciated.

I would also like to extend my warmest thank you to Dr. Khaled Derbashi for his support and motivation during my journey of study.

I am also very grateful to Prof. Sufian Forawi for the inspiration he has been to me through his writings during my journey of study.

Lastly, I hope that the insights presented in this thesis might drive policies which will develop the good work being accomplished by this research study.

Table of contents

List of Tables

List of figures:

List of Abbreviations:

Abbreviation	Definition
AC	Associative Classification
AEN	Action state Evaluation Network
AI	Artificial Intelligence
ANFIS	Adaptive Network-Based Fuzzy Inference System
ANN	Artificial Neural Networks
ARM	Association Rule Mining
ASN	Action Selection Network
CBA	Classification Based on Association
CBI	Confederation of British Industry
CS	Computer Science
DM	Data Mining
EDM	Educational Data Mining
FALCON	Fuzzy Adaptive Learning Control Network
GA	Genetic Algorithm
GARIC	Generalized Approximate Reasoning based Intelligence
GEI	Graduate Employability Indicators
GISMO	Graphical Interactive Student Monitoring System
GPA	Grade Point Average
IG	Information Gain
IT	Information Technology

KDD	Knowledge Discovery in Databases
KNN	K-Nearest Neighbor Algorithms
LA	Learning Analytics
LMS	Learning Management System
LSE	Least Square Error
MCA	Master of Computer Applications
MDL	Minimum Description Length
MOHE	Ministry Of Higher Education
NFC	Neuro Fuzzy Classifier
PPI	Pre-Professional Identity
PSO	Particle Swarm Optimization
ROC	Receiver Operating Characteristic
SMO	Sequential Minimal Optimization
SVM	Support Vector Machine
SVM	Support Vector Machine
WEKA	Waikato Environment for Knowledge Analysis
WUM	Web Usage Mining

1 Chapter 1. INTRODUCTION

The current research study provides a multi-faceted domain for statistical analysis and data analytics in order to define different attributes influencing the future of Computer Science, Information Technology graduates' employability in the Middle East. The concept of Employment is known as a paid job contracted between an employer and employee and relates to an individual who is hired for a salary to achieve certain work for an employer (Curtis, 2015). Simply, the concept of employability is about being able to get and keep achieving particular work. In a more comprehensive way employability is the capacity to shift self-sufficiently into the work force market to recognize the ability through maintainable employment. In a personal view, employability is based on the knowledge, skills and viewpoints they have, the different way they apply those possessions and introduce them to employers, and the contexts within which they look for work, which is known as the personal status and the work force settings (Nauta et al., 2009). On the other hand, Employability is defined as having the ability to get a certain work as a type of employment, keep the same type of employment and attain new role if necessitated. Adding to the same point of view, employability for an individual focuses on how individuals can deploy their knowledge, skills, and attitudes in order to convince their employers (Vanhercke et al., 2014). Hence, the definition of employability is varied between individuals and institutional relationships with the labour market itself. As in a number of countries all over the world, the governmental sector concentrates more on the vocational skills than the soft skills (de Guzman and Choi, 2013). For all of the above, in this thesis the concept of employability is highlighted for Computer Science, Computer Information System, and Computer Engineering graduates and the most effective factors affecting their abilities and skills to get a job. On the other hand, many tools and techniques are used to analyse the educational data in order to reach profitable decisions in the educational sector (Osofisan, Adeyemo and Oluwasusi, 2014). One of the major tasks of data mining in education is to analyse students' behaviours (Dawson and Dawson, 2019). Furthermore, recent years have witnessed a great rise in the application of electronic tools in the field of education. From nursery classes at the preschool stage to the postgraduate programs at the universities, electronic tools are being used extensively to support and enrich the quality of education. Although the implementation of computer networks is an integral characteristic of online learning, the face to face schools and universities are also using extensive network-connected electronic tools such as mobile phones, tablets, and computers, which directly and indirectly affect graduates' employability in all fields and majors. However, practitioners and academic administrators can gain from their colleagues in business and service industries a multifaceted system of methods and tools, usually indicating data mining which is being applied to analyse a huge dataset in decision-making. Scientists and researchers have begun paying attention to the use of data mining and data analytics to manage big data created by the educational sector. From the educational perspective, these tools are specifically defined as educational data mining (EDM) and learning analytics (LA). Generally, EDM seeks new patterns in big data and provides new algorithms and/or new models, while LA uses known predictive models in certain systems.

From all of the above, the term of Employability is considered as a two-sided equation and many people should have several kinds of assistance to overcome the physical and mental obstacles for both the learning and development scales. Moreover, Employability is not only about vocational and academic skills, but individuals should also have relevant and usable workforce information to assist them to take right decisions about the workforce alternatives available to them. They also need help to understand when such information may be important and to read that information and transfer it into intelligence (Schnell and Rodríguez, 2017).

1.1 Problem statement:

Data analytics has been used in several areas because of its ability to rapidly analyse different amounts of data. More recently, researchers and higher educational institutions have also started to discover the potential of data analytics in analysing academic data. The main goal of such endeavors is to discover the means to develop the services offered by these institutions and to improve their organization.

At present most research studies focus on retention issues (Romero and Ventura, 2013), and studies that may improve student academic performance (Desmarais and Pelczer, 2010), from student academic performance, retention policies to develop graduation level and programs to improve graduates employability. The application of data analytics in graduate employability is to search for significant relationships such as patterns, association and changes among variables in databases. One of the main goals of this research study is to predict graduates employment after graduation from university and to compare a number of classifiers in order to select the ones with high accuracy.

The research is intended to assist the higher educational organizations in preparing their graduates with sufficient skills to enter the job market. This study provides a graduate employability model that applies a classification task to consider the most important reasons for graduate employability specifically for CS graduates using the educational domain in Jordan.

The educational field generates huge amounts of data annually; this data is considered by researchers and data analytics as an indicator for many things in the learning outputs, such as students' failure or distinction in some courses (Mishra, Bansal and Singh, 2017). While it is a great challenge to use this data to improve both the educational process and students future, the current research study would compare a group of classification data mining techniques; these techniques include a combination of neural network approach and fuzzy set approach, in order to discover the hidden information from the generated data. The discovered information may support decision-makers in the educational field and students' parents alike.

Historically, taking the right decision by educational stakeholders is a great risk in the educational process. From this idea, developers would use the generated data as an indicator for future decisions. While these decisions are directly and indirectly affecting students, their current academic performance would be the guidance for their future. The main objective of the current research study is to build a classification model by using neuro-fuzzy techniques on the graduates of CS and IT majors by labelling them into employed and unemployed categories based on the most effective factors and skills of the graduates. This model would be applied for newly registered CS and IT students and predict their employability future in the Jordan. A set of recommendations will be provided to high educational institutions and universities in order to develop educational plans to enhance students' skills required by CS, IT labour market.

Moreover, there are many research studies in the educational field that explored the ways of implementing data mining techniques for different educational aims. A classification considered as a very familiar data mining task that has been used widely in educational field. There are many approaches that have been implemented to perform the job of classification, such as decision trees, neural networks, naïve Bayesian, support vector machines, and fuzzy set. This study combined the neural network and fuzzy approach, a combination of both approaches called Neuro-fuzzy method. Artificial Neural networks are classified as one of the artificial intelligence techniques that imitates the human brain, in which different nodes connect together to create new data layers; the hidden layers produce the outputs known as node value. Fuzzy set approach can handle the problems relating to ambiguous, subjective and values' estimation.

Additionally, most of the research studies that study the employability of IT graduates came from Southeast Asia and Middle Eastern countries. One study investigates the importance of employability skills as perceived by employers from manufacturing industries. The results of the study indicated that employers give great importance to communication skills, problem solving skills, team work skills and personal qualities (Rasul *et al.*, 2013).This means that there is a problem with employability in these countries as it will be highlighted in the literature review part. Therefore, we decided to choose a country from the Middle East and study the CS and IT market as well as select a data sample of graduates in IT specialization from this country. Unfortunately, the numbers of IT graduates in the UAE are relatively small. Due to that fact, we have chosen Jordan, which gave us the ability to get good data samples from several graduates' tracer offices of different universities. Moreover, we managed to get statistical data from official sources, such as the Ministry of Digital Economy and Entrepreneurship and the Ministry of Labour in Jordan. The current study would compare a number of classifiers to predict graduates' future employability in Jordan. Figure 1.1 shows the study problem and shapes the study problem framework. Creasey (2013) assured the existence of recent debate about students' employability. The current research study focuses on students of CS as educational input, who after acquiring a set of skills and knowledge will be educational output (fresh graduates). Then they will face different challenges to join the job market and seek their employers' and institutions' satisfaction. This satisfaction needs to be fulfilled through a certain number of personal, academic, generic, and communication skills. In the current study and as shown in figure 1.1, the researcher divided the employment into soft skills and demographics. Hence, a number of classifiers are compared to determine the effective factors that affect the future of CS graduates' employability.

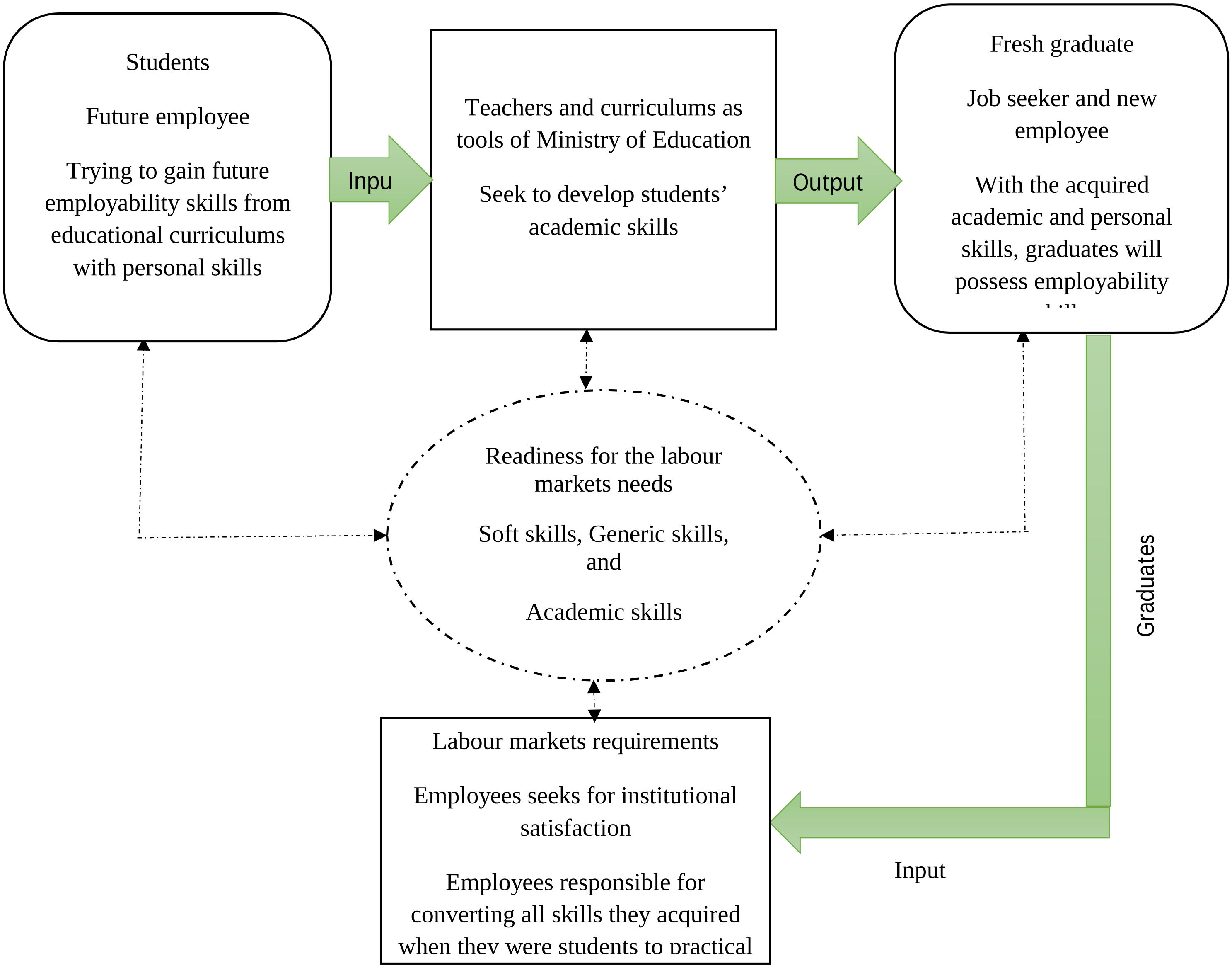

Figure 1. 1 Problem statement framework

1.2 Significance of the study:

The employability of graduates has emerged as a major concern in recent years. In Jordan the demand for labor, especially for skilled and qualified employees, is expected to stay high and robust in the coming years. Internationally, Tuma and Pratt (1982) stated that employability markets' requirements has become one of the most critical issues in our societies. Regionally, the Ministry of Digital Economy & Entrepreneurship of Jordan (2019), mentioned employability as one of the top concerns of developing both the educational and economical systems. Jordan believes that successful reform of initiatives, which give skilful graduates to the labour market can highly reinforce the goal of the Fourth Industrial Revolution (Bagdasarian et al., 2019)

In this essence, a number of official research studies found an employment rate in Jordan of only 40% of total graduates in the IT field for the year of 2018. This is an indication of a significant gap between graduates' numbers (supply) and employability (demand). To the best of our knowledge, the number of research studies concerned with employability of CS and IT graduates is growing steadily in international literature. While a set of research studies were conducted internationally in order to minimize the gap between Computer Science (CS) and Information Technology (IT) graduating students' numbers and the ability of getting employed (Marchante et al., 2003), very few such studies are carried out in the Middle East region, especially in Jordan, which has encouraged recent research study to be conducted in the region. Most of the international studies in the field of employability focused on demographics and how gender affects the ability of getting hired (Kasler, Zysberg and Harel, 2017); few of them focused on the soft skills and how they affect employability of the graduates (Andrews and Higson, 2008), and very few emphasized the importance of written communication skills (Moore and Morton, 2017). Yet the number of such research studies is rare in the Middle East; examining the combination of all factors with a different cultural framework will make this research significant.

The current research study serves in providing the national literature with a clear perspective about the needed skills for the labour market in CS and IT fields. Also, it reveals the existing number of factors affecting self-employment as an essential domain to provide a reasonable argument on why there is a gap between the demand and the supply in the field of employment (Branine and Avramenko, 2015).

Additionally, the labour market is extremely competitive, with high demands for high quality and qualifications from new employees. Despite the theoretical emphasis that supports employability, none of the research studies mentioned that this theoretical base was able to articulate the relationship between educational curriculums, graduates' skills, and labour market demand in Jordan. This research study serves to recommend a list of the needed skills for computer science graduates to match the market's requirements in Jordan. Adding to that, Brown, Hesketh and Williams (2013) advocated that graduates' employment remains a national and international issue because of the vast numbers of graduating students in all fields that are produced every year from higher education institutions.

As a response to all of the previous calls, the current study attempts to focus on the required skills for CS graduates to get hired in Jordan and it may be used as a source or a reference in the development of the educational curriculums. Additionally, this study may participate in providing a list of new fields of research such as the effect of cultures on such topics which will profoundly affect the future of employability through an analysis of the currently required skills. The findings will add to the body of knowledge, of how to further integrate the economic and educational fields by providing a classification model for CS graduates. All countries seek to the highest employability rates in all fields; which came as a serious demand to the researchers conducting studies that assess the educational stakeholders taking their decision towards the future of their students.

According to Chen (2017) and self-determination theory, the efforts made to CS students by their families and the ministries of education are in a deep need to be developed towards the students' future careers. When any person is unemployed, his/her family also will be affected, while the society as a whole misplaces its contribution to the economy with regard to the goods and services that are produced. Unemployed persons also lose their purchasing energy, which can lead to unemployment for other workers, producing a cascading effect through the economy.

On the other hand, in many places the economic fields are not generating enough vacancies for the growing number of graduates. As a result, there will be an uncertain future for the new generations of graduating students unless a plan is constructed to confront this issue (Gibson and Ifenthaler, 2016). Therefore, higher educational institutions need to provide their students with the required knowledge and soft skills to tackle the challenges of the labour market. All of this preparation would increase the graduates' opportunities to obtain decent jobs which match their qualifications, education, and soft skills.

Hence, the current research study uses several techniques with scientific approaches in order to reach a suitable predictive model, such as Data Mining (DM), Educational Data Mining (EDM), and a combination between Artificial Neural Network (ANN) and fuzzy logic, namely neuro-fuzzy. While extracting information from amounts of data is a simple definition of Data Mining (Sumathi, 2006), extracting information from the educational data is called Educational Data Mining (Asif et al., 2017). However, Artificial Intelligence has different types; these would be implemented to extend human thinking and the intelligence of machines (Negnevitsky, 2011). Deep neural network is one of the types that mimics the real human brain neurons (Neapolitan and Neapolitan, 2018).

In a nutshell, the study findings are likely to provide new insights that should develop and impact the CS graduates' employability future. Moreover, the study results will provide a set of recommendations to higher educational institutions of the required skills for CS graduates which would develop curriculums.

1.3 Conceptual framework:

The current study provides a conceptual framework for identifying what employers assume about the value of CS graduates with similar educational credentials in the workplace (their employability). As shown in figure 1.2, the conceptual framework of this research is designed in a graphical structure to best explain the natural progression of the study problem (employability) under discussion. It also illustrates the connection between the main concepts that are highlighted in this study and indicates how the research claim would be explained and discussed. Figure 1.2 demonstrates the conceptual framework of the study. Three main areas are addressed; the first area is the importance of employability. Past research studies on employability skills conducted nationally and internationally, showed that a lot of technical graduates lacked employability skills (Rasul et al., 2013), which made it important to conduct such a study specifically in the Arab region where it is rare to find a discussion about employability. Also, examining the required skills affecting the CS graduates' employability is a critical topic for improving the economic side, where all countries and governments seek for the highest percentage of their graduates' employment. The second area of the conceptual framework is the practice, which indicates the importance of data analytics in employability.

Graduate employability refers to university alumni who have the ability to get a job. Furthermore, employability shows that institutions and employers have developed the knowledge, skills, attributes, reflective competences and identity that graduates need in order to succeed in the workforce.

Recently in universities, there have been two parallel roads guiding us to move from Point A to Point B, only there is a path between the two points. They would be more powerful and strategic if they were joined as a single road with two lanes. One of the two roads is that of data analytics and the other is that of student employability (Tomy and Pardede, 2019). Hence, data mining different steps are applied in this study in order to compare different classifiers and reach the set of needed skills for the CS graduates' employability. However, the third area is the outcome of comparing the classifiers. The "No Free Lunch" theorem indicates that there is no one model that works best for every problem (Wolpert and Macready, 1997). The expectations of a suitable model for one topic may not work for another topic, so it is normal in machine learning to examine multiple classifiers and find one that works best for a particular problem.

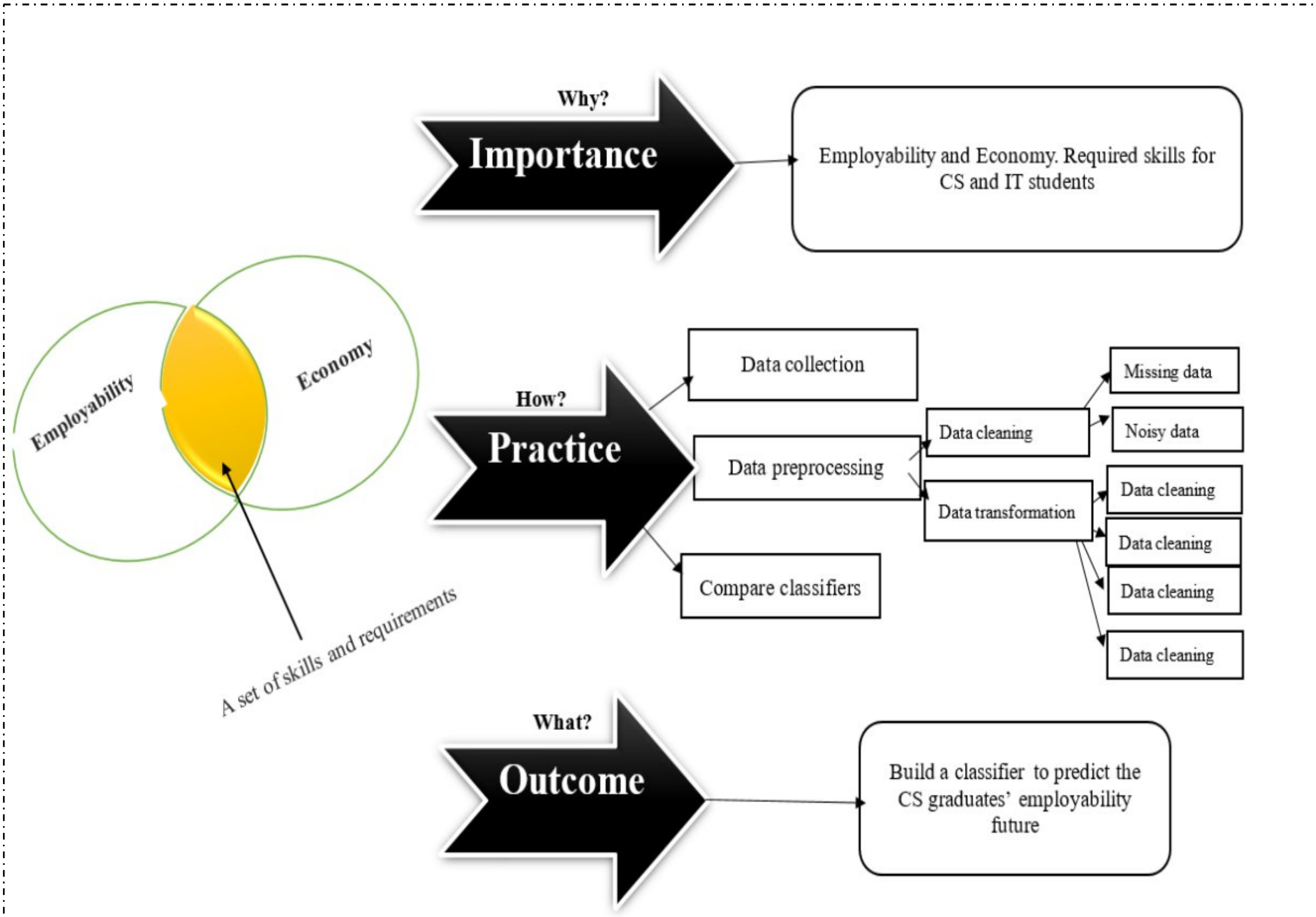

Figure 1. 2 Conceptual framework of the study

1.4 Research aims and objectives:

The current research study provides a multi-faceted statistical analysis and data mining in order to consider different attributes affecting future employability in Jordan. Since data mining techniques have been used in different fields, they can rapidly analyse the massive datasets. The current research study aims to build a classification model for CS and IT graduates' future employment by using data mining techniques. In this regard and based on the following statistical facts, this study aims for comparing a number of algorithms to reach a suitable range of accuracy and efficiency.

1. A model is an efficient illustration of reality (Choi, Sun and Heng, 2004), and representations are made to discard redundant details and allow us to focus on the feature of reality that we want to understand. These simplifications are reformed on hypotheses; these hypotheses may work in some situations, but may not work in others. This indicates that a model that illuminates a certain situation well may fail in another situation. In both statistics and machine learning, researchers need to check their assumptions before depending on a model.
2. The "No Free Lunch" theorem (NFL) states that there is no one model that works best for every problem. The hypothesis of a great model for one problem may not work for another problem, so it is normal in machine learning to try different models and discover what works best for a particular problem. This is especially true in supervised learning; validation or cross-validation is commonly used to help the predictive accuracies of multiple models of varying complexity to find the best model (Köppen, Wolpert and Macready, 2001). Depending on the problem, it is important to assess the trade-offs between speed, accuracy, and complexity of different models and algorithms to find a model that works best for that particular problem. Similarly, the current research study compares a number of classifiers to decide on the suitable model for predicting the CS and IT graduates' employability.
3. Furthermore, the NFL theorem states that with all possible performance measures, no search algorithm is the best or better than another when its performance is averaged over all possible discrete functions. In 2005, Wolpert and Macready stated that any two optimization algorithms are equivalent when their performance is averaged across all possible problems. Also the 1997 theoremms of Wolpert and Macready are mathematically technical. Hence, the folkloric NFL theorem is an easily

illustrated and easily comprehended consequence of theorems that Wolpert and Macready actually prove. It is weaker than the proven theorems, and thus does not encapsulate them. Accordingly, in the current study, a number of data mining algorithms will be compared to classify CS and IT alumni students in Jordan according to their employability status. More specifically, the major aims and objectives of the present research study are as follows:

- The primary objective of this research study is to build a classification model by comparing neuro-fuzzy techniques with a number of classifiers on the alumni graduates in both of CS and IT majors by labelling them into employed and unemployed categories. This model would be used for newly registered CS and IT students and predict their employability future in the Jordan.
- Moreover, it is challenging and important at the same time to come up with a new paradigm in research as the issue of employability rates and the graduates' numbers has a limited set of research papers in the Middle East in general. From this point of view, the current research study would be considered as a reference for this type of research studies in the Middle East.
- Building on the ideas stated above, collecting the required data from Jordanian universities about graduates' career journeys would be authenticated and trusted because it is collected from individuals not institutions (Witten, Frank and Hall, 2011).
- It can be noticed in the majority of developing countries that the ministry of higher education in these countries struggles to achieve 100% employment rates from the graduating students (Md Razak et al., 2014). Also, graduating students struggle to join the most appropriate field which has vacancies for them according to their academic qualifications and soft skills (Stott, 2007). These two reasons came obvious to apply classification techniques to determine the relation between the outcomes of higher education institutions as a supply with the requirements needed for employment as a demand and to reach a suitable predictive model for the newly joining students in both CS and IT majors for their employment future. CS and IT are the two majors that will be emphasized in this research study. Also, a type of classification will be applied on the recent employees from alumni students of different Jordanian universities in the two fields, and then a neuro-fuzzy approach will be implemented to find out a suitable predictive model for students who are still on the seats of studying CS and IT.
- Moreover the importance to understand the human needs to achieve the highest rate of self-employment (Cairns et al., 2015) which is cited in many research studies as a basic factor of self-

determination theory (Deci and Ryan, 2015). This came as an issue with an urgent need to be discussed in the research field because both self-employment and self-determination affect the internal social movement (Heller et al., 2011).

- This research study presents a type of artificial intelligence, machine learning, and educational data mining studies which have a multi-faceted approach in order to predict university students' future employability in Jordan.
- From the definition of data mining by many researchers as a type of technology, its main function would serve the current research study to illustrate knowledge acquirement and to reach meaningful relationships to form patterns, associations and differences among variables in databases. There are many data mining approaches that could be applied in order to mine applicable and attractive knowledge from huge datasets. Also, data mining has numerous tasks, for example classification and prediction, and association rule mining with clustering. Additionally, classification is considered as one of the most essential approaches in data mining, to construct classification models using an input data set.
- The applied classification approaches usually construct models that are familiar with predicting future data trends. Moreover, there are different algorithms for data classification such as decision tree, Naïve Bayes classifiers, SVM, MLP, and ANFIS. In this research we will compare these classifiers in order to reach the high accuracy and efficiency.
- Based on the current research findings, a set of recommendations will be provided to the ministry of higher education and higher education institutions in Jordan with the most effective factors affecting CS and IT graduates employability.

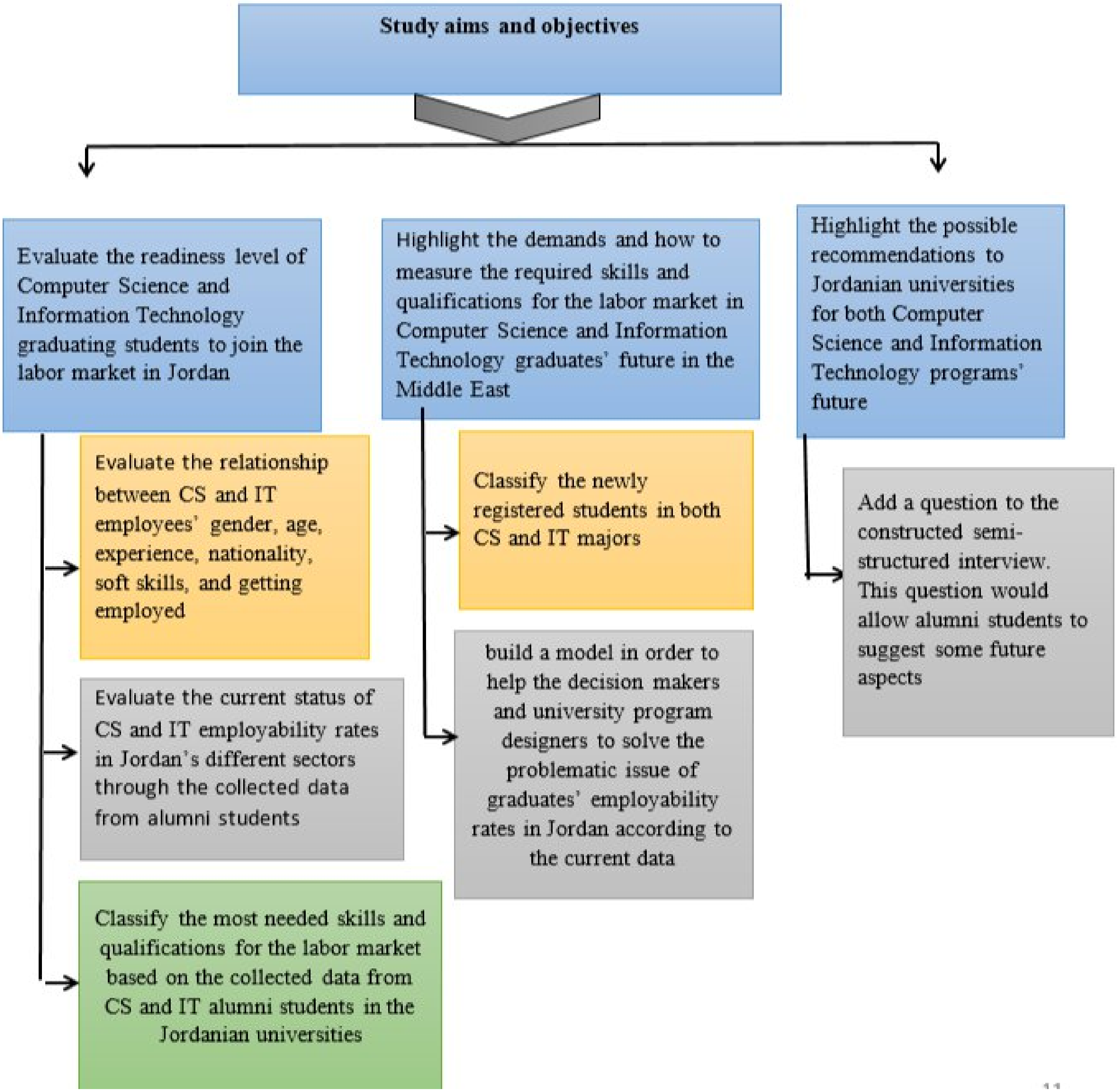

figure 1. 3 Study aims and objectives.

To sum up, the major aim of this research study is predicting the registered CS and IT major students' future employability according to the classification of graduating their' employment rates into three different categories: employed, unemployed or in an undetermined situation.

1.5 Research questions:

The current research study will address the following questions:

1. How can we apply a neuro-fuzzy approach in predicting CS, IT graduates' future employability?
2. Which predictive technique would be most suitable for predicting CS, IT graduates employability?
3. What are the IT recruitment experts opinion of the proposed system?
4. Which factor most affect the CS, IT students' employability future?
5. Which factor most affect the CS and IT graduates' future employment rates?
6. How can the prediction model be used to evaluate the graduates' future employment?

The required data for this research study will be collected from different sources quantitatively and qualitatively. The sample population for the current study will be the CS and IT alumni graduating students from different universities in Jordan. The data of the graduate students was collected from the graduates data; demographics, soft skills, technical skills, and academic skills; which was built from a tracer study carried out by the Jordanian universities.

1.6 Organization of Dissertation:

The reminder of the thesis is organized as follows:

Chapter 2 presents an organized review of the literature that includes computational terms in computer science, previous research studies of data mining tasks and major applications, and data mining classification techniques. Also, this chapter goes through the fundamental theories related to the prediction models, previous research studies in the field of educational data mining and processing techniques for data analytics with the determination of the successful prediction models and classification tools. The most important techniques and applications of educational data mining in employability will be discussed in this chapter. Literature of neuro-fuzzy approach in classification and its applications in EDM will be discussed.

Chapter 3 introduces and illustrates classification, including decision tree classifiers, attribute selection measures and tree pruning. Chapter 3 also discusses Bayesian classification methods and several rule-based techniques, as well the common techniques for assessing accuracy. The final part of chapter 3 demonstrates the advanced classification techniques, such as Bayesian belief networks, classification by backpropagation, support vector machines and neuro-fuzzy classifications methods called ANFIS which is the main algorithm in this thesis through which the prediction model will be built.

Chapter 4 presents the process of applying our methodology and data analysis for alumni students, their labels (employed, unemployed, and undetermined), and the factor which most affects their employment rates, with associations generated from the prediction model.

Chapter 5 presents the findings of the data analysis with the discussion of the results. Different obstacles to data driven research of the employment market and the most effective factors will be highlighted.

Chapter 6 presents the discussion of the final results that were produced from the experimental study that was performed in Chapter five. The more important attributes that affect the prediction of the employability of the graduates will be shown. The required skills for IT graduates that affect their employability will be discussed, which may prompt the educational institutes to modify their curriculums and teaching strategies.

Chapter 7 presents the final conclusions produced from the study outcomes. The limitations of implementing ANFIS algorithm on an educational dataset which contains more than 20 attributes will be demonstrated, as well as the researches that can be carried out according to the problems that appeared during experimental study will be demonstrated as proposed future works after this thesis. One more thing to be presented in this chapter is a set of recommendations for Jordan's Ministry of Education to discern the direction of future research.

2 Chapter 2. LITERATURE REVIEW

Currently, employability has gained a lot of attention from higher education institutions. Employability data enables these institutions to better plan their educational strategies, enhance the curriculum, as well as improve students' performance. Studying employability can be achieved by analysing the data extracted from educational resources. Nowadays, there are a lot of resources of educational data such as learning management systems (LMS), online learning system, admission systems, and social media, which provide a vast amount of data. Handling and analysing this huge amount of data cannot be carried out without sophisticated tools, and data mining is the most important technology for this purpose; the task of data mining predicts interesting and useful information that helps in making future strategy plans. Data mining has been used in many fields such as banking, stock exchange, marketing management, retail sales, health, and recently in education. Additionally, and as noticed from the literature, several research studies in different countries have investigated issues related to employability. Most of these studies have been applied to countries with high unemployment rates, such as Vietnam (Tran, 2015). The Arab world suffers from the some of the highest unemployment rates in the world, especially non-oil countries (Henderson, 2013). Therefore, I decided to choose a country from the Middle East and study the Computer Science (CS) market as well as selecting a data sample of graduates in CS specializations from this country.

On the other hand, employers are often in search of skills that go beyond credentials, knowledge, and experience. An individual's education and experience may make him/her suitable to be hired and join a job, and to be effective in the majority of tasks, he/she will need certain skills which are likely to progress over time. Some skills will be precise to the job, but the greatest amount will be so-called 'soft skills' that can be needed in any job or employment fields. These soft skills are 'employability skills': they are what makes anyone employable (Lin, Korsakul and Korsakul, 2012). Furthermore, higher education has a basic role in the reinforcement of any people's economy as it is an industry branch in itself and it keeps the rest of the industry going by offering trained labourers (Laine, Leino and Pulkkinen, 2015). In the past, the basic issues for the educational institutions were the reduction in the students' success rates (Araque, Roldán and Salguero, 2009), reduction in students' retention (Pegrum, Bartle and Longnecker, 2015), growth in students shifting to other likely institutions and shortage of guiding students in subject choice. Since education has become more employment focused, graduating from any university has become an essential aspect in developing its reputation (Pool, Qualter and Sewell, 2014a). Educational institutions produce and gather vast amount of data (Sedkaoui and Khelfaoui, 2019). This may be about students' academic progression, students' personal profile, their communication abilities, their web log interests and also graduates' profiles (Bharambe et al., 2017). The task of predicting students' employability may support identifying the students who are at a higher risk of unemployment and consequently management may make timely interference by introducing vital steps to train their students in order to develop their performance skills (Mishra, Kumar and Gupta, 2017). Computer scientists have tried to identify relations between academic achievement, socioeconomic surroundings, job skills of the graduates and employability, however, mostly by implementing statistical methods (Abas and Imam, 2016). Those statistical methods have a direct relation with the field of computer science and data analytics (Cao, 2017).

Educational Data Mining (EDM) is the process of applying Data Mining (DM) techniques to educational data (Wu et al., 2014). EDM has emerged as an important research area in recent years for researchers all over the world (García et al., 2011) . Due to the importance of data mining in the field of education, the International Educational Data Mining Society[1] was founded in July 2011. This society provides much useful research related to EDM, as well as helpful datasets.

This chapter surveys the essential background to understand the DM and EDM. Firstly, here is discussed the data mining definitions as part of Knowledge Discovery in Databases (KDD), and also present data mining tasks and major applications. Furthermore, educational data mining (EDM) is illustrated as a main part of this thesis. A lot of literature shown in this section demonstrates the most important research studies that have included the concept of EDM. Then, the most important techniques and applications of applying educational data mining in employability are investigated. Finally, a very important part of literature, which is neuro-fuzzy approach backgrounds and related works is illustrated.

2.1 Data mining

This section deliberates the data mining definition as part of KDD, and also presents data mining tasks and its main applications.

[1] http://educationaldatamining.org/

2.1.1 Definition

Data mining is the process of automatically discovering useful information in a huge dataset. Other definitions for DM are, "the non-trivial extraction of implicit, previously unknown and potentially useful information (such as rule, constraints and regularities) from data in a database" (Kaur and Madan, 2015), and "efficient discovery of previously unknown patterns in large datasets" (Anand, Bell and Hughes, 1996), and "process of employing one or more computer learning techniques to automatically analyse and extract knowledge from data contained within databases" (Pechenizkiy et al., 2011).

Not all tasks of information elicitation from databases are considered as data mining tasks. For instance, the retrieval of information about web pages using a search engine is considered an information retrieval task, not a data mining task.

Many people use data mining terms as a substitute for the term "knowledge discovery from data" (KDD), but actually DM is a step of KDD. Data mining is an integral step of the KDD process which is the major process of KDD: converting raw data into useful information. This process involves a series of transformation steps from data cleansing to data mining to knowledge representation, as in Figure 2.1

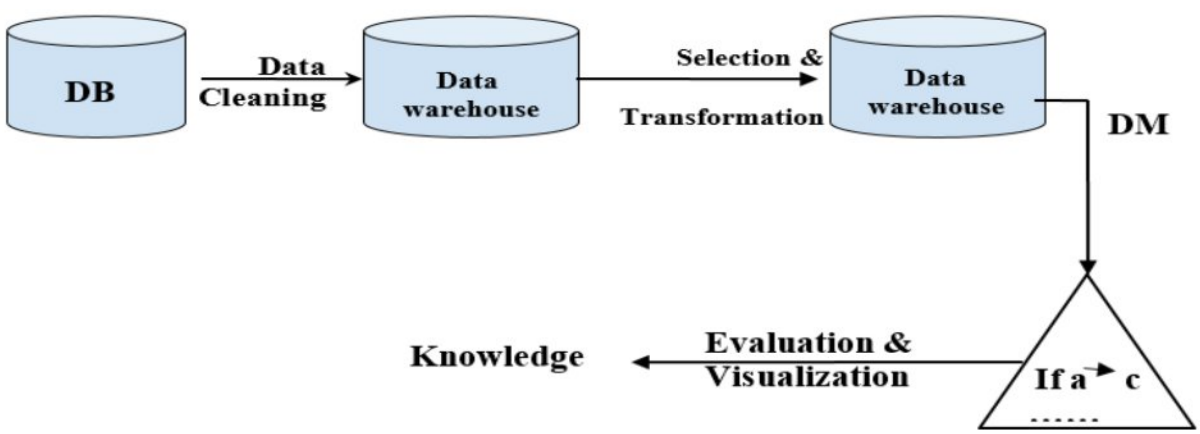

Figure 2. 1 Knowledge Discovery from database (KDD)

The KDD steps are:

1. Data cleaning: removes noise and duplication of data.
2. Data selection and transformation: remove irrelevant data for the given task and transform the remaining data into a form suitable for mining.
3. Data mining: extracts patterns from the data.
4. Evaluation: ensures that only valid and useful patterns are retained
5. Visualization: presents visual representations of mined patterns to users.

DM use sampling, estimation and evaluation from statistics and machine learning, it uses search methods from artificial intelligence (AI) and it uses pattern extraction from pattern recognition. Other research areas adopted by DM: see Figure 2.2.

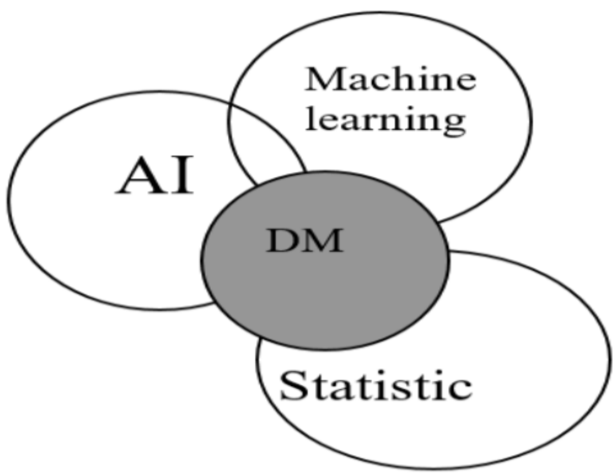

Figure 2. 2 DM discipline

2.1.2 Tasks

Some research studies classify them into two essential tasks (Kesavaraj and Sukumaran, 2013):

- **Predictive** tasks: where the major goal is to predict the unknown values based on known (e.g. classification or regression).
- **Descriptive** tasks: the major goal is to extract patterns that define the correlation in data (e.g. association analysis).

Data mining can be used for tasks such as classification, clustering and outlier analysis. The most common data mining tasks are (Zhao, 2013):

1. **Association rule mining:** the task of discovering interesting relationships between items in a given transactional dataset (HarvinderChauhan and AnuChauhan, 2014). Useful applications of this task are market basket analysis and determining web pages that are accessed together.
2. **Classification:** The classification model is discovered from analysis of a certain dataset. Useful applications of this task are fraud detection, medical diagnoses, email categorization, etc.
3. **Clustering**: the task of discovering groups of closely related objects (Chattopadhyay et al., 2020), where the objects that belong to the same group (cluster) are more related to each other than to objects belonging to another group. Unlike classification, clustering analyses and classifies object without using class labels (Saturi, Dara and Prem Chand, 2019). A number of researchers use the term "unsupervised classification" instead of clustering to distinguish between clustering and classification. Useful applications of clustering include market segmentation to distinguishing characteristics of a group of customers.
4. **Outlier analysis**: Outliers are intense values that diverges from other observations on data, they may show a variability in a measurement, experimental errors or a novelty. In other words, an outlier is an observation that deviates from an overall pattern on a sample. They are sometimes discarded as a noise. In some applications the analysis of outlier objects is essential, such as fraud detection application where the fraud is an exception.

2.1.3 Classification

Classification aims to predict categorical labels for unseen data. Classification is the process of predicting the class of given data points. Classes are sometimes defined as targets, labels or categories (Wedyan, 2014). Supervised learning has two steps: the first is the learning step (also called training step) where the set of characteristics are selected from the training data that represent and describe these data; this set of rules is called a classifier. Training data is the major and most important data which assists machines to learn and make the predictions. This data is used by machine learning to develop a certain algorithm and more than 70% of the total data used in the project. A big quantity of datasets is used to train the model at the best level to get the best results. Because the class label exists in the training data the classification is considered a supervised learning.

In this section we will review the most popular supervised classification techniques briefly, which are decision tree learning, naïve Bayes, k-nearest neighbour, support vector machine, and neural network.

2.1.3.1 Decision Tree learning

A decision tree is a decision support tool that has a tree-like graph or model of decisions and their possible significances, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only has conditional control statements. A decision tree is also a flowchart-like structure in which each internal node represents a "test" on an attribute. The first decision tree algorithm was ID3, developed by Quinlan during the late 1970s, and later Quinlan developed C4.5 (Kohavi and Quinlan, 1999). In C4.5 the decision tree is built in a top down recursive approach, where the algorithm starts by selecting a training set of items and the class label which is associated with the items. It chooses the attribute that best discriminates the items (i.e. the training dataset). After that it creates a tree node whose value is the chosen attribute. Then the child node is created for the selected node where each node represents the unique value for the chosen attribute as the "best" one. The items of the training set distribute to the children based on the value of the children. The algorithm is then applied in recursive manner at each child node. Hence, the main goal is to maximize data generalization in the decision tree to reduce the number of tree nodes and tree level in order to load the tree in the memory. So, to do that we need to select the "best" attribute which discriminates the items. Additionally, decision tree-based algorithms use different techniques to select a suitable attribute, for instance C4.5 uses Information Gain (IG) measure for selection process, a measure taken from information theory. Decision tree-based algorithms receive much attention because they have been successfully applied to real problems, also the outcome of the decision tree is explicit and interpretable as the decision tree is a set of rules.

2.1.3.2 Naïve Bayesian (NB) Classification

Bayesian classifier is a statistical classifier; based on Bayes' theorem (Stern, 2015). It predicts the probability of a class of given data. Naïve Bayesian classifier proposes the effect of an attribute value on a given class to be independent of the value of other attributes; this proposal is called class conditional independence (Schetinin et al., 2007). Let X denote to attribute-set and Y denote to class. In Bayesian statistics, the posterior probability of a random event or an uncertain proposition is the conditional probability that is assigned after the relevant evidence or background is taken into account. "Posterior", in this context, means after taking into account the relevant evidences related to the particular case being examined. For instance, there is a ("non-posterior") probability of a person finding buried treasure if they dig in a random spot, and a posterior probability of finding buried treasure if they dig in a spot where their metal detector rings (Schetinin et al., 2007).

The conditional probability independence is expressed formally in equation 2.5:

2.5 $$P(X/Y=y) = \prod_{i=1}^{d} P(x_i/Y=y) = P(x_1/y) * P(x_2/y) * \ldots\ldots * P(x_d/y)$$

to classify a test data. The naïve Bayes classifier computes the posterior probability for each class Y using equation 2.6.

2.6 $$P(Y/X) = \frac{P(Y) \prod_{i=1}^{d} P(X_i/Y)}{P(X)}$$

Then the classifier predicts the class ci of the data X if P (xi /ci) P(ci)> P(X /cj)P(cj); that means the ci is chosen as the class label of data x if it has the maximum value (Sivasankar and Rajesh, 2012). The naïve Bayes classifier requires a small training dataset. Naïve Bayes also can handle the noise and missing data.

Bayesian classifier theoretically has the minimum error rate when compared with other classifiers. Several empirical studies found the naïve Bayes classifier is comparable in some applications with decision tree and neural network.

2.1.3.3 K-Nearest Neighbor Algorithm (KNN)

K-nearest neighbour (KNN) algorithm is a classification method. It classifies objects based on the closest (nearest) training examples (Richman, 2011). Informal interpretation of KNN for classification is the following saying: "When I see a bird that walks like a duck, and quacks like a duck and looks like a duck, this bird probably is a duck." In KNN, the object is classified by the majority voting of its neighbour, i.e. the object is classified to the most common class for its k-nearest neighbours (Chen and Shah, 2018). Figure 2.3 illustrates nearest neighbour of the circle object.

A KNN classifier hasn't got an explicit training phase. It is an example of lazy learning, which does not require re-training when new data becomes available.

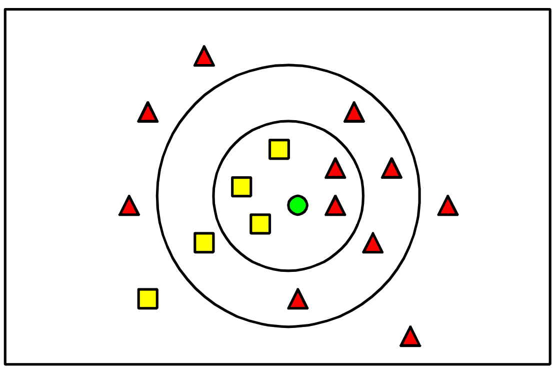

Figure 2. 3 K-Nearest Neighbour

The nearest neighbour classifier is based on local information, not like a decision tree where it tries to find a global classifier, therefore nearest neighbour classifier is apt to noise, especially if k is small. KNN is applied in many applications such as text categorization problem and it shows promising results when 'compared with other statistical methods like Bayesian (Al-Shalabi, Kanaan and Gharaibeh, 2006).

2.1.3.4 Support Vector Machine

The support-vector machine is introduced in (Vapnik, 1995; Vapknik, 1998a; Vapnik, 1998b) as a new learning machine for two-group classification problems. In the field of machine learning, support-vector machines (SVMs, also support-vector networks) are presented as supervised learning examples with associated learning algorithms that considers data used for classification and regression analysis.

The Support Vector Machine (SVM) algorithm is a general machine learning instrument that presents solutions for both of classification and regression problems. SVM is also firstly presented at AT&T Bell Laboratories by Vapnik with colleagues (Boser et al., 1992, Guyon et al., 1993, Vapnik et al., 1997), the statistical learning framework or VC theory recommended by Vapnik and Chervonekis (1974) and Vapnik (1982, 1995) shows that SVM offers one of the most robust prediction techniques.

SVM model is a representation of the samples as objects in space, charted so that the samples of the discrete categories are separated by a clear gap that is as extensive as possible. New examples are then charted into that same space and predicted to a certain category based on the side of the gap on which they fall.

Additionally, SVMs can effectively achieve a non-linear classification by the kernel trick (Schölkopf, 2001), implicitly charting their inputs into high-dimensional feature spaces.

Also when data is unlabelled, supervised learning is not achievable, in this case an unsupervised learning approach is needed, which tries to reach natural clustering of the data to clusters, and then chart new data to these formed clusters. The support-vector clustering algorithm, created by Hava Siegelmann and Vladimir Vapnik, employs the statistics of support vectors, presented in the support vector machines algorithm, to classify unlabeled data, and is one of the most widely applied clustering algorithms in industrial applications (Ben-Hur et al., 2001).

2.1.3.5 Artificial Neural Network

Artificial neural network (ANN) come from attempts to simulate the biological neural system. The human brain consists of neurons linked together via axons. Each neuron is connected to the axon of other neurons via dendrites. The point of convergence between axon and dendrites is called a synapse. Scientists discovered that the learning process of the brain is carried out by changing the strength of synaptic connections between neurons (Agatonovic-Kustrin and Beresford, 2000). ANN corresponds to a set of nodes and links. ANN processes a certain instance at a specific time, classifies the record, then it compares the classification process with the actual class of the instance. If the error exists in classification, the error is fed back to the network to modify the second iteration, and so on.

Artificial neural networks are computational models that act similarly to a human nervous system. There are different types of ANNs. These types of networks are applied based on varied mathematical operations and a group of parameters needed to decide the output. These types are:

- Feedforward Neural Network – Artificial Neuron:

This neural network is a simple type of ANN, as the data or the input goes in one direction. The data goes through the input nodes and exit on the output nodes. This neural network may or may not have the hidden layers. This type of neural networks is applied in computer vision and speech recognition as the classification of the target classes is complicated (Hagan and Menhaj, 1994).

- Radial basis function Neural Network:

This type considers the distance of a point with respect to the centre. It has two layers, first where the features are joined with the Radial Basis Function in the hidden layer and then the output of these features are considered while calculating the same output in the next time-step which is essentially a memory (Matera, 1998).

- Kohonen Self Organizing Neural Network:

Kohonen map objects to input vectors of random dimension to discrete map included neurons. The map is required to be trained to provide its own group of the training data. It contains either one or two dimensions (Ultsch and Siemon, 1990).

- Recurrent Neural Network(RNN) – Long Short Term Memory:

It works on the rule of keeping the output of a layer and feeding this back to the input to help in predicting the outcome of the layer (Cho et al., 2014).

- Convolutional Neural Network:

It is similar to feed forward neural networks, as the neurons have weights and biases those can learn. Its application has been in signal and image processing in the field of computer vision (Kalchbrenner, Grefenstette and Blunsom, 2014).

- Modular Neural Network:

Modular Neural Networks have a group of various networks acting independently and contributing towards the output. Each neural network has a group of inputs that are unique compared to other networks constructing and performing sub-tasks. These networks do not interact or signal each other in achieving the tasks (Azam, 2000).

2.2 Educational data mining:

“EDM converts raw data coming from educational systems into useful information that could potentially have a greater impact on educational research and practice.” (Peña-Ayala, 2014)

Data mining is strongly used in education to discover helpful information in order to improve the quality of education (Ramanathan, Geetha and Khalid, 2015). EDM uses educational databases, such as admission systems, learning management systems or any other systems related to the students. Analysing the extracted data from these systems can classify the students, or make a decision to improve student performance (Bhaskaran, Lu and Al Aali, 2016) .

Romero and Ventura (2007) sort work-educational data mining into the following groups:

- Statistical and visualization tasks.
- Web mining tasks:
    - Clustering, Classification, and outlier detection.
    - Association rule mining and sequential pattern mining.
    - Text mining.

The above perspective is concentrated on EDM applications that use extracted data from the web.

Another classification of EDM has been done by Baker (2009), which categorizes work in educational data mining as follows:

- Prediction
  - Classification
  - Regression
  - Density estimation
- Clustering
- Relationship mining
  - Association rule mining
  - Correlation mining
  - Sequential pattern mining
  - Causal data mining
- Distillation of data for human judgment
- Discovery with models (Romero, Ventura and García, 2008).

Baker has added two more extra categories - distillation of data for human judgment and discovery with models, which are not in the common DM category. The use of models for information discovery has been used widely as a technique in EDM research and is applied to enhance complicated analyses.

The rest of this section will show the most important techniques and research of data mining and EDM. The demonstration will be according to the above-mentioned categories.

2.2.1 Statistics and Visualization:

Statistics in educational data mining have been used as standard tools developed to analyse the extracted data from log files and web servers such as Access-Watch, Analog, Gwstat, Web-Stat, etc. (August 2012). On the other hand, there are some specific statistical educational data mining tools such as Synergo/ColAT (Romero and Ventura, 2007). For example, Pahl and Donnellan (2002) used statistics to measure visits to individual pages, average session length in time or in number of requests. The authors presented a set of mining methods suitable for the educational situation. They divided the usage evaluation of web-based systems into two dimensions: time and space (Pahl and Donnellan, 2002). Meanwhile, Silva and Vieira (2002) developed a tool called MultiStar for knowledge discovery. MultiStar is used to analyse data involving facts that belong to the same category. The way of representing the facts as hierarchies involving the relationship of inheritance among the facts does not need the user to understand the concept on which it is based. Using this tool provides information about the student's actions and communications, and information about the student's activities in the course (Silva and Vieira, 2002). Thus, some other general statistics tools have been utilized to show the connected learner distribution over time, the most frequently joined courses, and how many learners log in to learning sessions over time (Pérez Zorrilla, 2005). On the other hand, some specific statistical tools in adaptive and intelligent web-based educational systems (AIWBES) can demonstrate the average number of constraint violations, the degree of problem complexity, and the total time spent in the trials (Nilakant and Mitrovic, 2005). Furthermore, another research study discussed the complexity and benefits of a huge data amount collected automatically by an intelligent tutoring system. The researcher conducted simple statistical operations such as standard deviation, mode, sample size, etc. They showed the importance of analysing large volumes of data compared to few data (Boyer, 2010).

The information extracted from statistical operations is complicated to explain to teachers. For that reason, information visualization techniques can be used to convert complex and numerical student tracking data extracted from web-based educational systems into graphical representation. These techniques help in analysing large amounts of information by representing the data in some visual displays, such as diagrams, graphical trees, spreadsheet charts, scatterplot, 3D representations, etc. (Mazza and Milani, 2005). There are many examples of visualization tools in educational data mining, such as GISMO (Graphical Interactive Student Monitoring System) (Romero, Ventura and García, 2008). This tool allows the user to display data from courses collected in real settings as graphical representations. The purpose of building such a tool is to help instructors monitor what is happening in distance learning classes (Mazza and Milani, 2005). Nowadays, most LMSs contain GISMO tools and plugins. Moodle, which is a well-known LMS and widely used, provides a GISMO tool for analytic purposes. Figure 2.4 shows students' login overview screen in Moodle's GISMO tool (Mazza and Milani, 2004).

Figure 2. 4: Students login overview screen in Moodle's GISMO

One more research study developed a tool called LISTEN to browse a database of students' interactions with an automated tutor. Using databases logged by LISTEN's Reading Tutor, they demonstrated how student interaction events occurred and represented them as a graphical expandable tree (Mostow and Beck, 2001). Furthermore, a group of researchers developed a tool to allow the discovery of information in Moodle data. They conducted an experiment to demonstrate the ease of obtaining data visually and quickly for the user. The results provide useful information so that users can provide faster feedback on students' progress in distance e-learning courses (Kuosa et al., 2016). Another research paper created a visualization tool that allows user to filter numeric variables by one or more, which makes the chart respond immediately according to filtered variables. Data are represented as a 2D scatterplot and some parameters of the representation can be interactively amended (Chan, Correa and Ma, 2013).

2.2.2 Web mining:

According to Romero and Ventura (2013), there are three categories that have been used in web mining: web content mining, which is the process of extracting useful information from the contents of web pages; web structure mining is the process of finding structure information from the web; and lastly, web usage mining (WUM) is the finding of meaningful patterns from data generated during the process of communication between client and server (Romero and Ventura, 2013). There are several web mining techniques used on educational applications, most of which can be grouped into the following categories: clustering, classification and outlier detection; association rule mining and sequential pattern mining; and text mining.

2.2.2.1 Clustering, classification and outlier detection:

Clustering is a process of collecting physical objects into classes that have similar objects. Both clustering and classification are considered as classification techniques. Clustering is an unsupervised classification and classification is a supervised classification. Furthermore, an outlier is a measurement that is unusually large or small relative to the other values in a dataset. Outlier methods are usually used when measurements are observed, recorded, or entered into the computer wrongly. Also the outliers' detection can be used if the measurements come from another population or represent unusual events (Romero and Ventura, 2013). Following is an illustration about some work that applied these techniques. First we demonstrate clustering techniques. Over the last few decades, a set of research studies use clustering to mine student data in order to discover patterns reflecting student behaviours. The presented models focused on collaboration management to distinguish similar behaviour groups in unstructured collaboration spaces. They emphasized that the data used in this model should be carefully designed and tuned to contain all useful information (Romero et al., 2009) because the main task of clustering is to group objects based on their similarities (Talavera and Gaudioso, 2004). Accordingly, Mor and Minguillón (2004) use a clustering technique for grouping students based on SCORM standard paradigm. They proposed a method for analysing student behaviour through interacting with an e-learning system to include the concept of recommended itinerary, by combining instructor's' expertise with learned experience acquired by system usage analysis (Mor and Minguillón, 2004). The main goal of such studies is to design a usage mining tool for analysing such student navigational behaviour and for extracting applicable information that can be implemented to validate varied aspects related to virtual campus structure and usability but also to decide the optimal scheduling for each course based on user profile. The combination of teachers' expertise with learned experience acquired by system usage analysis came obvious. While another proposed models tended to detect a typical behaviour in the grouping structure of the students of a virtual campus. They used a clustering and a generative topographic mapping model to conduct experiments using the extracted data. The experiment results indicated that the proposed model neutralizes the negative impact of outliers on the data clustering process (Castro et al., 2005). More recently, two clustering methods; Hierarchical Clustering (Ward's clustering) and Non-Hierarchical Clustering (k-Means clustering) are presented the literature. Researchers built profiles of student behaviour from learner's' activity in an online learning environment and also to create click-stream server data. They proposed a new approach for implementing data mining services in SCORM-compliant LMSs (Psaromiligkos et al., 2011). Accordingly, several clustering techniques such as Expectation Maximization, Hierarchical Clustering, Simple k-Means and X-Means as provided in WEKA software are proposed. They applied these methods to predict the potentiality of students' performance: who could fail during online courses in a Learning Management System (LMS) to fill in the research gap of system integrations. Experiments in clustering were conducted using real data obtained from various online courses. That is why the authors have compared several classic clustering algorithms on several group of students using their defined features and analysed the meaning of the clusters they produced. (Bovo et al., 2013). Despite the fact that many methods have been proposed to classify data in order to analyse and predict facts and relationships, classification is the most important technique in data mining. In this study we used classification techniques in data mining to predict the employability of students (Syeda Farha Shazmeen, 2013). Many researchers have come to consider the idea of a classification technique is to categorize an attribute value into one of a group of possible classes and answer certain research questions that incorporate information to the readers such as classification definition (Wedyan, 2014). Therefore, Classification techniques are presented as supervised learning methods that classify data items into predefined class labels (Syeda Farha Shazmeen, 2013). Additionally, several techniques and algorithms have been proposed to perform classification tasks, some of which are listed below:

1. Decision Trees
2. Artificial Neural Networks
3. Support Vector Machine
4. K-Nearest Neighbor

5. Naïve-Bayes

These techniques enhance the efficiency of the algorithm to classify the data properly (Syeda Farha Shazmeen, 2013). For that reason, classifications have been widely used in EDM due to their high accuracy in prediction. In this regard, Chen et al. (2000) used a decision tree technique and applied C5.0 algorithm and data cube information processing methodologies to monitor students' behaviours and find the pedagogical rules on students' learning performance from web logs. The induction analysis found potential student groups that have shared characteristics and reaction to a certain pedagogical strategy (Chen et al., 2000). After few years, another research study used four classifiers to categorize students based on features extracted from the logged web data in order to predict their final grades. By using a genetic algorithm weighing the features, they optimized a combination of classifiers (Minaei-Bidgoli and Punch, 2003). Hence, researchers applied a Bayesian classification technique to a student database to predict students' performance with emphasis on identifying the difference between high fast learners and slow learners. According to the authors, the results of their work help to identify students who need special attention to reduce failing (Ahmed and Elaraby, 2014). Data mining is widely used in the educational field for different purposes, one of the most important topics are discussed such as using several classifications to study students' performance, focusing on several factors that may impact students' performance in higher education. Qualitative predictive models which were effectively able to predict students' grades from a training dataset are presented. In one study, four decision tree algorithms were applied, in addition to the Naïve Bayes algorithm. The study found that not solely academic efforts impacted the students' performance, but that many other factors also have an influence (Abu, 2016). However, Ahmed et al. (2014) implemented the decision tree (ID3) method as a classification technique in data mining to predict students' final grades. They used different factors that were collected from the student's' database. According to the authors, the study assists to enhance the student's' performance and reduce the failure rate (Badr, Din and Elaraby, 2014). Unlike classification and clustering techniques, outlier detection methods handle unusual data, and can detect students with learning problems (Kou and Lu, 2016). Thus, another research study applied two classification methods, Rule Induction and Naïve Bayesian classifier. The dataset was collected from graduate students' data collected from the college of Science and Technology – Khanyounis. The students are clustered into groups using K-Means clustering. To detect all outliers in the data, they used Distance-based Approach and Density-based Approach (Tair and El-Halees, 2012). On the other hand, a model for automatic analysis of student interactions with a web-based learning system is proposed in one more study. The result of automatic analysis provides useful information including decision trees that were used for predictions in later experiments. Also the generated decision trees were used in analysing data using machine learning techniques (Muehlenbrock, 2005). In this regard, Ueno (2004) proposed a method of online outlier detection of learners' irregular learning processes by using a Bayesian predictive distribution. The outlier detection method uses a students' response time data for the e-learning contents. This method can be used for short samples, and it helps a two-way instruction by using the results of mining processes (Ueno, 2004).

2.2.3 Text mining

Text mining focuses on web content mining, and uses text data to perform the data mining tasks. It is used in many fields of our life such as are highlighted as a bridge between free-text and well-structured information in healthcare to represent cancer information. Many different techniques could be used to achieve text mining tasks such as machine learning, data mining, statistics, information retrieval and natural language processing (Spasić et al., 2014). Text mining can use unstructured or semi-structured text data such as full-text documents, HTML, XML, and JSON files (Grobelnik, Mladenic and Jermol, 2002). Furthermore, it is presented in a group of research studies as a method for building an e-textbook using web content mining. The generated e-textbook allows systematic learning over the messy web. The authors used a ranking technique to filter descriptive web pages' content in order to estimate the web pages' suitability and they produced concept features and built user-specified topic hierarchy (Chen, Li and Jia, 2005). On contrast, Tane, Schmitz, and Stumme (2004) developed an ontology-based tool that allows web resources' availability. They used text mining and text clustering methods to find and organize material using semantical information in order to categorize information due to their subjects and relationships (Tane, Schmitz and Stumme, 2004). Moreover, another approach is presented to apply text mining as a method for evaluation of asynchronous conversational forums to obtain important information from a negotiated discussion forum depending on Allaire's Cold Fusion forum software. The authors used temporal contribution guides to discern how text mining different techniques in the query process could enhance the instructor's skills to assess the progress of the discussion (Ellis and Dringus, 2005). More recently, Hammouda and Kamel (2008) proposed a technique to apply text mining to documents. This helps as a root for knowledge extraction in e-learning settings. They used graph theory to facilitate phrase representation and efficient matching. By using this text mining technique, documents are grouped according to their topic. The recent literature also proposed an efficient Lovins stemmer in combination with snowball-based stop-word calculation and word tokenizer for text pre-processing. Researchers conducted several experiments on publicly available very well-known text datasets and opines the effectiveness of the proposed approach in terms of accuracy, F-score and time, in comparison with many baseline methods available in the recent literature (Panda, 2018).

2.3 Neuro-Fuzzy Approach:

In this research, we adopt a neuro-fuzzy approach to perform the classification task. Therefore, we will study the literature of neuro-fuzzy in this section and illustrate how several studies defined neuro-fuzzy and use it in classification. Also, we discuss how they build a prediction model using a neuro-fuzzy classifier.

The first research using the neuro-fuzzy system started at the beginning of the 1990s, with Jang, Lin and Lee in 1991, Berenji in 1992 and Nauck from 1993. Most of the first applications were in process control. Its application is known for all the fields of knowledge such as data analysis, classification techniques, imperfections detection and support to decision-making (Jang, 1992). While neuro-fuzzy is a combination of neural networks and fuzzy set approaches, the idea of neuro-fuzzy uses heuristic learning techniques derived from the domain of the neural network approach to assist the development of a fuzzy system. Modern neuro-fuzzy systems are usually described as a multilayer feed-forward neural network, but fuzzifications of other neural network structures are also considered, for example, self-organizing attributes maps (Seising, 2007). According to the literature, the neuro-fuzzy systems include the following properties:

- A neuro-fuzzy system is a fuzzy system that is trained by a learning algorithm (usually) obtained from neural network approach. The (heuristical) learning procedure operates on local information and causes only local amendments in the underlying fuzzy system. The learning process is not knowledge-based, but data-driven (Vieira, Dias and Mota, 2004).
- A neuro-fuzzy system can be viewed as a three-layer feed-forward neural network. The units in this network use t-norms or t-conorms instead of the activation functions shared in neural networks. The first layer describes input variables, the second (hidden) layer indicates fuzzy rules, and the third layer represents output results (Çaydaş, Hasçalik and Ekici, 2009) .
- A neuro-fuzzy system can always be analysed   a system of fuzzy rules. It is both possible to generate the system out of training data from the beginning, and to initialize it by prior knowledge in the shape of fuzzy rules (Yu-Chi Ho, 2005).

- The learning process of a neuro-fuzzy system takes the semantical features of the underlying fuzzy system into consideration. This results in constraints on the possible enhancements applicable to the system parameters.
- A neuro-fuzzy system estimates an n-dimensional function that is partially provided by the training data. The fuzzy rules encoded within the system represent fuzzy samples, and can be formed as unclear prototypes of the training data. A neuro-fuzzy system should not be shown as a kind of (fuzzy) expert system, and it has nothing to do with fuzzy logic in the strict sense (Jang and Sun, 1995).

The main idea of neuro-fuzzy combinations is to obtain adaptive systems that can use prior knowledge and can analyse by means of linguistic rules. Neuro-fuzzy models can be separated into cooperative models, which use neural networks to define fuzzy system parameters, and hybrid models which generate a new structure using concepts from both techniques. In addition, there are concurrent neural/fuzzy models that use neural networks and fuzzy systems separately. Most approaches adapt the backpropagation learning rule (Nauck and Kruse, 1997).

Most of the combinations of techniques based on neural networks and fuzzy logic are known as neuro-fuzzy systems. The different combinations of these techniques can be classified, in accordance with in the following types:

Cooperative Neuro-Fuzzy System: In the cooperative systems there is a prior process phase where the neural networks techniques of learning define some sub-blocks of the fuzzy system. For example, the fuzzy sets and/or fuzzy rules (fuzzy associative memories or the use of clustering algorithms to determine the rules and fuzzy sets position). After the fuzzy sub-blocks are calculated the neural network learning methods are taken away, executing only the fuzzy system (Vieira, Dias and Mota, 2004).

Concurrent Neuro-Fuzzy System: In the concurrent systems the neural network and the fuzzy system perform continuously together. In general, the neural networks pre-process the inputs (or post-process the outputs) of the fuzzy system. A concurrent system is not exactly a neuro-fuzzy system indirect way, because the neural network performs together with the fuzzy system.

Hybrid Neuro-Fuzzy System: In this category, a neural network is employed to learn some inputs of the fuzzy system (parameters of the fuzzy sets, fuzzy rules and weights of the rules) of a fuzzy system in a repeated fashion. Most of the researchers applies the neuro-fuzzy technique to refer only to hybrid neuro-fuzzy systems. There are several techniques to create hybrid neuro-fuzzy systems; as this is a recent research subject, each researcher has proposed their own particular models. These models are similar in their core, but they introduce basic uniqueness (Vieira et al., 2004).

There are several neuro-fuzzy architectures, such as:

Fuzzy Adaptive Learning Control Network (FALCON) was proposed by Lin and Lee (1991). They introduced an approach based on fuzzy control systems and decision/diagnosis methods. This approach differs from classic fuzzy logic control and decision method in its network design and training potential. FALCON consists of five layers, including multiple linguistics nodes for every output variable. The first node is for the patterns and the second node is for the actual output of the FALCON. The initial invisible layer performs the matching process of the input values related to each membership actions. The second invisible layer sets the predecessors of the rules. The third invisible layer includes the consequents. FALCON uses a crossbred learning algorithm consisting of unsupervised learning to set the initial membership procedures and initial rule base and it employs a learning algorithm based on the gradient descent to adapt the final variables of the membership functions to provide the wanted output (Lin, Lin and Lee, 1995).

Berenji and Khedkar (1992) proposed Generalized Approximate Reasoning based Intelligence Control (GARIC) (Mishra, Sahoo and Mishra, 2018). GARIC is a model for learning and adapting a fuzzy logic controller based on reinforcements from a dynamic system. The GARIC applies a Neuro-fuzzy framework with multiple neural systems modules, these are; ASN Action Selection Network (ASN) and evaluation method which is called Action State Evaluation Network (AEN). The AEN is an adaptive estimator of ASN functions. The GARIC ASN is a developed system consisting of five layers. The layers' relations are not ranked. The initial invisible layer has the linguistic values for each input attribute. Every input can only link to the initial layer that describes its relation to linguistics variables. The invisible layer number two describes the fuzzy rule attribute that assign the connection level of every rule applying a soft-min operator. The invisible layer number three describes the linguistics variables of the output attributes. The result of every rule is computed based on the power of the predecessors of the rules computed in the rule nodes. GARIC applies the mean of the local mean of the extreme way to compute the output of the rules. This technique requires a numerical value in the removal of every rule. Therefore, the outcomes should be converted from fuzzy variables for numerical variables before being gathered in the final output variables of the system. GARIC applies a combination of gradient descending and reinforcement learning for a fine adaptation of its inner parameters (Zarandi, Jouzdani and Turksen, 2007).

Nauck et al (1996) proposed a NEFCLASS model which is used to derive fuzzy rules from a group of data that can be divided into several crisp classes. The goal of NEFCLASS is to introduce an interpretable fuzzy classifier (Nauck and Kruse, 1999) .

Nauck and Kruse (1999) presented learning algorithms to work out the architecture and the inputs of a fuzzy system to approximate a function given by a supervised learning problem. The resulting NEFPROX model is an enhancement of the NEFCON and NEFCLASS approaches. This approach is used for control or classification purposes (Nauck and Kruse, 1999).

Tano` et al. (1996) developed FINEST (Fuzzy Inference Environment Software with Tuning) (Oyama et al., 1995). The major features are (1) enhanced generalized modus ponens, (2) technique that can tune the inference method as well as fuzzy predicates, and (3) software environment for debugging and tuning. They give an outline of the software and show an important concept, a deep combination of the fuzzy inference and the neural network in FINEST, which makes FINEST tune the inference way itself. According to the authors, now it can be used as a tool for quantification of the meaning of natural language sentences as well as a tool for fuzzy modelling and fuzzy control (Arnould et al., 1995).

Sulzberger et al. (1993) proposed a fuzzy net approach (Siddique and Adeli, 2013). This technique is introduced for the optimization of fuzzy rule-based systems using neural networks. A neural network model with particular neurons was provided so that the conversion of fuzzy rules and membership functions into the network can be done. The efficiency of this technique, and hence the results of the original rule base, is then enhanced by training the network by combining neural network learning algorithms. The optimized rules and membership functions can be derived from the net and used in common fuzzy inference methods. This net has been evaluated on the Wall-Jumper-Over and the problem of local navigation for mobile robots (Jang, 1993).

Dynamic Evolving Neural Fuzzy Network (dmEFuNN) (Nauck and Kruse, 1996) modified EFuNN with the concept, that not only the gaining rule attribute's activation is increased but a group of rule features that is dynamically chosen for each new input attribute and their activation values are employed to compute the dynamical variables of the output function. Whereas EFuNN applies Mamdani-type fuzzy rules, dmEFuNN applies Takagi Sugeno fuzzy rules (Nauck and Kruse, 1997).

Ghosh *et al* (2000) proposed a method of performing the neuro-fuzzy classification task; the main idea is to employ the function of feature-wise level of classes' patterns that are extracted from a fuzzification function. A membership matrix created by a fuzzification process has a total number of items equal to the number of attributes product and classes exist in the data set. These matrix members are the input to neural systems. Building this method with different four data sets and multiple remote sensing images makes this method robust and efficient (Ghosh, Uma Shankar and Meher, 2009).

Hoffmann (2004) proposed a new approach using genetic algorithm and fuzzy-rules classifier. This approach iterates the rule to generate fuzzy rule base system prototype. The fuzzy rule base is created in a gradual way, in that the evolutionary algorithm enhances one rule of fuzzy classifier at a time. The boosting technique decreases the weight of the learning items that are categorized properly by the new rule. The following generation of rule cycle concentrates on fuzzy rules that explain the incorrectly classified variables. Assigning weight for fuzzy rules represents the proportional power the boosting method defines to the rule class when it gathers the cast votes (Hoffmann, 2004).

Another important neuro-fuzzy architecture is Adaptive Network-based Fuzzy Inference System (ANFIS) (Jang, 1992). In our study, we will apply ANFIS to build a classification model. Therefore, we will focus on the studies that used an ANFIS approach.

ANFIS applies a conclusion framework known as the Takagi Sugeno technique. This method contains five levels, and the initial invisible level is in charge of comparing the values of the given attribute connected to every membership task. The second invisible level which is called T- norm factor is used to compute the predecessors of the rules. The invisible level number three handles the rules' powers. The invisible level number four defines each rule's outcomes. The last level which is the output level computes the overall output by gathering all entered indexes to this level (Jang, 1992).

The adaptive Neuro-fuzzy inference system applies the backpropagation training method to assign the elements of membership procedure and the least mean square technique to assign the filtered parameters. Every phase of the repeated backpropagation method includes two sections. In section number one, the tuning values are increased and the output variables are computed via the iterative minimum squared technique, whereas the global network variables can be a constant value. In the section number two, the tuning values are increased one more time; for each loop, the learning backpropagation method is employed to amend the premises variable, whereas the consequents keep constant (Vieira et al., 2004)

ANFIS techniques have been used in several fields such as the medical field. ANFIS also has provided useful information that assists in decision-making. It is considered a helpful method to classify patients according to their diagnoses. Hence, ANFIS methods have been used to predict having certain diseases. For instance, proposed a classification model based on ANFIS is proposed for heart disease prediction. Data was collected from the University of California at Irvine (UCI) machine learning repository (Kalaiselvi and Nasira, 2014). Seven attributes were used as input for the prediction method. To evaluate the ability of the trained ANFIS techniques to discover the heart disease diagnosis, they used k-fold cross validation technique (Ziasabounchi and Askerzade, 2014). Manogaran, Varatharajan and Priyan (2018) presented Multiple Kernel Learning with Adaptive Neuro-Fuzzy Inference System (MKL with ANFIS) based deep learning method for heart disease diagnosis. MKL technique is used to separate parameters between heart disease patients and usual individuals. The proposed MKL with ANFIS is also compared with various existing deep learning methods to evaluate its performance in prediction (Manogaran, Varatharajan and Priyan, 2018). Many studies have applied ANFIS classifiers for prediction of the cancer disease: Kalaiselvi and Nasira (2014) proposed a classification technique based on ANFIS approach to predict having cancer by discovering a complex relationship between diabetes and cancer. They trained the model using ANFIS technique. Seven features have been selected to act as input to the network. According to the authors, the presented model decreases the costs of several medical tests and assists the patients to conduct precautionary measures well in advance (Kalaiselvi and Nasira, 2014).

Mahmoudi, Lahijan and Kanan (2013) introduced a method that employs both Genetic Algorithm and ANFIS. They used ANFIS as a classifier for chosen genes from Particle Swarm Genetic Algorithm methods, applying six datasets of microarray gene expression data for several cancer cases. The performance of ANFIS is compared with several classifiers, include; K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Classification and Regression Trees (CART).

According to the authors, ANFIS performs the best results for actual data of all the datasets. ANFIS performs well for a fewer number of genes (Mahmoudi, Lahijan and Kanan, 2013). Furthermore, several breast cancer prediction models have been built based on ANFIS. Atashi, Nazeri, Abbasi, Dorri and Alijani (2017) proposed an ANFIS model for breast cancer detection. In this study the risk factors were divided into three priorities according to their degree of risk level, the fuzzified and the subtractive clustering method were used for their input with the same order. Randomly, the dataset was divided into two sets of 70% and 30% of the total items, and used for training and testing the new model. After the training, the system was separately tested with the Wisconsin and real clinical data; the system achieved high performance using the second dataset (Atashi et al., 2017). Huang, Hung, Lee, Li, and Wang (2012) proposed a CBR system with two feature ranking algorithms, PSO-Based ANN and ANFIS for breast cancer classification application. According to the authors, the ANFIS shows characteristics of fast and proper learning with the ability to use both linguistic information and the data information and good generalization capability (Huang et al., 2012).

As shown in the above-mentioned literature, the Adaptive Network based Fuzzy Inference System (ANFIS) has been used significantly in the medical domain which needs accurate results in prediction due to the critical data it used. According to the previous literature, the ANFIS approach has proved good performance and high efficiency in classifying and predicting medical problems.

ANFIS classification techniques have been used widely in EDM, especially in predicting students' performance. The following literature used ANFIS as a classifier in EDM:

Altaher and BaRukab (2017) proposed the Adaptive Neuro-Fuzzy Inference system (ANFIS) to assist students in improving their academic achievements by predicting their academic performance. They built a model consisting of two steps. First, the grades of the students in the previous test are pre-processed by normalizing their grades in order to optimize the accuracy and performance of the prediction. Second, the ANFIS is employed to discover the students' performance in the next semester. They used three separate ANFIS models, ANFIS-GaussMF, ANFIS-TriMF, and ANFIS-GbellMF, that applied various membership functions to optimize fuzzy rules for the prediction process of the students' performance. The experimental results showed the superiority of ANFIS-GbellMF compared with other ANFIS models with a Root Mean Square Error (Altaher and BaRukab, 2017) .

Iraji, Aboutalebi, Seyedaghaee, and Tosinia (2012) proposed a smart method which can divide and categorize students based on students' performance factor. The system is proposed through learning vector quantization (LVQ) methods, based on ANFIS technique to categorize students according to their learning performance factors. In the proposed system, ANFIS achieved the lowest error ratio and can be employed as an efficient classification model for categorizing students (Saber Iraji *et al.*, 2012).

Ghosh *et al* (2014) used an ANFIS approach with several datasets to predict students' performance. The used methodology applied a fuzzification matrix as there was an association between the input patterns and the degree of membership in several classes. According to the level of membership degree a model would be assigned to a certain class. They implemented this technique to 10 benchmark data sets from the UCI machine learning warehouse for classification. They evaluated the proposed model by comparing its achievement with multiple robust supervised classification algorithms, Radial Basis Function Neural Network (RBFNN) and ANFIS algorithm. They then tested the efficiency by using several measurements, like Kappa statistic -accuracy, root-mean square error, TP rate, FP rate recall, precision recall, and F-score (Kar, Das and Ghosh, 2014).

Taylan & Karagözoğlu (2009) presented an approach based on ANFIS to building a fuzzy conclusion network to help in predicting learners' performance. The authors used exam, major, second, final and achievements evolution as attribute inputs, and the output was a learner's academic achievement. The authors applied an actual dataset collected from the learner's performance in a certain course. The results indicated the ability of the ANFIS model to create the similar outputs as the statistical technique (Taylan and Karagözoğlu, 2009).

Yusof *et al*., (2012) used a Neuro-Fuzzy system based on fuzzy set theory to build a model to define a learner's academic performance. The authors applied four features, grade, time, number of ties, and help, to divide learner performance into three classes. In their work, a fully fuzzy rule base has been decreased to a summarized fuzzy rule base. This model was able to define the most effective features that could be employed to describe the decision system (Yusof *et al.*, 2012).

Do and Chen (2013) conducted a comparative study of hierarchical ANFIS and ANN in predicting student academic performance. They used two models, the hierarchical ANFIS and ANN, to predict student academic performance. The performance statistics of the hierarchical ANFIS model were compared to those of the ANN technique by using several measures such as MAPE (Mean Absolute Percentage Error), RMSE (Root Mean Square Error), MAE, and R achieved. The hierarchical ANFIS model gained the best overall performance in all measures (Do and Chen, 2013).

Nagpal, Mehrotra and Bhatia, (2016) proposed a model based on an adaptive neuro-fuzzy Inference system (ANFIS) in the prediction of the anxiety of students, using a hybrid learning algorithm to enhance the prediction based on the conventional model using a questioner. Here, first-order Sugeno fuzzy model considered whose inputs are adapted via a hybrid learning technique. The efficiency of the proposed method was tested in terms of prediction errors. It was discovered that both MAPE and RMSE decreased significantly. The results indicate that the fusion of fuzzy logic and neural network with a hybrid learning algorithm can be very helpful in Psychological research (Nagpal, Mehrotra and Bhatia, 2016).

Pujianto and Sunyoto (2018) proposed an ANFIS model to build a support system for scholarship prediction. This model can act like the human brain's neural network to convert a problem into a form of neurons or equations as a basis for finding a decision or prediction. The authors applied a technique of artificial neural networks applying the backpropagation method with an accuracy of 0.9090 and MAE 0.0001017 on the epoch to 329. They tested the performance using 3000 data via the K-3 Fold Validation test and 0.2 learning rate configuration and 0.2 momentum (Pujianto, Kusrini and Sunyoto, 2018).

Rathore and Jayanthi (2017) used an adaptive neuro-fuzzy inference system (ANFIS) to develop a predication classifier to categorize the students' data taken from placement and training systems from large institutions. The proposed ANFIS prediction model will categorize the student data in a simple way and will be helpful to several educational institutions. Their fuzzy inference system has been used to discover student achievement which will help to find the achievement of all students and give a chance to enhance performance. They categorize the student's dataset for placement and non-placement classes (Rathore and Jayanthi, 2017).

Maitra, Madan and Mahajan (2018) presented an adaptive neuro-fuzzy inference system with backpropagation method in order to build a model for predicting student performance progression, which depends on the trends of previous performances. The model used MATLAB R2018a software and they tested this model by using a real-time data set. Also, the testing phase included comparing the result with other actual data sets from several universities. According to the authors, this model performs well in predicting students' performance (Maitra, Madan and Mahajan, 2018).

Eti (2018) introduced a simplified implementation, simulating ANFIS training errors using data of students results. The major aim of this approach is defining ANFIS training errors. He used a predefined laptop and Matrix Laboratory (Matlab) in the experiment as simulation tools. Matlab as a simulator provided the training error interface, the validation interface, and the fuzzy inference interfaces. The main goal of the proposed models to train the ANFIS provides a training error with 8.3252e-005 confirmed in the first period. This was subsequently followed by an average testing error of 8.325e-005. The minute errors indicate good efficiency using ANFIS.

Shana and Abdulla (2015) used an ANFIS technique to discover the graduation grades of a student in first semester at the university. The dataset used contained 200 students' data from two different colleges. The ANFIS was trained with 150 training samples. They concluded two important points: the collected student data such as grades used for training usually formed as ranges value that is not fit for all machine learning techniques. The second point is that the prediction model uses a neuro-fuzzy approach that could predict the AGPA with a high accuracy rate (Shana and Abdulla, 2015).

Yusof, Zin, Yassin and Samsuri (2009) proposed an evaluation model based on ANFIS to estimate learners' achievement and efficiency level in programming method courses. They used an ANFIS approach to investigate the possible terms to create a fuzzy rule-based system of a learner model. The backpropagation is applied as the learning approach for the neural network to fix the shortage in the decisions taken by human experts. According to the authors, by training the neural network with 18 human experts that are clear, the neural network has correctly obtained other decisions to create a completed fuzzy rule base and can modify its parameter by learning approach. Some of the decisions are illogically categorized (Yusof *et al.*, 2009).

Ouguengay, Elfaddouli and Bennani (2015) studied a certain issue of educational assessment of reading and writing competencies' acquisition assessment in education for the Amazigh language, according to the activities of the students, with exam assessment. This exam consists of varied items and is done in a learning management system platform. They proposed an online model applying artificial intelligence practices, specifically multi-input–multi-output adaptive neuro-fuzzy inference system, also CANFIS for co-active ANFIS. To optimize the performance of the proposed work they used the fuzzy analytical hierarchy process (FAHP) approach for the preparation of fuzzified training data which may enhance the illustration of the linguistics of competency evaluation problem (Ouguengay, El Faddouli and Bennani, 2015).

As stated in the previous literature, the neuro-fuzzy approach has been used significantly in classification problems and has achieved high accuracy in prediction rates for almost all the above-mentioned studies. After reviewing the literature about ANFIS, we have noted that there are no studies using the ANFIS approach to build classifier models for the employability problem. The previous two reasons motivate us to build a classification model that can predict CS and IT graduates employment status using the ANFIS approach.

2.4 Use of Data Mining and Neural Networks in employability:

Data mining has been used in different fields due to its ability to efficiently deal with a huge amount of data. Currently, researchers, universities and higher educational institutions have begun investigating the power of data mining in analysing academic data. The purpose of this thesis is to predict the chance of graduate employability after graduation from Jordanian universities. Although many research studies have attempted to predict students' performance that may lead to employability, employability prediction itself is still not widely known in research fields. Before discussing data mining techniques that have been used in employability we should define the term of employability in general. The following illustrates the definitions and history of employability. Additionally, the problem of CS graduates' employability in a country like Jordan is important because, generally, the expectation from parents, graduates, curriculum designers, and government sponsors and employers is that organizations of higher education should emphasize the alignment between what is learnt at university and what is required by the work force. For graduates, acquiring employability skills is significant because while many employers are willing to fund training to assist employees acquire more sophisticated job-related skills, it has been stated that employers are not prepared to help people develop the skills that are regarded as a basic requirement for employment (Ramlall, 2004). Furthermore, in a developing country like Jordan, many families are dependent on the income of university/college graduates, leading to dissatisfaction when their children find it hard to find employment upon graduation; which made it urgent and complex at the same time to reach the most effective skills affecting the CS employability.

Many classification techniques are used for employment purposes. Bayesian method and the Tree method are used in a study to build a classification model for the employment problem and the data collected from graduate profiles at the Maejo University in Thailand (M. and A., 2016). The authors reached a model to be used for predicting if the graduates would graduate or not, while the current study aims for comparing numbers of DM algorithms to predict CS graduates' employability which would help all of graduates, students, parents, and educational institutions as well. All of the compared algorithms are applied with the neural networks as neural networks are machine learning algorithms that provide accuracy in many cases (Svozil, Kvasnička and Pospíchal, 1997). Also, these techniques are considered as effective in increasing the model's ability to learn features (Neapolitan and Neapolitan, 2018) which is needed for the current research study.

Adding to all of the above, the following points show the importance of using NNs in the current research study;

1. NNs are able to learn and model non-linear and multifaceted relationships (Várkonyi-Kóczy, Tusor and Bukor, 2014), which is really significant because in real life, many of the relationships between inputs and outputs are non-linear as well as complex (Ehret et al., 2015).

2. NNs also can imply hidden relationships on unseen data as well, thus making the model generalize and predict on unseen data (Pimparkar et al., 2019).

3. Unlike many other prediction techniques, NNs do not force any restrictions on the input variables (like how they should be distributed) (Schaap and Bouten, 1996). In addition to that, many research studies have shown that NNs can better model data with high instability and non-constant modification, given its ability to learn hidden relationships in the data without imposing any fixed relationships in the data (Jia et al., 2019).

4. A neural network is a sequence of algorithms that endeavours to consider underlying relationships in a group of data by a function that imitates the way the human brain works. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature (Jang et al., 2019). Neural networks can adapt to altering input; so the network makes the possible result without needing to redesign the output criteria, which made it suitable to be used with the current research problem.

5. The data structures and functionality of neural networks are designed to simulate associative memory. Neural networks learn by dealing with examples, each of which has a known input"; which in our study is "a CS graduate", and "result" is "employed or not in our research" forming probability-weighted relationships between the two, which are saved within the data structure of the net itself.

2.4.1 Data Analytics, Educational Data Mining, and Employability definitions

Many large organizations gather great amounts of data each day in a geographically distributed fashion, at data centers around the globe. Despite their geographically different sources the data must be processed and analyzed as a whole to extract insight. In one study, the author called the task of providing large-scale geo-distributed analytics Wide-Area Big Data (WABD) (Maroosis, 2018). From this perspective, and the growing interest in data and analytics in education, teaching, and learning, the need for high-quality research into the models, methods, technologies, and influence of analytics is increased. Two research groups, Educational Data Mining (EDM) and Learning Analytics, have developed separately to achieve this priority. Literature argues for increased and proper communication with collaboration between these communities in order to share research, methods, and tools for data mining and analysis in the service of developing both fields (Siemens and Baker, 2012). For more than a decade, much business discourse has occurred about the issue of employability specifically about Jordan and, generally, about the world's labour markets. None of the discussions has investigated enough to provide a strong, actionable, framework for labour market stakeholders, education, business, government and skill training suppliers, to take real steps toward better employability. Currently, the term "employability" is reduced to being a word that hardly helps scratch the surface of the mammoth problem. Consequently, handling employability as an analytical task helps researchers get to the multiple attributes that affect it. Moreover, it helps figure the interdependencies among attributes so researchers and scientists can create the significance of cause and effect (Sheikh, 2013).

More recently, a set of research studies used data analytics techniques to find a solution for the employability issue. Twelve strategies are identified in the literature that have been empirically connected to improvements in graduate employability. A survey methodology was applied to examine self-reported use on these strategies among four stakeholder communities. A set of questions was addressed to students and to higher education career improvement professionals and educators and to employers. Across the four stakeholder groups, 705 responses were obtained and analysed. The findings were inconsistencies between the strategies stated in the literature and those indicated in the surveys, as well as discrepancies between stakeholder groups in regard to which strategies were indicated (Kinash et al., 2016).

The term "employability" has so far not had a clear definition, and many researchers have defined the concept of employability differently:

Hillage and Pollard (1998) defined it as a set of skills, understandings and personal attributes that give graduates more chance to get employment and be successful in their chosen occupations, which benefits themselves, the workforce, the community and the economy (Hillage and Pollard, 1998). According to Hillage and Pollard (1998), the following factors can be recognized within most definitions of employability:

- The capacity to obtain an initial employment
- The ability to have a job while at the same time keeping the transitions between jobs and tasks within the same institution to meet new job requirements.
- The ability to find new employment if needed
- The quality of work or employment.

Yorke (2006) defined it as a group of features, competences and knowledge that all labour market members should own to certify they have the capability of being applicable for the workplace to the benefit of themselves, their employer and the institution economy (Yorke and Knight, 2006).

As noted from the above definitions of employability, the factors that may affect employability are varied and are not clear enough, which opens the door for researchers to explore and discover more factors that may influence employability.

2.4.1.1 Conceptual Distinctions about Employability definitions:

The concept of employability has been differently conceptualized and operationalized in a number of managerial and technical research studies that support an explanation with a list of emphasized employability attributes. The following table presents different definitions of employability with focused terms from different points of view:

No	Term	Definition	Source

1	Employability with individual organizational levels	Employability is a set of accomplishm which represent a critical but insuffic circumstances for the acquisition employment (that is dependent on the cur situation of the economy). At the same ti the author defined the term 'employability considerably more multifaceted than so advocates of 'core', 'key' and 'transfera abilities have been recommended by so authors, and is intensely aligned with academic valuing of powerful learning.	(Yorke, 2005)
2	Employability and adaptability	Employability is holding a group of sk knowledge, understanding and indivic characteristics that make an individual m likely to select occupations in which they be satisfied and successful. As an emplo it is important to carry on freque developing the labour market. They will' n to display that they respond to cha positively and can adjust rapidly while working productively to a high level.	

			(Pool, Qualter and Sew 2014b)
3	Employability and entrepreneurship	It has become obvious that the t 'employability' is habitually u inaccurately and carelessly with 'enterpri which in turn is confused v 'entrepreneurship'. Confusion betw enterprise and entrepreneurship acknowledged some years ago, as t suggested that it was likely to discrimi between: 'business entrepreneurship' – encouraging learners to carry out their c businesses; 'working in enterprises' 'being enterprising' – being advan distinguishing opportunities and having r and challenges. The concept 'enterprise' used for a number of years in hig educational institutions to label many ta that we now incorporate with the t 'employability'. Moreover, since the con 'employability' is being used m extensively in the higher educational sec the area for confusion has become bro used and the necessity for clarity more urg	(Sewell and Dacre P 2010)
4	Employability and remaining competitive	Studying at university supports an advant in the future career. Someone's qualificati such as being a graduate, and the sub	

		he/she specialized in or the final grade t attained may be related to some employ Moreover, despite having a first-class deg and a related major for the career they w they will most likely be competing aga other employees who have the same or sim academic qualifications. Therefore, it's t employability, the distinctive mix qualifications, abilities and personal skills they have, which would allow them to distinguished from the group.	(Kundi et al., 2017)
5	Employability and individual goals	People should improve their skills in orde handle the great range of obstacles which presently found in the world: me authorization for a real and democr citizenship, knowledge acquisition discussion for lifetime and life-wide learn 21st century skills for corresponding dem and supply in workforce markets, employability for random future ca success. One area of equipment towa fulfilling these goals is online education which learners have the opportunity to l their own time, content, and objectives.	(Martínez-Cerdá et al., 2018)

Table 2. 1 Employability definitions

2.4.1.2 The History of Employability and its standards in the Middle East:

Employability plays a vital part in updating workforce market policy all over the world's regions (Fransen et al., 2019). A number of research studies have investigated previous implementations of the term of employability and explained its value as an exploratory conception and a theory for policy analysis. Also they have outlined the improvement of the concept, showed its importance in the recent workforce market and its different strategies (with particular reference to the UK) and tried to determine an approach to explaining employability that can better suggest workforce market policy, by exceeding explanations of employment and unemployment status that emphasize either supply-side or demand-side variables (Nickson et al., 2012). Despite that fact that the literature provides a range of definitions of 'employability', many policy-makers currently use the concept as explanation for 'the individual's employability skills and attributes' (McQuaid and Lindsay, 2005).

On the other hand, another number of research papers define employability as conceptualized as a multidimensional form of work-certain adaptability that allows employees to determine and understand career opportunities. According to other research studies, people who hold high levels of employability are expected to earn the reimbursements of active adaptability. Another point cited through this group of research studies is that not only does the person-cantered and active adaptation support the conceptual fundamentals for the paradigm of employability, but it also supports the conceptual cement that adheres the element dimensions of employability. Also, the authors emphasize that career individuality is similar to paradigms of role individuality, occupational individuality, and institutional identity in that they all indicate how individuals introduce themselves in a specific work context. Personal adaptability donates both institutional performance and career achievement as it allows individuals to stay creative and attractive to employers in frequently changing work areas (Helens-Hart, 2015).

Tymon (2013) stated that in the face of recent deliberations about what should be done in the field of employability, most higher education organizations embrace the improvement of employability skills and qualifications within their prospectuses. Despite all of this, employers state that graduates are not ready to join the labour market, having a lack of some required and basic skills for fruitful employability future. Research studies on why this might be including the viewpoints of several shareholders: governmental sectors, managers, higher educational institutions and graduates themselves. Remarkably though, the opinions of undergraduates, the beneficiaries of this employability progress, are not well defined. Undeniably, learners' insights are very important as learning theory informs us that the enthusiasm and commitment of learners are vital requirements for effective consequences. So the main question here is whether undergraduate students are involved in the progress of employability skills (Tymon, 2013). Tymon (2013) investigated the opinions of over 400 business studies, marketing, and human resource management undergraduate students about the future of employability. Results recommended that there is only little arrangement between the opinions of students and other stakeholder communities. There are differentiations between first, second and final year students, which could explain an observed lack of involvement with employability-related progress. Some recommendations for developing involvement are given, together with ideas on what can, persuasively, be done within higher educational organizations.

From all of the above, employability as an international and national term it needs a certain set of standards and qualifications in order to achieve its goals (Winterton and Haworth, 2013). A number of countries in the Middle East region follows certain standards and it has its own laws and regulations in hiring people. In this regard, a very limited number of research studies had been conducted to investigate the level or females' readiness to join the field of work and the results mostly referred to the significant effect of the final grade point of them on getting hired in different sectors (El-Temtamy, Kathleen O'Neill and Midraj, 2016). Achieving a high rate of employment is a critical challenge for any government. Jordan is such any Middle East country seeks to achieve this rate but it is all the most difficult in Jordan as more than half of its people are teenagers. Young people unemployment speed has continued in the double figures for more than 100 years, and females' contribution in the work force is between the smallest in the world. For this reason, Jordan developed The National Employment Strategy (NES) to recognize the points by which the interference may has an impact for fixing unemployment irregularities and improving employability.

In order to make this unit effective, it should cover the following points:

- The first point came from the most effective factor affecting the employability rates in Jordan such as social norms, education, and working conditions, business owners
- The second point came from Jordan's restricted sources and the persistent description of the unemployment problem.
- lastly, boosting employability necessitates a reliable and continuous exertion for the medium and long term. (Hassan, 2011/2020)

2.4.1.3 The Historical Revolution of Employability:

In the last few decades, employability has had different definitions, along with a number of required skills and qualifications. This point became obvious through a number of research studies conducted all over the world; one study investigated the skills required for the employability of engineering graduates from employers' points of view. Around 180 employers from different sectors of engineering in Peninsular Malaysia participated in this research study. The researcher adapted a questionnaire from the SCANS model as an instrument of the study. The findings showed that the employers insisted that graduates must have a high level of academic skills to get employed, also the company size came as a non-significant variable affecting employability rates. Additionally, the author emphasized the importance of having a high level of technological information to get employed, according to the respondents' answers (Husain et al., 2010).

Another research study discovered that technical graduates are lacking in employability skills and qualifications. Also, employability competences are vital in outcome-based education; the basic objective of this research study was to build an Employability Skills Assessment Tool to assess learners and assist instructors to develop graduates proficient in the employability skills required by the workforce. The employability skills assessment tool was built by the Kepner-Tregoe (K-T) approach. Participants were 107 employers from five kinds of Malaysian manufacturing industries. The results revealed that employers strongly agreed on the importance for all seven of the employability skills: personal skills, intellectual skills, interpersonal qualities/ethics, resource qualities, system and amp; technology skills, basic skills and informational qualities. These skills were categorized and selected as components for the employability skills evaluation tool. The tool, once established, was tested and validated by industrial employers and lecturers in organizations. The agreement quantity was found to be between substantial agreement and almost absolute agreement (Rasul et al., 2012).

Despite endeavours to widen the terms of graduate employability, there remains a primary focus on rising industry-relevant employability requirements. The skills-based approach is, however, too limited and does not completely resolve the complication of graduate work-readiness. Another type of research paper argued for the re-conceptualizing of graduate employability by supporting pre-professional identity (PPI) development. The author stated that PPI associates to an identification of and link with the skills, qualities, behaviour, culture and thought of a student's proposed occupation. The 'communities of practice' model was extended upon to prove how PPI can be built during university years. Also, a learner makes sense of his/her future career through numerous memberships and differing types of engagement with various institutions within higher education's 'landscape of practice'. Moreover, the writer of this research study gave an example of institutions with professional connotations, learner societies, job facilities and employers, and a number of suggestions for stakeholders are deliberated (Jackson, 2016).

Another type of research study came as systematic reviews from literature provided the idea of generalizing the concept of employability as a major issue discussed in different research gates such as Springer Science + Business Media Dordrecht. As professionals, they are concerned with the employability issue in different nations: those facing unfavourable economic situations. The employability hypothesis characterizes a scientific obstacle in order to improve the understanding of the relationship between what matters to the career searchers 'and the prospects of the world of vocation. This research study presented a review of the term of employability. Three main views (educational and governmental, institutional, and personal) that are not exclusive can be recognized. The determined systematic review focuses on the importance of assuming a systemic integrative method and a broader clarification of employability. The researcher recommended a research plan to improve the theory and implementations of the term of employability for the future use (Truong, Laura and Shaw, 2018). From all of the above, it can be seen that the concept of employability comes in different areas of the literature with different meanings, obstacles, and requirements. Also, the employability requirements are different according to the graduates' academic and personal skills as classified by the researchers.

2.5 The evolution of predictive models in the fields of Education and Employability:

The field of education carries out an essential task in the improvement of a population's economy and development. It promises to refresh the regions by providing a consistent and qualified workforce to different societies. Also, higher education is the fundamental for adopting the talent, the basic factor for enriching the quality of public human capital, and the major technique to promote a population's competitive status. Consequently, research in improving higher education is an essential field and is fundamentally needed. To come up with a clearer point of view, different organizations are using cutting edge technologies such as artificial intelligence, deep neural network, and data mining on the huge amounts of data produced from the educational domain, like academic, social, demographic data of learners, and teachers data as well. The data created in the educational organizations can provide deep insight into the educational system's relationship with the employment system (Dutt, Ismail and Herawan, 2017). Furthermore, improving employability skills has been noted as a critical task within any management organization in the world's different sectors (Little, 2011). Employers' requirements and also the augmentation of learners' skill competences should be taken into consideration to convey future skills assessments (Zwane, Du Plessis and Slabbert, 2014). While the requirements of employers differ from the skills that are presented by graduates, a number of research studies with different techniques have been conducted in order to address this gap between what is supplied by the educational sector and what is demanded by the employment sectors (Zwane, Du Plessis and Slabbert, 2017).

Moreover, the relation between graduation rates and employability markets requirements has become one of the most critical issues in our societies (Tuma and Pratt, 1982). From this perspective many research studies are conducted in order to minimize the gap between Computer Science (CS) and Information Technology (IT) graduating students' numbers and the desire to get employment (Marchante et al., 2003). The following key ideas are stated by different authors in the literature about the most effective factors that would affect their prospects for employment:

- A group of research studies results confirmed three effective factors that had a direct influence on IT employability, which consist of IT core, IT_professional, and gender (Piad, 2018). Another point of view which has become clearer is that national employers consider the absence of essential soft skills to be hindering the graduates from national public colleges from securing employment (Williams, 2015).
- In addition, some research studies stated that the policy of the countries focused on places of employment, with the idea that the economic benefits of people and the reasonable improvement of countries are approached to emphasize the knowledge, skills and initiatives of the workforce. These

points together with degree-level credentials come to play an essential role in handling the ‘knowledge-driven’ economy of the future (Brown, Hesketh and Williams, 2013).

- Another group of researchers combined both of the market demands (employability) and the university graduates (supply) to improve the future of both sides (Aamodt and Havnes, 2008).

A further group of research studies found that the most effective factors affecting the term of self-employability is its consideration as the research area in this domain because it can provide researchers with some arguments and recommendations on how to narrow the gap between what is supplied and what is demanded (Branine and Avramenko, 2015). From this view point, the current research study will provide a real statistical analysis of CS and IT graduating students who joined the labour market, did not join, and are in an undetermined situation. After that a prediction model will be formed in order to classify the newly registered students in the same majors in all UAE universities which will be coming as a compilation of a number of research study approaches, such as (Piironen and Vehtari, 2017).

2.6 The historical relationship between Education, prediction models, and the evolution of Data Mining:

Nowadays graduates’ employability is a main issue for the organizations proposing higher education and a technique for early expectation of employability of the graduates is usually necessary to take an appropriate decision (Bridgstock, 2009).

A number of research studies apply several classification procedures of data mining, such as Bayesian methods, Multilayer Perceptron’s and Sequential Minimal Optimization (SMO), ensemble techniques and decision trees, to predict the employability of Master of Computer Applications (MCA) learners and find the algorithm which is most suitable for this issue (Mishra, Kumar and Gupta, 2016a).

In the same vein, another study assured the importance of data mining techniques in the educational field related to the history of graduates’ employment (Berland, Baker and Blikstein, 2014). Further, data mining has been implemented in diverse fields because of its capacity to promptly study massive volumes of data. The cited study in this regard aimed to construct the Graduates Employment Model with a classification task in data mining, and to contrast some data-mining techniques such as Bayesian method and the Tree method. The Bayesian technique contains five algorithms, such as AODE, BayesNet, HNB, Naive Bayes, WAODE. The Tree technique contains five algorithms, such as BFTree, NBTree, REPTree, ID3, C4.5. The study research handles a classification task in WEKA, and the authors compared the findings of each algorithm, as different classification models were produced. In order to authenticate the produced model, the experiments were constructed with authentic data gathered from graduate profiles at the Maejo University in Thailand. The intended model is used for expecting whether a graduate was employed, unemployed, or in an unidentified case (Pääkkönen et al., 2013).

Moreover, a number of studies have focused on the required qualifications for improving the future employability (Thijssen, Van Der Heijden and Rocco, 2008). With the current massive focus on academic principles, features and graduate employability consequences, Australian Higher Education organizations have an enhanced need to improve and develop feedback instruments to help and support graduate employability results. Another paper states the improvement of the Graduate Employability Indicators (GEI), a set of surveys for graduates, employers and instructors, of the significance of 14 employment competences for graduate place of work accomplishment and their determination by prospective graduates until five years out. These surveys were assigned through an ALTC permit, Building course team ability for graduate employability, a collaborative project between Curtin University, RMIT University, University of Southern Queensland and Victoria University. The paper sets out the relationships and differentiations between the GEI and other pointers; such as the Australian Graduate Pathways Survey (GPS), the Australasian Survey of Student Engagement (AUSSE) and the National Survey of Student Engagement (NSSE), revealing its conceivable implementation in local and international benchmarking actions. Summary visual data on the viewpoints of graduates from one of the preliminary surveys is also constructed to determine the kind of data that can be collected from the surveys (Oliver and Jorre de St Jorre, 2018).Also, in Educational Data Mining (EDM) there is too much imprecise input information, ambiguity or vagueness in input data, thus a lot of problems can occur during the classification process (Hernández-Blanco et al., 2019). Fuzzy logic is a mathematical model proposed by Lotfi Zadeh in 1965 to solve this problem observed with conventional computer logic while manipulating the imprecise and vague data (Zadeh and Aliev, 2018). Fuzzy logic is an approach to computing based on degrees of truth (degrees of membership of specific class) rather than crisp logic, true or false, on which the modern computer is based (Anderson et al., 2009).

Fuzzy logic is a computational paradigm based on the manner of human thinking; it deals with the problems in the same manner as the human brain works, where it takes imprecise input then produce the precise (e.g. grade is excellent, income is high).

Fuzzy logic can make development and implementation much simpler, and the techniques can give higher accuracy than other conventional techniques (Lingala et al., 2014).

Neuro-fuzzy algorithms combine neural networks with fuzzy logic where the fuzzy logic system builds according to the structure of the neural network.

There are two neuro-fuzzy systems:

1. Mamdani approach. The main characteristics of this approach are:
    a. There is output membership function in this technique
    b. The output of this approach is crisp value which is produced through the defuzzification process
    c. This approach can take multiple inputs then produce single or multiple outputs
    d. High interpretability
    e. The design of the system in this approach is not inelastic
2. Sugeno's approach: the main characteristics are:
    a. There is no output membership function in this technique
    b. No need for defuzzification because the output of the last layer is crisp value
    c. It takes multiple inputs then produce single output and it cannot produce multiple outputs
    d. There is a problem with interpretability
    e. The system design can be elastic.

The two techniques, Neural networks and fuzzy logic could be integrated by two different ways:

1- Neuro-Fuzzy System (NFS)

A fuzzy logic is represented using the structure of the neural networks (NN) and trained using either a backpropagation (BP) algorithm or genetic algorithm (GA). The purpose is to enhance the performance of the fuzzy reasoning tool by representing the fuzzy system using NN structure and the network is trained using BPGA.

2- Fuzzy Neural Network (FNN)

The neurons of the neural networks are built using fuzzy set theory ('Proceedings of the International Conference on Soft Computing for Problem Solving, SocProS 2011', 2012); this method actually can be developed using three different ways:

1- Fuzzy input with real weight

2- Real input with fuzzy weight

3- Fuzzy input and fuzzy weight.

The fuzzy neural network has not become popular; but the neuro fuzzy system is very popular and has a lot of real-life application.

The current research study uses the neuro-fuzzy algorithm called Adaptive Network-Based Fuzzy Inference System (ANFIS), which uses Sugeno's approach in the fuzzy system. Additionally, an interesting topic presented by researchers in higher education is to examine the importance of data mining in studying data in order to improve the quality of provision and address the demands of their graduates. Therefore, educational data mining appears as one of the most appropriate instruments to investigate academic data to classify patterns and assist in decision making influencing the educational field (Han et al., 2007). This cited article predicts the employability of IT graduates, applying nine variables. First, a number of classification algorithms in data mining were examined producing logistic regression with accuracy of 78.4 is applied. According to logistic regression analysis, three academic variables clearly have an effect: IT_core, IT_professional and gender are illustrated as meaningful predictors for employability. The data was gathered from the five-year profiles of 515 graduates randomly selected from the employment office tracer study (Piad et al., 2016).

Adding to all of the above, it is essential to take a look at the other wing of employability which is the designers and constructors of the careers (Larson and Lockee, 2009). Many research studies all over the world had been conducted in this area; one study intended to emphasize improving a career proficiency model by exploring the relationships between career proficiency and career success from a career improvement viewpoint. The authors assured that career competencies explained in the cited study are an essential orientation for the designing of basic or general education programs in the hospitality area. Hospitality courses can propose a "hospitality career and employability" program that introduces courses such as career identification, career designing, self-management, job-seeking and success approaches, problem solving skills, ethics and security in the place of the work, manner, collaboration, and coordination and networking skills. The study implemented a questionnaire survey to gather data from a group of 277 participants at 36 international places and applied the AMOS statistical software group to do structural equation modelling (SEM) for analysis. The findings of this study explained that career competency model is a multifaceted concept including four competency scopes that affect the career fulfilment of some department employees in international tourist hotels. Particularly, the competencies related to "career modification and management" competency measurements were the most effective competencies for career success. The career competencies in this study can serve as an orientation for people designing basic or general education programs in the hospitality field (Wang and Tsai, 2014).

2.7 Data Mining Techniques Used for Employability:

This section delivers a detailed review of the literature regarding employability problem and how data mining techniques are widely implemented to support the study claim. Thus, the main sections introduced in the literature review highlight and discuss the most effective factors affect employment future.

Some researchers focus on data mining methods and algorithms to predict graduates’ employability; others study the attributes that impact employability. The following literature studied attributes of employability.

Al-Janabi (2010) proposed an approach depending on features (knowledge areas) gained from the logged data of employment and university graduates. He presented a model for analysing data of the IT graduates according to the employability knowledge areas in order to predict feedback recommendations to enhance the IT programs’ teaching and learning resources and processes towards the improvement of the programs’ learning outcomes. Furthermore, Artificial neural networks came from attempts to simulate the biological neural system, where the human brain consists of neurons linked together via axons. Each neuron connects to the axon of other neurons via dendrites. The point of convergence between axon and dendrites is called a synapse. Scientists discovered the learning process of the brain is carried out by changes to the strength of synaptic connection between neurons. Artificial neural network (ANN) corresponds to a set of nodes, linked as in a brain. ANN process one record at a time; it classifies the record then it compares the classification process with the actual class of the record. If an error exists in classification, the error is fed back to the network to modify the second iteration and so on. ANN is a popular technique, usually used when high accuracy and superior learning capability is desired even if the available training data is not large. There are various ANN methods but the most popular one is the Multilayer Perceptron Backpropagation Network (MLPBPN) algorithm, which is used in most ANN research studies (Albawi, Mohammed and Al-Zawi, 2018). Bezuidenhout (2011) developed the employability attributes framework (EAF). This framework illustrates a group of eight employability attributes that are considered as essential for boosting the probability of securing and sustaining employment opportunities (Bezuidenhout and Jeppesen, 2011). The EAF has the following eight measures: career self-management, cultural competence, career resilience, proactivity, entrepreneurial orientation, sociability, self-efficacy, and emotional literacy. Sriram, Srinivas and Thammi (2014) presented a model to predict the attributes which play the main role in the employability of students. They used Maximally Specific Hypothesis in order to reduce the representation of rules. These hypotheses can be used to identify the key attributes needed for employability among graduates.

Isljamovic and Suknovic (2014) used different artificial neural network algorithms in order to find the best suited technique for prediction of students' performance. Also, they studied which factors had a crucial influence on overall performance. The authors developed a method of multilayer neural network that has the ability to predict the success of students at the end of their studies. The main idea of their research was to make early prediction of students' performance.

Thakar and Mehta (2017) studied the role of secondary attributes to enhance the prediction accuracy of students' employability using data mining. They proved that prediction accuracy for students' employability can be enhanced with the applying of secondary attributes such as personal, social, psychological and other environmental variables in the dataset.

The following are related to work that focused on the data mining methods or algorithms to predict graduate employability.

Piad (2018) proposed a technique to predict the employability of IT graduates. His study defines the influential attributes for supervised learning using data mining methods. He conducted a comparison between several classification data mining algorithms. These algorithms are Naive Bayes, J48, Simple Cart, Logistic Regression and Chaid Algorithms. The author proved that the Logistic Regression achieved the highest accuracy, and he found that three possible predictors with a direct effect on IT employability are the IT_core Subjects, IT_professional subjects and gender (Piad, 2018).

Jantawan et al. (2013) used real data of graduate students of Maejo University in Thailand over three academic years. They conducted several experiments using algorithms of Bayesian Network and Decision Tree to predict whether a graduate has been employed, remains unemployed, or is in an undetermined situation after graduation (Jantawan and Tsai, 2013).

Sapaat et al. (2011) built the graduates employability model using a classification method in data mining. To perform the classification, they used extracted data from web-based survey system from the Ministry of

Higher Education, Malaysia (MOHE) for the year 2009. Bayes algorithms were used to achieve classification. In addition, they compared the performance of Bayes algorithms against a number of tree-based algorithms. The comparison shows the superiority of the Decision Tree classification model over Bayes Network Classification Models (Sapaat et al., 2011).

Mishra et al. (2016) applied several classifiers to predict the employability of students and build an employability model based on proper classifiers. The authors used different classification methods of data mining such as Bayesian methods, Multilayer Perceptron's and Sequential Minimal Optimization (SMO), Ensemble Methods and Decision Trees. They conducted a comparison between the classifiers to find the best classifier. A comparative study shows that J48 (a pruned C4. 5 decision tree) is most suitable for employability (Mishra, Kumar and Gupta, 2016b).

Khadilkar and Joshi (2017) proposed a predictive technique on employability using machine learning. In screening the resumes, they used text mining and appropriate weighting. They used several classifiers such as decision tree, K-NN, Naïve based approach, and Random Forest for employability prediction. Naïve based has the highest accuracy for the prediction of employability.

Rahman, Tan and Lim (2017) used supervised and unsupervised learning in data mining for employment prediction of fresh graduate students. These techniques were applied in features selection and determined the best model that can be used to predict the employment status of fresh graduates, either employed or unemployed. The algorithms in supervised and unsupervised learning, K-Nearest Neighbor, Naive Bayes, Decision Tree, Neural Network, Logistic Regression and Support Vector Machines, were compared to find which one achieved the best accuracy. They proved that K-Nearest Neighbor achieved the highest accuracy (Rahman, Tan and Lim, 2017).

Othman, Shan, Yusoff and Kee (2018) proposed a model that uses data mining techniques to discover the most important features that affect graduates' employability. They collected seven years of data (from 2011 to 2017) through Malaysia's Ministry of Education tracer study. The authors applied a set classification algorithms (three), Decision Tree, Support Vector Machines, and Artificial Neural Networks to develop the classification model, then compared them to reach the perfect performance. According to the authors, the decision tree J48 algorithm achieved higher accuracy compared to other algorithms, with a classification accuracy of 66.0651%, and it rose to 66.1824% after the process of parameter tuning. In their work, they discovered seven variables affecting graduate employability: age, faculty, the field of study, co-curriculum, marital status, industrial internship and English skill. In addition to these variables, attribute age, industrial internship and faculty hold the information that influences the employability status (Othman et al., 2018).

Kumar and Babu (2019) applied supervised Machine Learning algorithms to analyse data collected from educational institutions to predict the employability of current students (not graduates). They collected the data from 500 students of several Engineering colleges in Hyderabad and used Supervised Machine Learning algorithms such as Decision Tree, Support Vector Machine, Gaussian Naïve Bayes and K-Nearest Neighbor to build an employability prediction model of students and to determine the factors affecting their employability. They found that Decision Tree and Support Vector Machine outperformed the Gaussian Naïve Bayes and K-Nearest Neighbor, by predicting the employability of the students with 98% accuracy. Also, they found that factors such as communication skills, aptitude and reasoning skills, mentor, family income status, and the quality of teaching in college affect the employability of students (Kumar and Babu, 2019).

2.8 Chapter Summary:

Data mining techniques and algorithms have been implemented significantly in the education field to predict helpful information. This information can be used by teachers to improve the performance of students, also to enhance learning materials. The abundance of data provided by different resources, such as LMSs, admission systems, and social media can be used in EDM techniques. Educational data mining can be categorized into two parts: the first part is statistics and visualization. The second part is web mining, which includes clustering, classification, outlier detection, association rule mining, and sequential pattern mining and text mining. Recently, employability prediction has attracted the attention of higher education institutions, due to the valuable information obtained that can assist in strategic planning. EDM is considered an essential technique in the employability prediction process. Studies of employability can be divided into two parts; the first part studied the techniques that have been used in performing the employability prediction task. The second part studied the impact of the attributes on employability prediction. The neuro-fuzzy approach which combines both neural networks and fuzzy logic approaches has been applied widely in performing classification tasks.

3 Chapter 3. THEORETICAL FRAMEWORK

Supervised learning and unsupervised learning are considered as Machine Learning types (Ayodele, 2010). The area of research which is emphasized in the current research study is the supervised learning area. Supervised machine learning happens when a model is trained on existing data that is correctly labelled (Kotsiantis, 2007). Regression and classification are two techniques applied when creating machine learning algorithms. Both regression machine learning algorithms and classification machine learning algorithms are categorized under the area of supervised machine learning. The main difference between classification and regression is that classification predicts a discrete value, while regression predicts a continuous value (Polamuri, 2014).

Classification is a technique that is used to assign items into categories or classes. This research tries to recognize data mining classification methods and apply them to predict the employability of IT and CS graduates. Machine Learning is a subsection of artificial intelligence which concentrates mainly on machines learning from their experience and making predictions due to this experience (Tiwari, 2017). It empowers the computers (machines) to make data-driven decisions rather than being explicitly programmed for achieving a certain task. These programs or algorithms are designed in a way that they learn and develop over time when they are exposed to new data.

Regression is the process of reaching a model that predicts a continuous value due to its input variables (Huang, Cabral and Torre, 2016). In regression problems, the task is to mathematically estimate a mapping function (f) from the input variables (y) to the output variables (x). On the other hand, classification is the technique of reaching a model that splits input data into multiple discrete values.

Consider the same dataset of all the students at a university. A classification task would be to use parameters, such as a student's weight, or major, to determine whether they fall into the "Above Average" or "Below Average" category. Note that there are only two discrete labels in which the data is classified. Consider a dataset that contains information about all the students in a university. An example of a regression task would be to predict the height of any student based on their gender, weight, major. We can do this because height is a continuous value and there is an infinite amount of possible values for a person's height.

As the current research study aims to predict the future of CS and IT graduates employability and categorize graduates into employed and unemployed which are discrete values, classification is used to reach the research goal. The Adaptive Network-based Fuzzy Inference System (ANFIS) approach is compared in the current study with a number of classifiers in order to handle the classification task. This chapter provides a theoretical framework of classification as a type of supervised machine learning. We aim to develop a binary classification model to categorize employability status as either the graduate student is employed, or not.

In this regard, two phases will be used to perform classification tasks; a learning phase and a classification phase. In the first phase, a classification method constructs the classifier by analysing training data generated from datasets and their linked class labels. The class label attribute is categorical. Due to the existence of the class label of each training data, this phase is also called supervised learning.

In the second phase, the model is applied for classification of unknown instances. In this phase the classification model performance is verified and evaluated as well. There is a two-step procedure:

- Learning step: for building the classification model's several algorithms in order to construct a classifier by training the model with the training set availability. After that, the model should be trained to predict the exact results (Tang, Alelyani and Liu, 2014).
- Classification step: the constructed model is applied to discover class labels and verification, then the model is tested on tested data and the classification rules' accuracy is approximated (Kesavaraj and Sukumaran, 2013).

3.1 Major types of supervised learning methods

Supervised learning is defined as the machine learning task of learning a function that draws an input to an output based on example input-output pairs. It implies a function from labelled training dataset contains a group of training instances. In supervised learning, each example is a pair containing an input object and a desired output value. A supervised learning algorithm studies the training dataset and creates an implied function, which can be used for mapping new instances. An optimal scenario allows the algorithm to correctly determine the class labels for unseen instances.

In order to solve any given problem of supervised learning, the following steps should be followed:

1. Determine the data type that is used as a training set.
2. Collect the training dataset.
3. Consider the input feature representation of the learned function.
4. Define the input feature representation of the learned function.
5. Design the model by running the algorithm on the training dataset.
6. Evaluate the model accuracy.

An extensive range of supervised learning algorithms are available, each with its strengths, accuracy, and weaknesses. There is no single learning algorithm that works best on all supervised learning problems based on no free lunch theorem (Wolpert, 1995). The following factors should be considered when selecting the supervised learning algorithm for certain problem;

1. Data heterogeneity; many algorithms are easier to apply such as Support Vector Machines, linear regression, logistic regression, neural networks, and nearest neighbor.
2. Data redundancy; some algorithms perform poorly such as linear regression, logistic regression, and distance based methods because of the numerical stability.
3. Presence of interactions and non-linearity; if there are complex interactions among features, algorithms such as decision trees and neural networks work better, because they are specially considered to find these interactions.

3.2 Classification methods:

Classification is a supervised machine learning type, in which the algorithm learns from the data input given to it. After that, this learning is used to classify new instances (C., 1998). In other words, the training dataset is applied to find better borderline conditions which can be applied to define each target class; once such borderline conditions are defined, the next task is to predict the target class. Binary classifiers perform with two classes only or possible outcomes (example: positive or negative sentiment; whether graduate will be employed or not; etc), and Multiclass classifiers work with multiple classes (e.g.: to which grade a student belongs, whether an image is a cat or rabbit or dog, etc). Multiclass supposes that each sample is assigned to one and only one label (Sokolova and Lapalme, 2009).

In this chapter we will discuss a set of classification methods, improving classification accuracy techniques, classifier evaluation techniques, and measure of attribute selection. In the first section, we will discuss a set of classification algorithms: DTC, MLP, SVM, Naïve Bayes, and ANFIS.

3.2.1 Decision Tree Classifiers

This classifier is represented such as tree structure (Safavian and Landgrebe, 1991), where the features represent a tree; the first node in the flowchart is known as a root node, inner nodes indicate a test on feature, leaf nodes at the bottom of the tree represent a class label. Figure 3.1. illustrates the term of employability, that is, it finds if a graduate is employed or not. inner attributes are represented by squares, and leaf attributes are represented by circles. There are kinds of decision tree methods generate only binary trees, on the other hand, other kinds can generate non-binary trees.

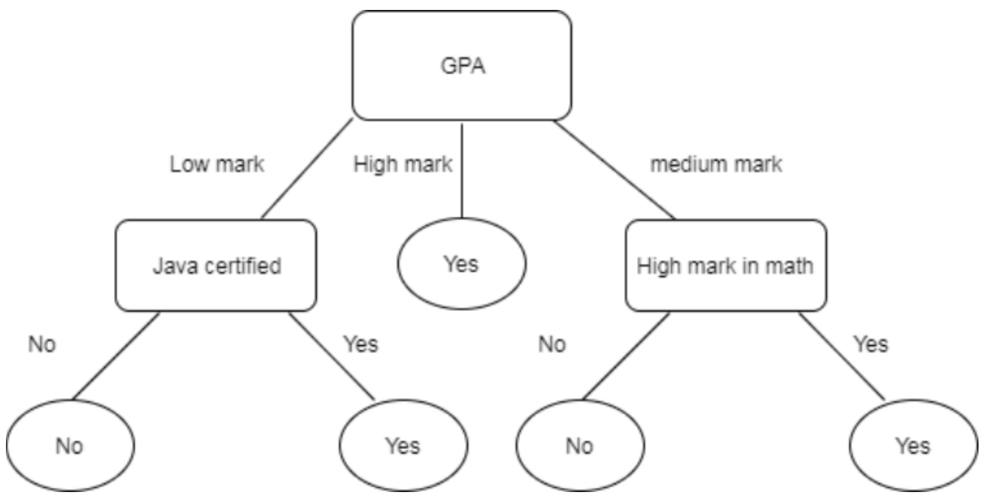

Figure 3. 1 illustration of decision tree that represents employability status, denoting if a graduate student is employed or not

Given a certain data, for which the related class label is not set, the features values of the data are validated versus the graph. A path is drawn from the top of the tree to a leaf node, this includes the class prediction for a given data. Converting Decision trees to classification rules can be done easily. The representation of decision tree classifiers does not need any domain information and for that reason it is good for information discovery. Tree structures can manage multidimensional data. Representing the attributes in tree shape makes it easy to be interpreted by humans. Also, and according to the literature, decision trees have achieved good accuracy. For the previous reasons, decision tree classifiers are popular and have been applied for classification in many application fields such as educational data mining applications.

Some of the decision tree algorithms are Hunt's Algorithm, ID3, C4.5, and CART. The first decision tree was ID3 (Amin, Indwiarti and Sibaroni, 2015); researchers then proposed C4.5 (a successor of ID3). ID3, C4.5, and CART follow a greedy method, in which decision trees are built in a top to bottom direction divide-and-conquer behavior. Decision tree algorithms begin with a training group of data and their linked class labels. The training set is repeatedly divided into smaller subsets as the tree is being constructed. Differences in decision tree algorithms include how the features are chosen in producing the tree and the techniques used for pruning (Bradford et al., 1998).

3.2.1.1 Tree Pruning:

After building the decision tree, many of the branches will reflect oddity in the training data due to missy data. Tree pruning methods solve this problem of overfitting the data. Pruning is a technique applied to deal with overfitting, that decreases the size of DTs by removing sections of the Tree that give little predictive or classification power. While overfitting is defined as the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably (Hawkins, 2004), an over-fitted model is a statistical model that contains more parameters than can be justified by the data (Rousseau and Mengersen, 2011). The essence of overfitting is to have unknowingly extracted some of the residual variation (the noise) as if that variation represented underlying model structure.

Such methods use statistical measures to clear the un-normal branches. An unpruned tree and a pruned version of it are represented in Figure 3.2. Pruned trees are smaller and, thus, easier to understand. They are usually faster at correctly classifying independent test data than unpruned trees.

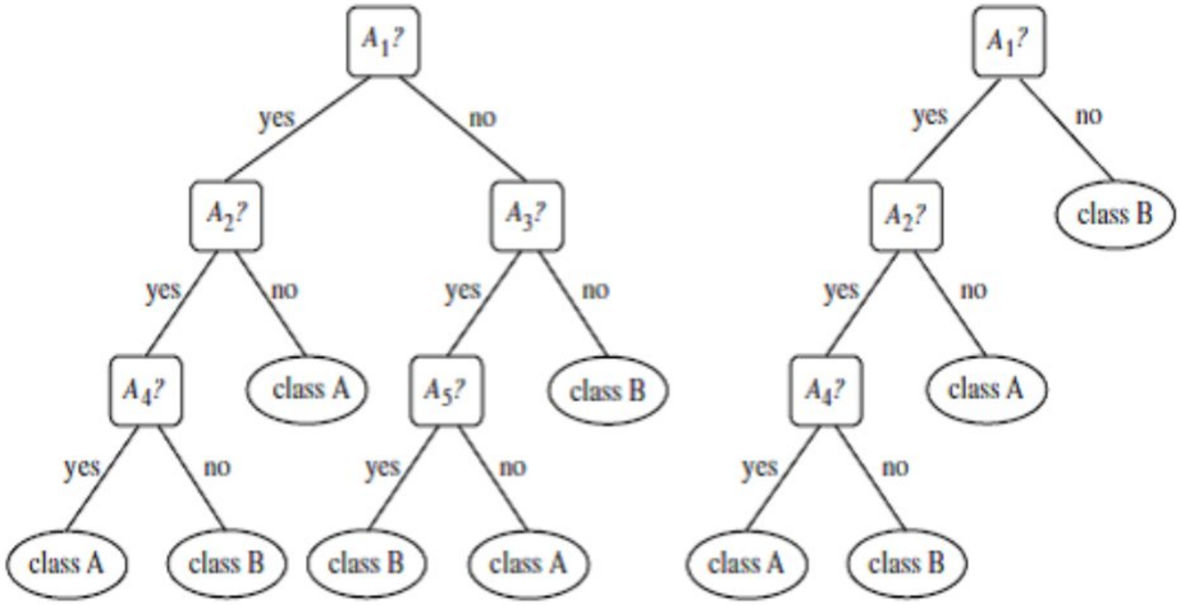

Figure 3. 2 An unpruned decision tree and a pruned version of it.

There are two methods to tree pruning: pre-pruning and post-pruning. In the pre-pruning method, a tree is "pruned" by limiting its construction in early stages, then the node is transformed to a leaf. The leaf may contain the most frequent class among the subset items or the prospect allocation of those items. The second method is post-pruning, which clears sub-trees from a "fully grown" tree. A sub-tree at a given node is pruned by removing its branches and replacing it with a leaf. The leaf is marked with the most frequent class among the sub-tree being replaced.

3.2.2 Bayesian Classifiers:

Bayesian classifiers are considered as statistical classifiers. They can discover class membership probabilities such as the probability that a given data belongs to a specific class. An example of a Bayesian classifier is naïve Bayesian classifier (Bielza and Larrañaga, 2014).

The main idea of Naive Bayesian classifiers is to suppose that the influence of an attribute value on a given class is not linked to values of the other attributes. This assumption is known as class conditional independence, because of that, we can call it "naïve". Let X be a data sample: class label is undefined. and H is a hypothesis that X belongs to class C. For the Classification problem, we need to define P(H|X), the probability that the hypothesis holds given the recorded data sample X. P(H), the initial probability e.g., X will be employed, regardless of GPA, English level, etc. P(X): probability that sample data is observed and P(X|H) (posteriori probability), the probability of observing the sample X, given that the hypothesis holds. Given training data X, posteriori probability of a hypothesis H, P(H|X), follows the Bayes theorem

3.1 $$P(H \mid \mathbf{X}) = \frac{P(\mathbf{X} \mid H)P(H)}{P(\mathbf{X})}$$

Predicts X belongs to C2 if the probability P (Ci |X) is the largest values among all the P(Ck|X) for all the k classes. A practical difficulty is that it needs initial information of many probabilities, at big computational cost.

The naïve Bayesian classifier performs as follows:

- D is a training set of items and their related class labels, and each data is described by an n-D attribute vector X = (x1, x2, …, xn).
- If we have m classes C1, C2, …, Cm.
- Classification is to find the largest posteriori, i.e., the maximal P (Ci |X).
- This can be represented as:

3.2 $$P(C_i \mid \mathbf{X}) = \frac{P(\mathbf{X} \mid C_i)P(C_i)}{P(\mathbf{X})}$$

- Because P(X) is fixed values for all classes, only P (Ck|X) = P(X |C I) P(C I) requires to be maximized.

- A simplified assumption: attributes are conditionally free:

3. 3 $$P(\mathbf{X} \mid C_i) = \prod_{k=1}^{n} P(x_k \mid C_i) = P(x_1 \mid C_i) \times P(x_2 \mid C_i) \times \ldots \times P(x_n \mid C_i)$$

To calculate $P(C_i|\boldsymbol{X})$, we consider the following:

- If Ak is categorical, P(Xk|Ci) is the numbers of items in Ci having value Xk for Ak divided by |Ci, D| (numbers of items of Ci in D).

- If Ak is continuous-valued, P(Xk|Ci) is commonly calculated according to Gaussian distribution with a mean μ and standard deviation σ, defined as:

3.4 $$g(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Also

3.5 $$P(\mathbf{X}|C_i) = g(x_k, \mu_{C_i}, \sigma_{C_i})$$

- A null probability value stops the influences of all the other probabilities (on Ci) engaged in the product. Naïve Bayesian prediction needs each conditional probability to be greater than zero. Else, the predicted probabilities will be zero. A zero probability stops the influences of all the other probabilities (on Ci) engaged in the product.

- Advantages of Naïve Bayesian Classifier (Dejaeger, Verbraken and Baesens, 2013) are:

- Easy to implement.

- Accepted outcomes achieved in most of the scenarios.

- Disadvantages of Naïve Bayesian Classifier (Xu, 2018):

- Assumption: class conditional independence, because of that less accuracy.

- Dependencies available among variables.

To handle the dependencies problem, Bayesian Belief Networks is used.

In our research we will apply Naïve Bayesian Classifier and compare its accuracy with our approach of neuro-fuzzy.

3.2.3 Random Forests:

This method acts as bagging, with a small difference. When determining the split point, bagged decision trees have the complete elimination of features. Even though the bootstrapped datasets may be a bit different, the data is extremely going to break off at the same features throughout each model. On the contrary, random forest models decide where to split according to a random selection of features. Instead of splitting at similar features at each node throughout, random forest models apply a level of differentiation because each tree will split based on different features. This level of differentiation gives a better ensemble to aggregate over, ergo creating a more accurate predictor (Pavlov, 2019).

Another type of random forest is Forest-RC, which applies random linear combinations of the input features. Rather than randomly selecting a subset of the features, it generates new features that are a linear combination of the present features (Wager and Athey, 2018).

The random forest method performs well for the class imbalance problem on two-class jobs. Threshold-moving and ensemble methods were empirically observed to outperform oversampling and under sampling. The disadvantage of the random forest method is that multiclass tasks are much harder when a class imbalance problem occurs, where oversampling and threshold moving are less effective (Biau and Scornet, 2016).

3.2.4 Bayesian Belief Networks:

These algorithms indicate joint restrictive likelihood circulations. They permit class contingent independencies to be characterized between subsets of factors. They give a graphical representation of causative relationships, on which training can be conducted. Trained Bayesian belief networks can be used for building classifiers. Bayesian belief networks are also called belief networks, Bayesian networks, and probabilistic networks (Cheng and Greiner, 2001).

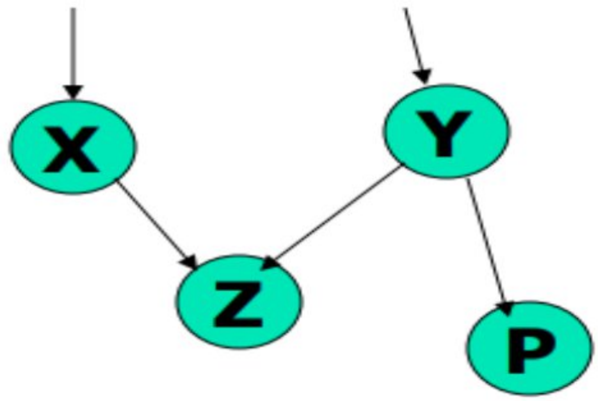

Figure 3. 3 A graphical model of causal relationships

Figure 3.3 is a graphical representation of relationships between attributes. It forms connection among the factors, which provides a specification of common probability distribution. Nodes represent random factors and links represent dependency. X and Y are the parents of Z, and Y is the parent of P. There are no links between Z and P and there are no loops.

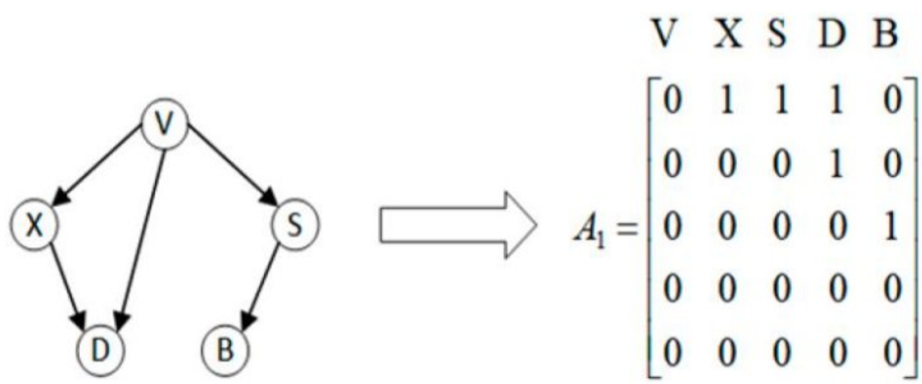

Figure 3. 4 Directed acyclic tree and conditional probability table (CPT)

A belief network can be represented by two forms; a directed acyclic graph and a group of conditional probability tables (Figure 3.4). Each node in the directed acyclic graph describes a random factor. The factors may be discrete-valued or continuous-valued. They may be linked to actual attributes given in the data or to hidden factors. Each arc represents a probabilistic link. If an arc is drawn from a node z to a node z, then y is a parent or direct grandfather of z, and z is a successor of y. Each variable is conditionally independent of its non-descendants in the graph, given its parents.

We can calculate the probability value of a certain combination of values as follows:

$$P(x_1,\ldots,x_n) = \prod_{i=1}^{n} P(x_i \mid Parents(Y_i)) \tag{3.6}$$

The included node in the tree can be chosen as an output node, forming a class label factor. There may be more than one output node. Several algorithms for deduction and training can be used in the tree. As opposed to restoring one class name, the classification procedure can restore a likelihood appropriation that gives the likelihood of each class. Coming up next is Bayes classifier algorithm.

```
Naive_Bayes_Classifier
START
The training datasets were split into two sets based on the outcomes of yes or no,
put sets into result_Map;
Traversal result_Map {
If the key is “yes” {
Assign the proportion of the sample with the result of “yes” in the total
training sample to yesCurrent;
Traversal testList{
yes_Current=yes_Current x The probability of this factor
value in the test_List that displays in the (yes) sample;
}
If the k is “no” {
Assign the proportion of the sample with the result of "no" in the total
training sample to noCurrent;
Traversal testList{
noCurrent=noCurrent x The probability of this factor value
in the test_List that appears in the “no” sample;
}
If yesCurrent>noCurrent
```

```
return "yes";
Else return"no";
END
```

The simplicity with which numerous applications can be diminished to Bayesian system deduction is beneficial in that it controls the need to design specific models for each such application.

- Training of Bayesian Belief Networks:

This stage includes several cases that can be handled. The tree structure may be built by human experts or inferred from the data. The tree nodes may be visible or hidden in all or some of the training items. The hidden data scenario is also called as missing values or incomplete data. Many methods are applied for training the tree graph from the visible training items. Human experts commonly understand the straightforward conditional connections that hold in the domain under analysis, which assists in building the tree (Landuyt et al., 2013).

If the tree graph is built and the factors are visible, then training the network is explicit. It contains calculating the CPT entries, as is comparably performed when computing the probabilities involved in naïve Bayesian classification. When the tree structure is built and some of the factors are invisible, there are several techniques to select from for learning the belief network (Landuyt et al., 2013).

There are several scenarios for the Learning Bayesian Networks as follows:

- If the tree graph and all factors are visible: train only the CPTs.
- If the tree structure is defined, some visible factors: gradient descent method, analogous to neural network learning.
- Tree structure undefined, all factors are visible: looking for through the tree graph space to rebuild tree structure.
- Undefined tree structure, all invisible factors: No proper technique proposed for this goal (Han, Y., & Lee et al., 2016).

3.2.5 Classification by Backpropagation:

Backpropagation is a neural system learning method. The neural systems field was initially ignited by analysts and neurobiologists who tried to create and test computational analogs of neurons. a neural network is a group of linked input and output units in which every association has a weight related with it. Through the learning stage, the network learns by adapting the weights to discover the right class name of the input items. Neural system learning is additionally alluded to as connectionist learning because of the associations between units (Miranda and Suñé, 2020).

Neural systems include long preparing occasions and are accordingly increasingly reasonable for applications where this is achievable. These systems require various parameters that are best defined empirically like the tree structure. Neural systems have been censured for their bad interpretability. For instance, it is hard for humans to analyse the representative meaning behind the learned weights and of invisible variables in the tree. These attributes made neural networks less preferable for data mining.

A lot of neural networks algorithms have been proposed. The most well-known neural network algorithm is backpropagation, which has spread in the 1980s. We will discuss these techniques in the following sections.

3.2.6 A Multilayer Feed-Forward Neural Network:

This method implements training on a multilayer feed-forward neural network. It repeatedly trains a group of weights for prediction of the class name. A multilayer feed-forward neural system includes an input layer, one or more invisible layers, and an output layer. A demonstration of a multilayer feed-forward network is illustrated in (Figure 3.5).

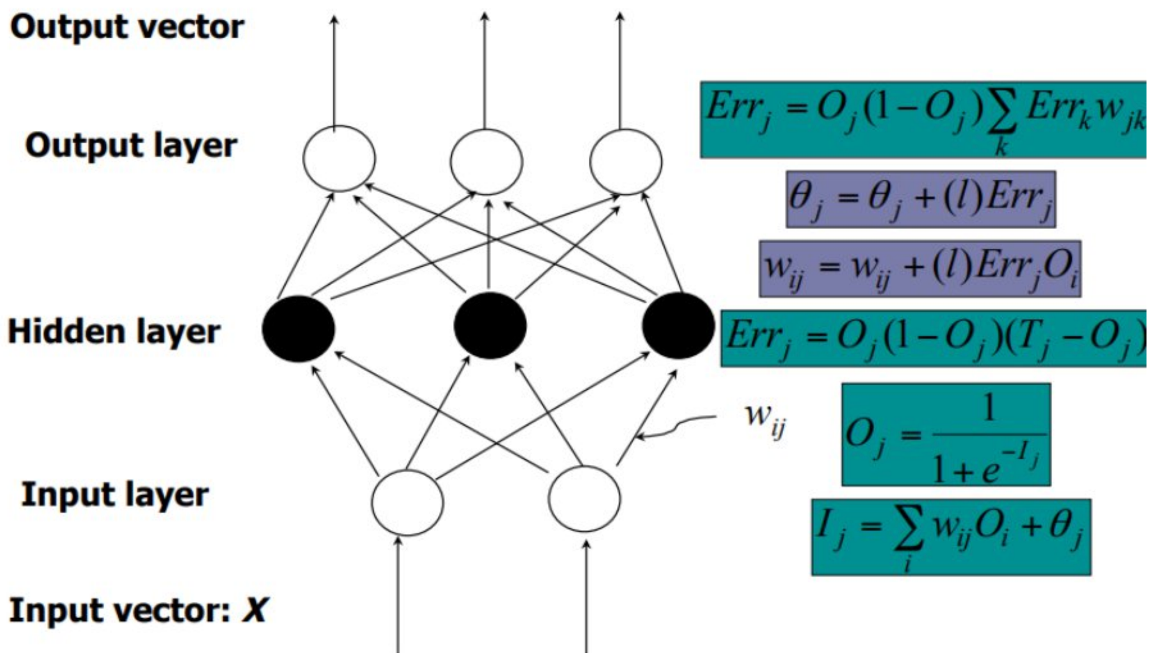

Figure 3. 5 Multilayer feed-forward neural network

The inputs in the above network agree with the features measured for each learning data. The inputs are entered at same time into the units to generate the input layer. These input values go through the input layer and are then assigned a weight to an invisible layer.

The outputs of the invisible layer units can be input to another hidden layer, and so on. The number of invisible layers is randomly; commonly only one layer is employed. The weighted outputs of the final invisible layer are input to units generating the output layer, which throws the network's prediction for given items. The units in the input layer are named input units. The units in the invisible layers and output layer are known as neurons, because of their representation in a biological way. The network of multilayer neural illustrated in (Figure 3.5) includes two layers of output units. Also, a network consisting of two invisible layers is known as a three-layer neural network.

- Network Topology Definition:

Before starting the training, we should determine NN structure by defining the number of units in the input layer, the number of hidden layers if greater than one, the number of units in each invisible layer, and the number of units in the output layer.

Normalizing the values of input for each factor evaluated in the training items will assist in making the training stage fast. Input values are normalized so as to be located between 0.0 and 1.0. Discrete-valued factors may be represented like there is one input unit per domain value.

Neural networks can be applied for building classifiers and numeric prediction. For building classifiers, one output unit may be applied to describe two classes. If there are more than two classes, then one output unit per class is used.

Network topology is based on try and catch errors technique and may influence the efficiency of trained models results. The initial values of the weights may also influence the performance results. Once a network has been trained and its efficiency is not performed well, it usually repeats the training stages with a different network design or another group of initial weights. Cross-validation techniques for efficiency evaluation can be applied to assist determine when a good network has been discovered. Several methods have been introduced that search for a proper network topology. These apply a hill-climbing method that begins with an initial design that is selectively amended.

3.2.7 Backpropagation Method:

This trains by repeatedly stages a data set of training items, matching the model's prediction for each data with the real defined target value. The target value may be the defined class name of the learning data or a continuous value. For each training data, the assigned weights are adapted so as to reduce the mean-squared error between the model's prediction and the real target value. These adaptations are generated in the backwards way by each invisible layer at the bottom to the root invisible layer at the top.

The processes of Backpropagation include the following steps:

- Assign random weights values for each node.
- Using activation function to increase the inputs forward.
- Repeat generating the errors by adapting the weights.
- Ending the condition.

To calculate the net input to the unit, each input linked to the unit is multiplied by its matching weight, and this is summed. Given a unit, j in an invisible layer, the net input, Ij, to unit j is (3.7):

3.7 $$I_j = \sum_i w_{ij} O_i + \theta_j,$$

wij represents the weight of the connection from unit i in the former layer to unit j; Oi is the output of unit i from the former layer; and θj is the bias of the unit. The bias represents a threshold in that it serves to change the value of the unit activity.

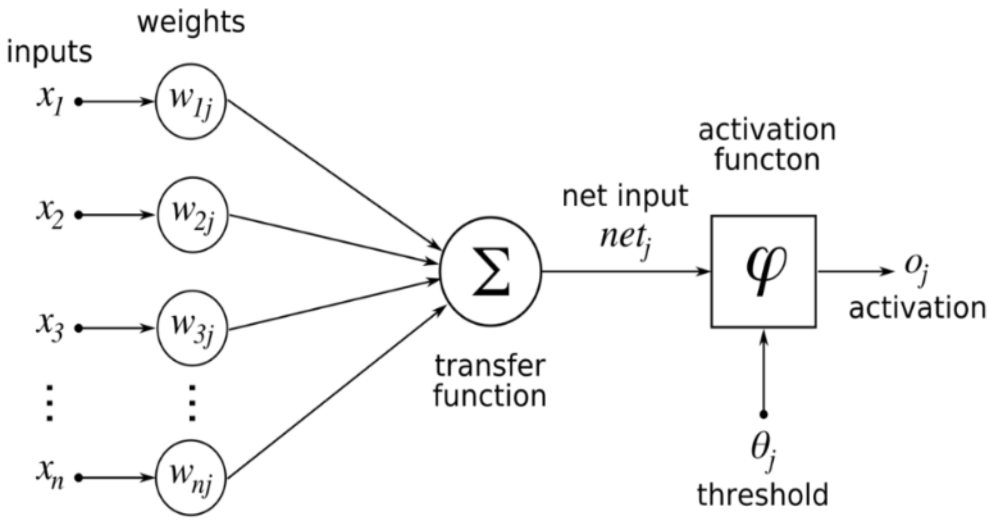

Figure 3. 6 Diagram of an artificial neural network to illustrate the layers used here

Invisible layers units assign their net input and then apply an activation function to it, as shown in Figure 3.6. The function encodes the activation of the neuron formed by the unit. The logistic, or sigmoid, function is applied. Given the net input Ij to unit j, then Oj, the output of unit j, is computed as (3.8):

3.8 $$O_j = \frac{1}{1+e^{-I_j}}.$$

The above procedure is also known as a squashing function, because it represents a huge input domain into a smaller range of 0 to 1. The logistic function is nonlinear and differentiable, permitting the backpropagation method to handle classification issues that are inseparable.

There are many backpropagation techniques that have been introduced for building classifier in neural systems. These methods include the dynamic adaptation of the network design and of the learning rate or other parameters, or the application of other error functions.

For the classification in our case, one output unit may be used to represent two classes: employed or not employed (where the value 1 represents the employed class, and the value 0 represents not employed). If there are more than two classes, then one output unit per class is used. The best numbers of hidden layers that perform good accuracy are not clear.

3.2.8 Support Vector Machines:

Known as SVM, it was proposed by Vapnik and et al, 1992. SVM is a supervised machine learning algorithm that can be used to build prediction classifiers of both linear data which is organized in a linear order and nonlinear data which is organized hierarchically. SVM uses a nonlinear mapping to transform the original training data into a higher dimension. After that, it searches for the linear best splitting hyperplane. Hyperplane is a line that linearly separates and classifies a set of data (Panja and Pal, 2018). SVM finds this hyperplane using support vectors and errors. The main idea of SVM is to find a hyperplane that best separates a dataset into two classes, as shown in (figure 3.7).

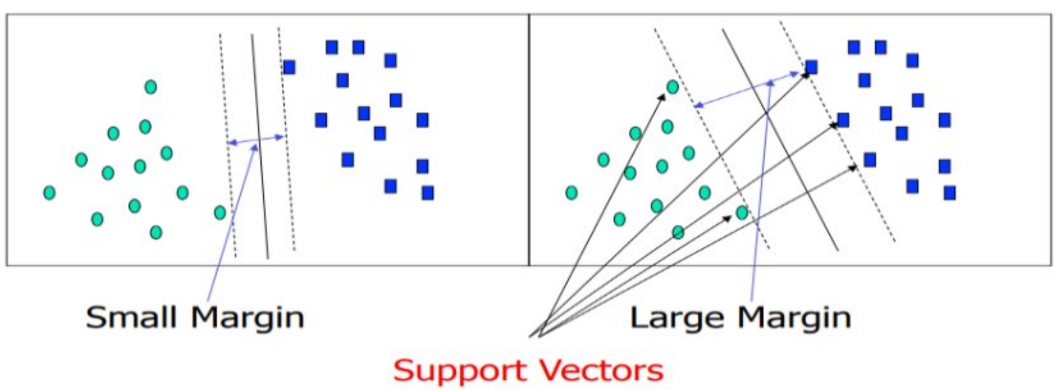

Figure 3. 7 Hyperplane separates dataset into two classes in linear case

Support vectors represent the data points closest to the hyperplane, the points of a data set that, if eliminated, would change the position of the dividing hyperplane. Because of this, they can be considered the critical elements of a data set. The further from the hyperplane our data points are located, the more confident we are that they have been classified correctly.

Bigger margins are indicating less errors in classifying data points.

- Linearly and not linearly Separable data:

In the linearly separable data, the hyperplane between data points is clear and easy to identify. For example, if we assign data D be (X1, Y1), …, (X|D|, Y|D|), where Xi is the set of training items associated with the class labels Yi , corresponding to the classes employed = yes and employed = no. There are many hyperplanes separating the two classes but we want to figure out the best one that reduces classification errors on hidden data. The concept of SVM is to search for the hyperplane with the maximum margin, known as maximum marginal hyperplane (MMH). A separating hyperplane can be written as:

3.9 $\mathbf{W} \bullet \mathbf{X} + b = 0$

where $\mathbf{W} = \{w_1, w_2, \ldots, w_n\}$ is a weight vector and b bias.

If the training items are 2-D we can rewrite Eq. (3.10) as:

3.10 $w_0 + w_1 x_1 + w_2 x_2 = 0$

The hyperplane defining the sides of the margin are:

H1: $w0 + w1\ x1 + w2\ x2 \geq 1$ for $yi = +1$, and

H2: $w0 + w1\ x1 + w2\ x2 \leq -1$ for $yi = -1$

Any training items that are located on hyperplanes H1 or H2 are considered as support vectors, because they are near the MMH. The support vectors are the most critical point to predict and provide the most information to build a prediction classifier.

In the case that the data are not linearly separable, as in (Figure 3.8), no straight line can be drawn that would split the classes. In this issue it is not allowed to make a straight line to separate the classes. Instead, the decision boundary is nonlinear.

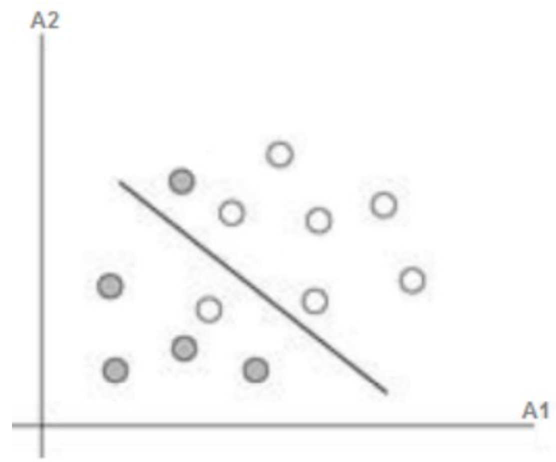

Figure 3. 8 A simple 2-D case showing linearly inseparable data.

The transformation process to the new dimension can be done using two steps; in the first step, original input data is transformed into a higher dimensional space using a nonlinear mapping. The second step searches for a linear separating hyperplane in the new space. The maximal marginal hyperplane found in the new space corresponds to a nonlinear separating hyper surface in the original space. Mapping of data into a higher dimension is known as kernelling.

Advantages and disadvantages of SVM

Advantages:

- Good accuracy
- Performs properly on small data.
- Good efficiency, because it applies a subset of training points

Disadvantages:

- Isn't good for huge datasets because the training time is long
- Isn't good for noisier datasets.

3.2.9 Associative Classification:

These methods are performed in a two-step process including frequent item-sets mining followed by rule generation. The first step searches for patterns of attribute–value pairs that are observed repeatedly in a data set. The second step analyses the frequent item-sets to create association rules. The associative classification includes the steps below:

- find the data for frequent item-sets, that is, find usually appearing attribute–value pairs in the data.
- Study the frequent item-sets to create association rules for each class, which fulfill confidence and support vector points.
- Arrange the rules to form a rule-based classifier.

The oldest and easiest algorithm for associative classification is CBA (Classification Based on Associations). This classification applies a repeated method to frequent item-set mining, where several passes created from frequent item-sets are used to create and test longer item-sets. The number of passes created is equal to the length of the tallest rule found. CBA applies a heuristic method to build the classification model, where the rules are sorted according to decreasing precedence based on their confidence and support.

```
KandomkowsList = Kandom( 1k);
For each row in RandomRowsList
PartionNumber = Mod (rownumber , nfolds);
PartitionedRandomRow= Concatenate (RandomRows , PartionNumber);
End for;
maxbucket = nfolds-1;
total_err=0;
for Iter in 0 to maxbucket loop
create CBA Model(TRi);
Error_rate=CBA_Model.get_error_rate;
total_err=total_err + Error_rate;
end for;
Average error rate=total err/ nfolds:
```

Classification based on Multiple Association Rules (CMAR) diverges from CBA in its technique for repeated item-set mining and its building of the classification's model. It uses several different rules' pruning methods with the assistance of a tree structure for dynamic storage and retrieval of rules. CMAR applies a distinct of the FP-growth algorithm to detect the whole set of rules achieving the minimum confidence and minimum support thresholds.

CPAR which stands for Classification based on Predictive Association Rules employs a unique method to rule generation, according to a rule creation algorithm for classification called FOIL (Gupta and Toshniwal, 2011). FOIL construct rules to recognize positive items (e.g., employed = yes) from negative items (e.g., employed = no). CPAR applies the best k rules of each group to predict the class label of X, according to expected accuracy.

3.2.10 Discriminative Frequent Pattern–Based Classification:

The discriminative power of a low-support feature is limited by a low value due to its shortage of coverage in the dataset. Therefore, the contribution of low-support features in classification is bounded, which explains the usage of frequent patterns in classification. Also, existing frequent pattern-mining algorithms can simplify the pattern generation, thus solving the scalability issue in the classification of huge datasets (Atsumi *et al*., 2007).

The general scope for discriminative frequent pattern–based classification includes these steps:

1. Attribute creation: The data, D, are partitioned according to the class name. Use frequent item-set mining to find frequent patterns in each partition, satisfying minimum support. The collection of frequent patterns F creates the feature candidates.
2. Attribute choosing: Use attribute choosing to F , resulting in Fs , the group of chosen frequent patterns.
3. Training of prediction classifier: A classifier relies on the data set D. Any training method can be applied as the prediction classifier.

Advantages of discriminative frequent pattern–based classifier:

- Performance accuracy is considered high.
- performs well when using samples include errors.
- Quick estimation of the trained target function.

Disadvantages:

- It takes a long time in the training process.
- Hard to figure out the assigned weights.
- Hard to incorporate domain knowledge.

3.2.11 Fuzzy Classifications Methods:

A fuzzy neural network introduced memory connections for classification and weight connections for selection, so that it tackles concurrently two foremost issues in pattern recognition that is pattern classification and feature selection. Fuzzy neural systems are used in pattern recognition applications. Lin and Lee presented a neural network in 1996 which composed of fuzzy neurons. In this section we will briefly discuss the techniques and methods that have been used in a fuzzy logic approach in classification.

3.2.12 Fuzzy Rule-based Classifier:

Fuzzy-rule-based control systems which are also called a fuzzy inference system (FIS) have been proposed to address various problems (Chaudhari and Patil, 2014). Fuzzy rules have been advocated as a key tool for articulating parts of knowledge in “fuzzy logic”. However, there does not exist a suitable type of fuzzy rules, nor is there only one type of “fuzzy logic”. This variety has caused many a misunderstanding in the literature of fuzzy control. Fuzzy rules in these systems were commonly derived from human experts. There are many methods introduced for automatically creating fuzzy rules from numerical data without domain experts. Adapting methods for the membership functions of antecedent and consequent fuzzy sets have also been introduced in several studies, such as adapting techniques relied on descent method (Bardenet et al., 2013), and other studies combined the learning ability of neural systems with fuzzy control systems to form self-learning fuzzy controllers. Genetic algorithms have also been employed for the learning of fuzzy rules. Also, fuzzy partitions were introduced by Ruspini 1969 (Mesiar and Rybárik, 1998).

In building the classifier, the basic role of fuzzy rules is to create the vague classification schemes, as commonly applied by a human, transparent in a formal and computer-realizable application. Fuzzy-rule-based classification systems (FRBCSs) may be assigned to two classes of classification systems: those which are supposed to perform autonomously, and those which are intended to be tools in the hands of the user to help him to make decisions (Liu et al., 2019).

Building classifiers in Rule-based systems have the problem that they include exact values for continuous features. The rule indicates that students who have two or more skills and who have a high grade (those with a grade of 90 or more) in math are employed:

if (no_of_skills>=2) AND (math_grade>=90) THEN employed = yes

By the above Rule, a student who has at least two skills will be employed if their mathematics grade is, say, 90 but not if it is 89. Such rule is too rigid in practice.

Alternatively, we can classify values into classes {low_grade, medium_grade, high_grade} and use fuzzy logic to make "fuzzy" thresholds or boundaries to be assigned for each class. Instead of having a precise cutoff between classes, fuzzy logic applies truth values between 0.0 and 1.0 to indicate the level of membership that a certain value is in a given class. Each class forms a fuzzy set. By using fuzzy logic, we can discover the notion that a mathematics grade of 89 is, more or less, high, although not as high as a math grade of 90. Fuzzy logic systems typically present graphical tools to help users in transforming attribute values to fuzzy truth values.

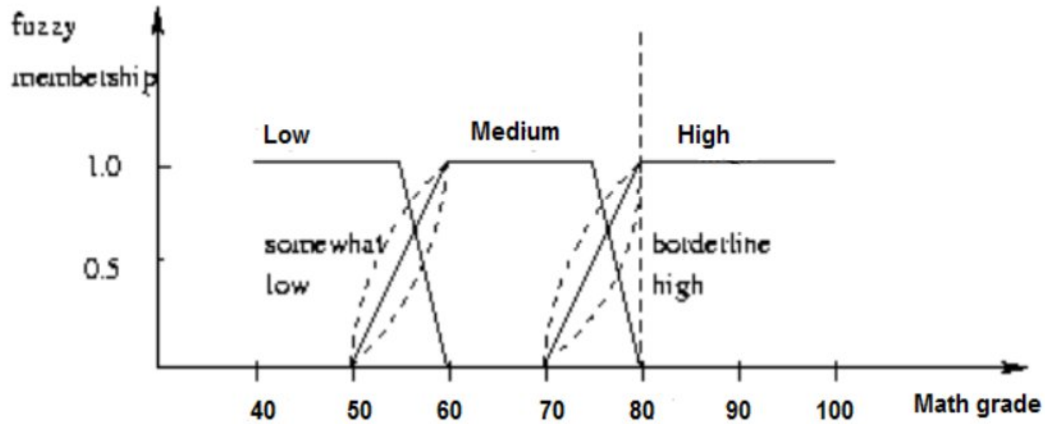

Figure 3. 9 Fuzzy truth values for math grade, representing the degree of membership of math grade with respect to the categories

Fuzzy set theory is also called possibility theory. It makes us perform at a high abstraction level and provides tools for dealing with imprecise data measurement. Most importantly, fuzzy set theory allows us to deal with vague or inexact facts. For that reason, we use a fuzzy set approach to build a classification model to predict the level of membership of employability to certain classes.

3.2.13 Neuro-Fuzzy Systems:

A neuro-fuzzy system is the combination of a fuzzy inference system and neural network (Walia, Singh and Sharma, 2015). A neuro-fuzzy system has been introduced to handle the shortcomings of neural networks, as well as the shortcomings of fuzzy systems, because it can represent knowledge in an interpretable way and learning ability. A neuro-fuzzy classifier (NFC) is considered as one of the neuro-fuzzy systems (Belaout et al., 2017), NFC combines the powerful description of FIS with the learning capabilities of neural networks to partition a feature space into classes.

The neuro-fuzzy modelling approach field is classified into two sections: the linguistic fuzzy modelling that concentrates on interpretability, mainly the Mamdani model; and the precise fuzzy modelling that concentrates on accuracy, mainly the Sugeno model or Takagi–Sugeno–Kang (TSK) model (Ghosh et al., 2014).

The rule based neuro-fuzzy classification approach usually uses the method of an adaptive neural system, namely adaptive neuro-fuzzy inference system (ANFIS) (Hamdan and Garibaldi, 2013). The ANFIS is a multilayer feedforward system that applies neural network learning algorithms and fuzzy logic to map an input space to an output space. It normally employs the Sugeno fuzzy model to produce IF-THEN learning rules. ANFIS techniques have the potential of training, building, and classifying. It has the advantage of tuning complex conversion of human intelligence to fuzzy systems. The disadvantage of the ANFIS predicting model is the time needed for training structure and determining parameters, which took much time. The structure of ANFIS comprises five layers (Figure 3.10)

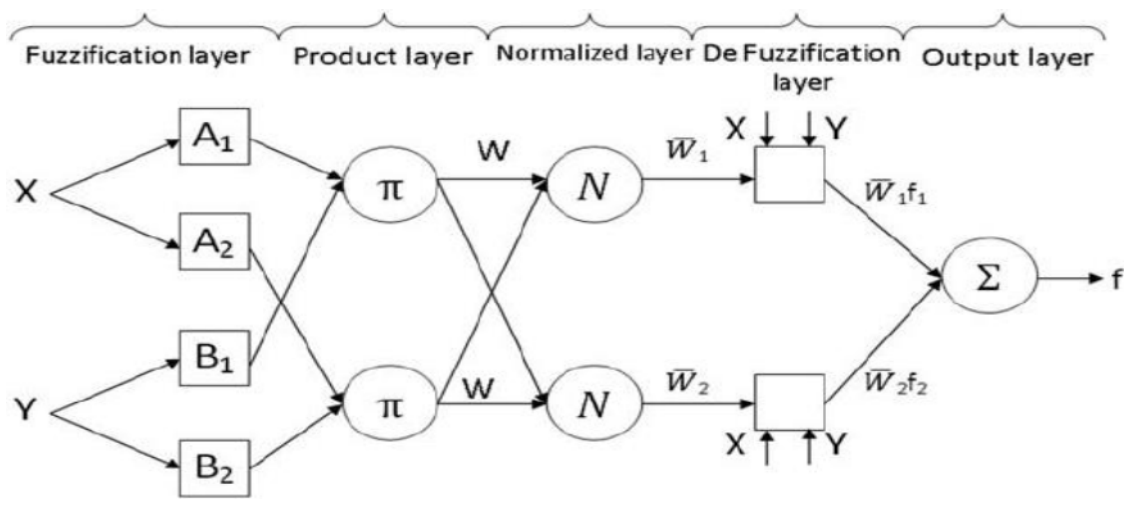

Figure 3. 10 ANFIS architecture for two-input Sugeno fuzzy model with four rules

The following are the main steps of ANFIS technique:

Step 1: Calculating Membership Functions:

Membership functions are used to find "likeness" or "degrees of membership" that a certain value has. These are typically represented by sinusoidal curves over a range of values. So if you are estimating energy consumption for a given year, you would break apart the values from least to greatest into five distinct categories with a deviation value causing a degree of spread and similarity between each of those values.

Step 2: Firing Strength of Fuzzy Rules:

The values, which were "fuzzified" in step 1 are now transported to a node layer and multiplied by the strength of an automatically generated fuzzy rule (Yao et al., 2009). Think of this as computing a "weight" based on the automatically generated rule and the data generated to that rule.

Step 3: Normalize Firing Strength Calculation:

The third step takes the output from step 2 and compares the value of the firing strength from the former node against the sum of all firing strengths (Yao et al., 2009). To simplify this description, think of the algorithm comparing the strength of an output rule of a single node, to the strength of other nodes and their underlying rules. If the strength is greater for a certain node, it is probably the "best possible" rule configuration for the dataset and is prioritized for the next step.

Step 4: Combine Premises with Consequents:

The fourth step takes the weighted values combined with the original inputs from the training data set to calculate an output based on the consequents data.

Step 5: Predict and Final Output:

The last step oversees the calculation of the sum of all incoming signals and uses them to a test dataset to generate a predicted value. This step also includes the de-fuzzification process for the data and translation back to meaningful values. The following is an ANFIS algorithm.

We have adopted an ANFIS technique in this study. Therefore, we will discuss this approach in detail in the next chapter.

3.2.14 Rule-based Classifiers:

In rule-based classifiers the learned model is represented as a group of IF-THEN rules. In the first section, we will explain how these rules are used for classification. The second section illustrates how to extract these rules using a decision tree. The third section describes how to use a sequential covering algorithm (Vishwakarma and Kapoor, 2012).

- Using IF-THEN Rules for Classification:

Knowledge and information are represented using rules. A rule-based classifier employs a group of IF-THEN rules for classification (Menardi and Torelli, 2014). An IF-THEN rule is a representation of the shape IF condition THEN conclusion. In the following statement R is rule:

R: IF GPA = A AND English level = advance THEN employed = yes

The left section (IF) of a rule is called the rule ancestor or precondition. The "THEN" section is the rule consequent. In the rule ancestor, the condition has one or more attribute tests (e.g., gpa = A and English level = advance) that are logically AND. The rule's consequent includes a class prediction (in this case, we are predicting whether a student will be employed). If all the attribute verifies in a rule antecedent assigned to true for a given data, then the rule antecedent is satisfied and that the rule covers the data. A rule R can be solved by its coverage and accuracy. Given a data, X, from a class-labelled data set, D, let ncovers = numbers of items covered by R; ncorrect = numbers of items rightly classified by R; and |D| be the number of items in D. We can assign the coverage and accuracy of R as (3.10):

3.10 $\boldsymbol{coverage(R) = n_{covers} / |D|}$

And as (3.15)

3.11 $\boldsymbol{accuracy(R) = n_{correct} / n_{covers}}$

The rule's coverage is the ratio of items that are covered by the rule. For a rule's accuracy, we calculate the percentage of them and the rule can rightly classify. If more than one rule is invoked, we should figure out which rule gets to fire and assign its class prediction to X. There are several methods. They are; size ordering and rule ordering.

The size ordering method defines the largest priority to the triggering rule that has the toughest requirements, where toughness is estimated by the rule antecedent size. The triggering rule with the most attribute tests is fired. The rule-ordering method gives the priorities to the rules in advance. The ordering of rules may be class-based or rule-based.

The class-based ordering, sorts the classes order of decreasing according to their importance. in this ordering technique, the rules are ordered into one long priority list, according to some measure of rule quality, like accuracy, coverage, or size, or according to advice from human experts.

3.3 Improving Classification Accuracy Techniques:

This part discusses the methods used for enhancing the accuracy of a classification model. One of the most important techniques that improves the classifier efficiency is ensemble methods. An ensemble method is a machine learning technique that combines several base classifiers to create one better classification model. Examples of ensemble methods are bagging, boosting, and random forests.

3.3.1 Bagging:

Bagging combines Bootstrapping and Aggregation to represent one ensemble model. With a dataset, several bootstrapped subsamples are pulled (Tiwari, 2017). A decision tree is represented on each of the bootstrapped subsamples. After each subsample decision tree has been generated, an algorithm is applied to aggregate over the decision trees to represent the most efficient classifier. Figure 3.11 will illustrate the steps of bagging algorithm processes:

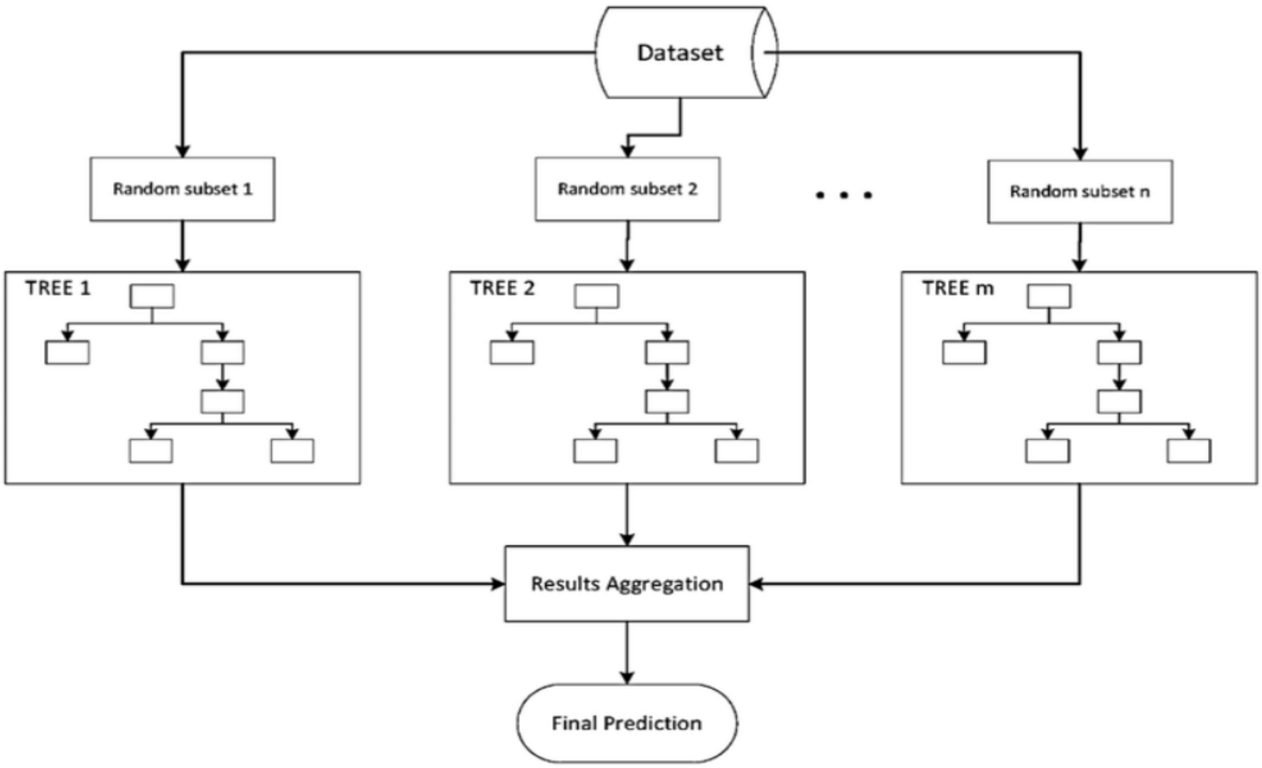

Figure 3. 11 Steps of Bagging algorithm

The bagged predictor usually has much better accuracy than one classifier derived from the actual training data. It will not be worse and is more robust to the effects of noisy data and over-fitting. The increased accuracy happens because the combined model decreases the difference of the individual classifiers.

3.3.2 Boosting:

Boosting is similar to bagging, but in addition to splitting the dataset, it assigns certain weights to each training data. Bootstrapped sub datasets are pulled from a larger dataset (Li, Wang and Sung, 2008). A decision tree is represented on each subsample. However, the decision tree is split into different features. In boosting, a sequence of k classifiers is repeatedly learned. After a classifier, Mi, is learned, the weights are adapted to make the subsequent classifier, M i+1, to consider the training items that were misclassified by Mi. The final boosted classifier, M*, merges the votes of each unique classifier, where the weight of each classifier's vote is a function of its accuracy.

Friedman et al. (2000) in (Chu, Lee and Ullah, 2020)presented a statistical view of the AdaBoost (short for Adaptive Boosting) algorithm. They analysed AdaBoost as stage-wise evaluation functions for matching an additive logistic regression model. They proved that AdaBoost was actually minimizing the exponential loss function.

Sometimes the result of a “boosted” model may have accuracy than an individual model that is generated from the same data. Bagging is less susceptible to model over-fitting. When both can greatly enhance the accuracy comparing to a single model, boosting tends to reach greater accuracy level.

3.3.3 Cost–Benefit and ROC Curves:

Cost-Benefit and ROC curves is a method to compare the classifiers’ accuracy. There are two methods of visualizing classifier’s accuracy, assigning proper thresholds based on the operating condition, and extracting useful aggregated measures like the place under the ROC curve (AUC) or the place under the optimal cost curve (Hernández-Orallo, Flach and Ferri, 2013). The cost related to a false negative is far larger than those of a false positive. To calculate classifier accuracy, we have defined similar costs and divided the summation of TP and TN by the total number of test items.

ROC curves are employed as a visual instrument for comparing two classifiers. A ROC curve for a certain classifier represents the trade-off between the true positive rate (TPR) and the false positive rate (FPR). To compare between two classifiers, a ROC curve allows us to visualize the trade-off between the rates at which the classifier can correctly predict positive cases against the rate at which it incorrectly discovers negative cases as positive for different portions of the test set. To plot a ROC curve for a given classifier model, M, the model should return a probability of the predicted class for each test data. Figure (3.12) shows the relationships between bad, good, and great models according to the ROC curve.

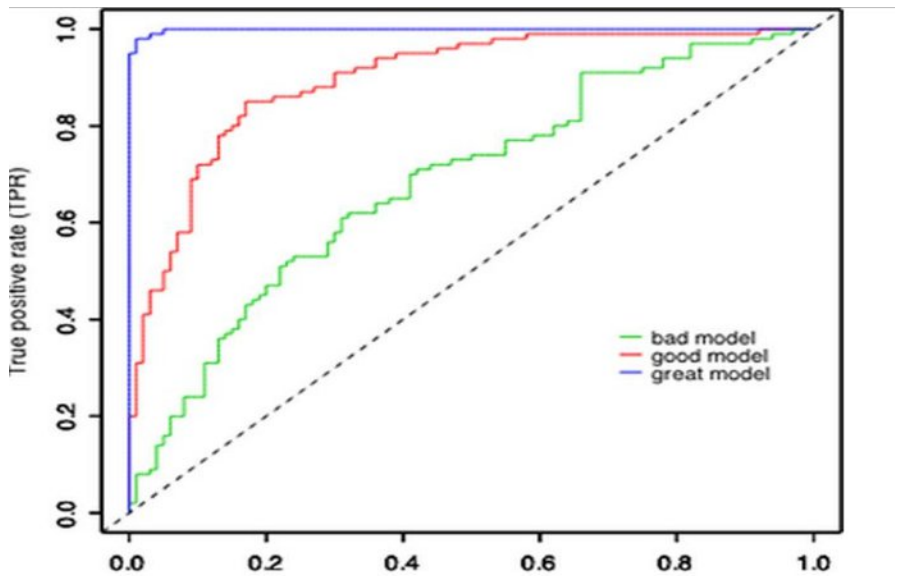

Figure 3. 12 Relationship between bad, good and great models according to ROC curve

3.4 Classifiers Evaluation:

After building the classification model, we need to estimate the accuracy of the classifier. In this study, we have used several techniques to create more than one classifier. Therefore, we need to compare their accuracy to find the best performance. There are common techniques for assessing accuracy, according to arbitrarily sampled partitions of the given data. We will discuss these techniques in the next sections.

3.4.1 Holdout Method and Random Subsampling:

In this method, the used dataset is randomly divided into multiple separated groups, a training set and a testing set (Roelofs et al., 2019). Two-thirds of the data are assigned to the training set, and the remaining one-third is allocated to the test set. The training set is employed to extract the model. The classifier's accuracy is then evaluated with the test set, as shown in (Figure 3.13).

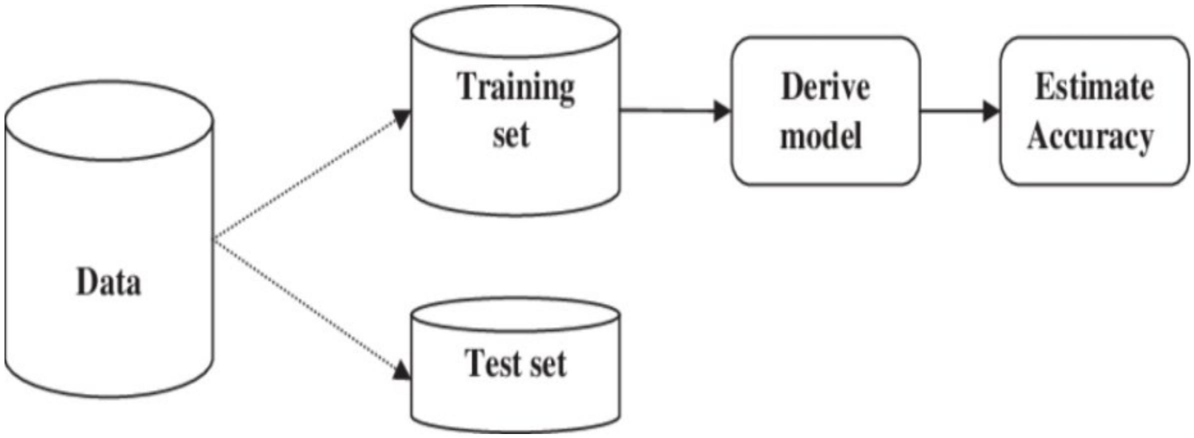

Figure 3. 13 Evaluating accuracy with the holdout technique.

Random subsampling is different from the holdout technique in which the holdout method is looped k times. The final accuracy evaluation is found by calculating the average of the accuracies taken from each iteration (Wald, Khoshgoftaar and Fazelpour, 2013).

3.4.2 Cross-Validation:

In this validation technique, the function has a one parameter called k that represents the number of groups that a certain data set is to be split into. This method is usually known as k-fold cross-validation. If a certain value for k is selected, it may be used in the location of k that is represented in the model, like k=10 will be 10-fold cross-validation. For classification, the accuracy evaluation is the final number of right classifications from the k iterations, divided by the whole number of items in the initial data (Hjorth and Hjorth, 2018).

3.4.3 Bootstrap:

The bootstrap strategy tests the given training items consistently with substitution (Hesterberg, 2011). Each time a data is chosen, it is similarly liable to be chosen again and re-added to the training data set. This strategy performs as follows, for a given data collection of d items. The data set index is divided into d times, with substitution, bringing about a training data set of d tests. The data items that didn't make it into the training set end up representing the test set. Almost 0.632 of the original data will end up in the bootstrap, and the remaining 0.368 will represent the test set (since $(1 - 1/d)d \approx e\text{-}1 = 0.368$). Then iterate the sampling procedure k times, final accuracy of the classifier:

3.12 $$acc(M) = \sum_{i=1}^{k}(0.632 \times acc(M_i)_{test\_set} + 0.368 \times acc(M_i)_{train\_set})$$

3.4.4 Confusion Matrix:

Is a table that is used to evaluate the efficiency of a classification model on a group of test data for which the true values are defined. It visualizes the performance of an algorithm (Ting, 2017). It is a helpful tool to estimate the performance of classifiers; it gives us a better understanding not only of the mistakes being produced by a classifier but it gives us kinds of mistakes that are being produced. In this study we will use a confusion matrix to test our classifier and compare it with other classifiers.

	Class 1 Predicted	**Class 2 Predicted**
Class 1 Actual	**TP**	**FN**
Class 2 Actual	**FP**	**TN**

Table 3. 1 Confusion matrix

- True positives (TP): represents the true items that were rightly assigned by the classifier. Let TP be the number of true positives. For example, if we have two classes, the TP may be employed = yes where the negative items are employed = no.

- True negatives (TN): Represents the correct negative items that were assigned by the classifier.

- False positives (FP): Represents the incorrect negative items that were assigned as positive.

- False negatives (FN): estimates the positive items that were mislabelled as false.

The accuracy of the classifier can be calculated as follows:

3.13 *Accuracy = (TP + TN) / (TP + TN + FP + FN)*

The value of Recall provides us an idea about when it’s really yes.:

3.14 *Recall = TP / (TP + FN)*

Precision value indicates when the prediction is yes.

3.15 *Precision = TP / (TP + FP)*

F-measure: We compute an F-measure which applies Harmonic Mean in place of Arithmetic Mean as it disciplines the extreme values more. The F-Measure will always be closer to the smaller value of Precision or Recall.

3.16 F-measure = (2 * Recall * Precision) / (Recall + Precision).

Table 3.2 shows how to utilize the confusion matrix by applying a real dataset on a classifier.

Class	Employed = yes	Employed = no	Total
Employed = yes	201	99	300
Employed = no	72	328	400
Total	273	427	700

Table 3. 2 Confusion matrix of applying real dataset on a classifier

As shown in the above table, confusion matrix contains two classes; (Employed = yes) and (Employed = no) with a dataset of 700 instances that represents graduates. The classifier classifies 201gradute as (Employed = yes) out of 300 and classify 328 graduates as (Employed = no) out of 400.

3.5 Attribute selection approaches:

The measure of attribute selection gives a ranking for each attribute representing the point of splitting the training into smaller subsets. The measure of attribute selection techniques is also known as splitting rules, to decide which way items are going to be split. An attribute selection measure is a heuristic for choosing the separating standard that best splits a certain data partition, D, of class-labelled training items into separated classes (Sun and Hu, 2017). If the data partition D was shifted into smaller partitions based on the outcomes of the splitting standard, if possible, each partition would be pure. The best splitting standard is the one that most accurately results in such a case. The feature selection approach decreases the number of input variables when building a predictive model (Iguyon and Elisseeff, 2003). It is desirable to decrease the number of input variables to both decrease the computational cost of modelling and, for some cases, to increase the performance of the model. According to this, the following points became obvious:

- There are two main kinds of attribute selection techniques: supervised and unsupervised. Supervised approaches could be divided into wrapper, filter and intrinsic.
- Filter-based attribute selection approaches use statistical measures to achieve the correlation or dependence between input variables that can be filtered to select the most suitable features.
- Statistical measures for attribute selection must be carefully selected, based on the data type of the input variable and the output or response variable.

Statistical-based attribute selection approaches involve assessing the relationship between each input variable and the target variable using statistics and choosing those input variables that have the deepest relationship with the target variable. These approaches can be fast and effective, although the selection of statistical measures depends on the type of the data for both the input and output variables. For instance, it can be critical for a machine learning expert to select a suitable statistical measure for a dataset when performing filter-based attribute selection.

In this regard, a number of predictive modelling problems involve a large number of variables that can reduce the development and training of models and need a large amount of system memory. Furthermore, some models cannot perform properly when including input variables that are not relevant to the target variable. So, feature selection approaches are in terms of supervised and unsupervised methods (Ladha and Deepa, 2011).

The variation has to do with whether features are chosen depending on the target variable or not. Unsupervised attribute selection approaches ignore the target variable; such as approaches that remove redundant variables using correlation. Supervised attribute selection approaches use the target variable, such as approaches that remove irrelevant variables. Another way to classify the attribute used is to determine features which may be divided into wrapper and filter approaches (Wang, Wang and Chang, 2016). These approaches are almost always supervised and are evaluated based on the performance of a resulting model on a dataset. Wrapper feature selection methods build models with different splits of input features and select those features that result in the best accuracy of the model according to a performance scale. These methods are indifferent with the variable types, although they can be computationally expensive. Recursive Feature Elimination RFE is a good example of a wrapper feature selection method (Inza et al., 2004). Filter feature selection approaches use statistical techniques to assess the correlation between each input variable and the target variable, and these degrees are used as the basis to select (filter) those input variables that would be applied in the model. Additionally, there is a set of machine learning algorithms that make feature selection automatically as part of learning the model. We can refer to these techniques as intrinsic feature selection approaches (Xue et al., 2016).

There are three main measures of attributes selection:

- Information gain
- Gain ratio
- Gini index (CART uses Gini Index as Classification matrix.).

3.5.1 Information Gain

This measure used the method proposed by Claude Shannon, which introduced the value or "information content" of messages. This technique measures the value of information content for the attributes and selects the attribute that achieves the highest value of information gaining. Let pi be the probability that an arbitrary data in D refers to class Ci, evaluated by |Ci, D|/|D|.

entropy required to classify a data in D:

3.17 $$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i)$$

where pi is the nonzero probability that a randomly data in D refers to class Ci and is evaluated by |Ci, D|/|D|. Info (D) is just the average set of information required to assign the class label of a data in D. The given information depends on the proportions of items of each class. Info (D) is also called the entropy of D.

The required information to classify D:

3.18 $$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times I(D_j)$$

The lower the expected information needed, the larger the purity of the partitions. Information gain is assigned as the margin between the actual information requirement and the new requirement. Information gained by branching on attribute A:

3.19 $$Gain(A) = Info(D) - Info_A(D)$$

Gain (A) shows how many branches generated on A. It is the expected decrease in the information requirement given by defining the value of A. The feature A with the largest information gain, Gain (A), is selected as the splitting feature at node N.

If the attributes are numerical values, then we have continuous-valued attributes. The given attribute A that is continuous-valued, rather than discrete-valued, for this scenario, we should define the "best" split-point for A, where the split-point is a threshold on A. To determine the split-point, we first sort the values of A in increasing order. Usually, the midpoint between each couple of adjacent values is assigned as a possible split-point. (ai +ai+1)/2 is the midpoint between the values of ai and ai+1. The stage with the smaller information requirement for A is chosen as the split-point for A. D1 is the set of items in D satisfying A ≤ split-point, and D2 is the set of items in D satisfying A > split-point.

3.5.2 Gain Ratio

This measure uses the C4.5 algorithms as an extension to information gain called gain ratio. The information gain measure is one-sided toward tests with several results. That is, it favors choosing features with larger numbers of values. C4.5 employs a gain ratio to solve the normalization problem:

3.20 $$SplitInfo_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times \log_2\left(\frac{|D_j|}{|D|}\right)$$

The above equation describes the information produced by splitting the training data set, D, into v divisions, identical to the v results of a test on attribute A. It varies from information gain, which measures the information with estimate to classification that is achieved according to the same division. The gain ratio is illustrated as (3.21)

3.21 *Gain Ratio(A) = Gain(A)/Split Info(A)*

The feature with the largest gain ratio is chosen as the splitting feature. In this research we used C4.5 to gain the information in order to apply a decision tree classification task.

3.5.3 Gini Index

The Gini index is used in CART. If a given data set D includes values from n classes, Gini index, Gini(D) is illustrated as:

3.22 $$gini(D) = 1 - \sum_{j=1}^{n} p_j^2$$

If D is shifted on A into two groups D1 and D2, the Gini index Gini(D) is assigned as:

3.23 $$gini_A(D) = \frac{|D_1|}{|D|} gini(D_1) + \frac{|D_2|}{|D|} gini(D_2)$$

The decrease in impurity can be represented in (3.8):

3.24 $$\Delta gini(A) = gini(D) - gini_A(D)$$

The feature that maximizes the decrease in impurity is chosen as the splitting point. This feature and either its splitting subset or split-point together form the splitting point.

Chapter Summary

Classification is a type of information analysis that extracts models describing data classes. The classification techniques can be divided into basic and advanced methods. Many methods and algorithms have been proposed to handle classification problems, such as classification rules-based, decision trees, and neural networks. The main advantage of rule-based classifiers is that they are easy to interpret by humans. There are methods for building classification models that can be easily implemented, such as naïve Bayesian classification; the drawback of Bayesian classifier is that it doesn't consider the relationships between attributes. Some other classification techniques such as support vector machines create a line or a hyperplane which separates the data into classes; the disadvantage of SVM is that the training time can be high. A neuro-fuzzy approach which is a combination of neural network and fuzzy logic can be applied to perform classification jobs. A neuro-fuzzy classification approach usually applies the concept of adaptive neural network, namely adaptive neuro-fuzzy inference system (ANFIS). The ANFIS is a multilayer feedforward network which uses neural network learning algorithms and fuzzy logic to map an input space to an output space.

4 Chapter 4. METHODOLOGY

This thesis focuses on employability problems. To the best of our knowledge, research studies that have worked in Arab countries on CS graduates employment using data mining are almost non-existent. Decision-makers in ministries of higher education and educational institutions realize the importance of getting all or most of their graduated students employed. Therefore, they strive to carry out continuous studies about the factors that may improve employment chances. Studying employability also helps these institutions to improve curriculums and study plans. It also helps them to open new specializations that achieve the requirements of the labour market and to reduce the focus on redundant specializations.

Jordan has a great number of universities with CS specializations and recruitment offices as well. Due to these facts, I chose Jordan, which gave me the ability to get a suitable data sample from several graduates' tracer offices from different universities. Moreover, I managed to get statistical data from official sources, such as the Ministry of Digital Economy and Entrepreneurship (MDEE) and the Ministry of Labour in Jordan (MLJ).

There are many research studies that have investigated unemployment without using mining the generated data. However, many others have approached the employability problem using various data mining classification techniques, for instance SVM, decision tree, KN-neighbor and neural network, but it appears that was rare for using neuro-fuzzy algorithm (NFA) to study the employability issue.

This chapter is divided into three sections; section 4.1 presents an introduction to our approach that will be used in the experiment. Section 4.2 introduces a statistical study about Jordanian graduates in different Computer Science (CS) fields. Section 4.3 demonstrates the processes of our methodology, which includes four steps as follows: data collection, data preprocessing, implementation of an Adaptive Neuro-Fuzzy Inference System (ANFIS) algorithm, a type of ANN which was developed in the early 1990s (Boyacioglu and Avci, 2010)) and how the evaluation and validation process will be carried out in this thesis.

4.1 Introduction:

The main objective of the current research study is to categorize the graduate students according to their employability status. The data is taken from the graduates' dataset which is built from tracer studies carried out by various universities in Jordan such as Al Balquaa university, Phladilphia university, and Al Zaytounah university. The dataset consists of 1095 instances and 22 attributes that refer to graduates' profiles.

As stated previously, many classification techniques are stated in the literature, and there are many research studies addressing the employability issues using various techniques, such as SVM, K-nearest, naïve based, and artificial neural networks (ANN). In this thesis we will study the employability problem comparing a neuro-fuzzy technique called ANFIS with a number of different classifiers. ANFIS applies both neural networks and fuzzy logic techniques, which takes some features from both of them. ANFIS is an inference system corresponding to a set of fuzzy IF-THEN rules that have learning capability to give non-exact values, but very close to real values from the original data.

A lot of data mining techniques can be used for classification; researchers have to select the technique that is compatible with their statement.

4.2 Statistical insights:

To better understand the Computer Science (CS) market in Jordan, and study the characteristics that may influence on the employment possibility of graduates in the CS field, we contacted the Ministry of Digital Economy and Entrepreneurship and the Ministry of Labour in Jordan, to collect useful data.

The total number of graduates from IT specializations for the year 2018 was 4180 graduates. The percentage of male graduates was 51% and 49% for female graduates. The percentage of graduates according to their academic degrees is as follows: 90% of graduates hold a bachelor degree, 6% a master degree, and 4% a diploma.

Figure (4.1) shows the percentage of graduates from CS the field for the year 2018 according to their specializations – bachelor

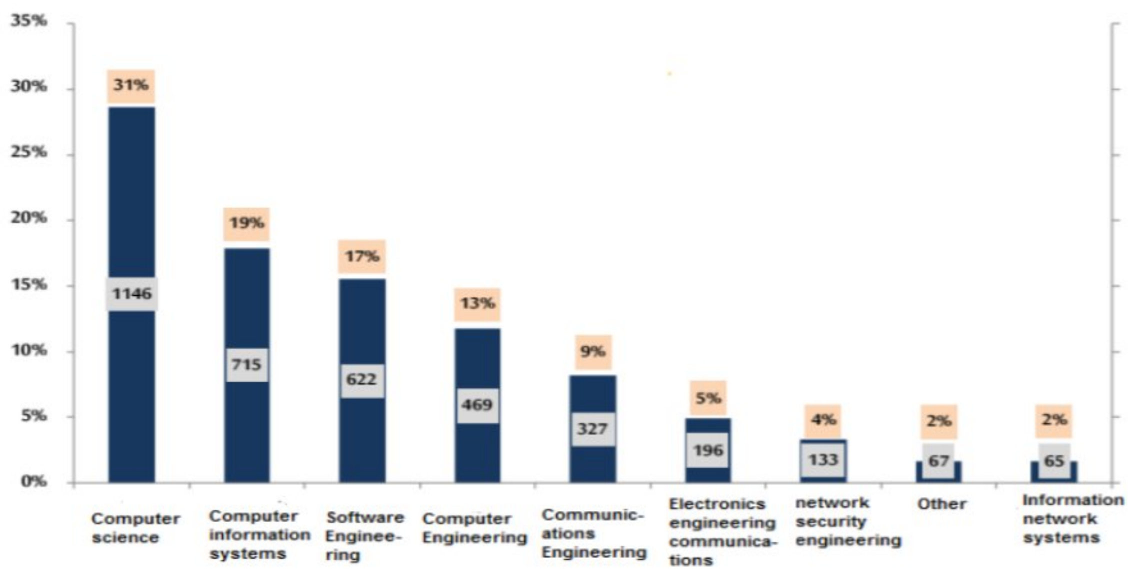

Figure 4. 1 The percentage of graduates from CS field for the year 2018 according to their specializations - bachelor

Figure 4.1 shows that computer science specialization has the largest number of graduates from the IT field with 31% of total graduates for the year 2018, and the lowest percentage goes to information network systems specialization. In the year 2017, the number of IT graduates was 2869; computer science specialization again got the highest ratio, with 29%.

The number of graduates who got jobs in IT in the year 2018 was 1512 which is 40% of total graduates, 64% for male and 36% for female. The percentage of employment for the year 2017 was the same (64%), but the number of male graduates who got jobs was less than 2018 with 61%, and 39% for female.

Most of the IT graduates are employed in the private sector with 92%. Only 8% of total graduates are employed in the public sector.

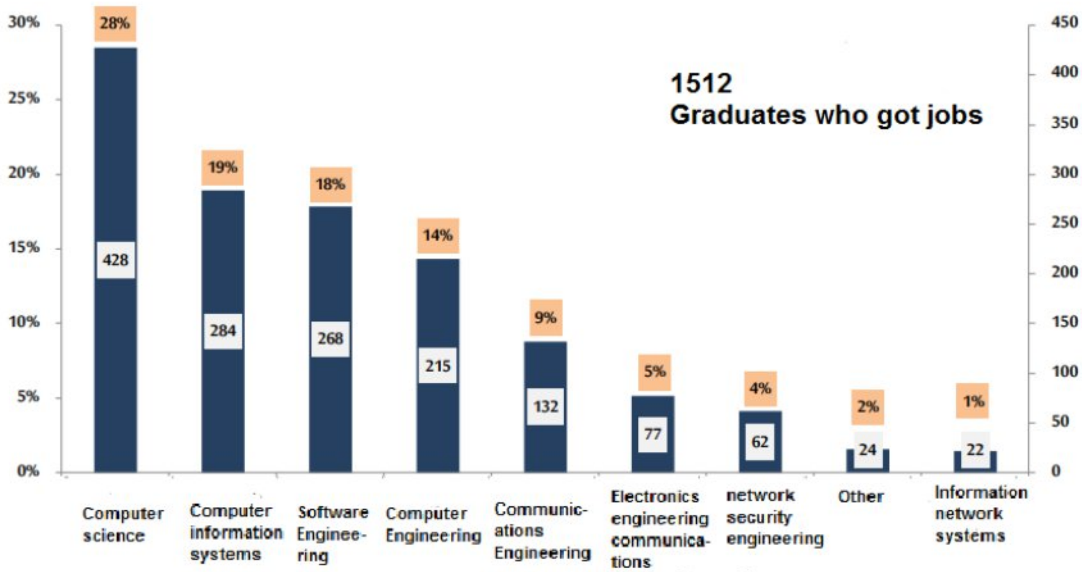

Figure 4. 2 Employment rates according to IT specialization for year 2018

Figure 4.2 shows that computer science graduates got the highest number of jobs, with 28% of total graduates for the year 2018, and the lowest number goes to information network systems specialization with 1% of total graduates.

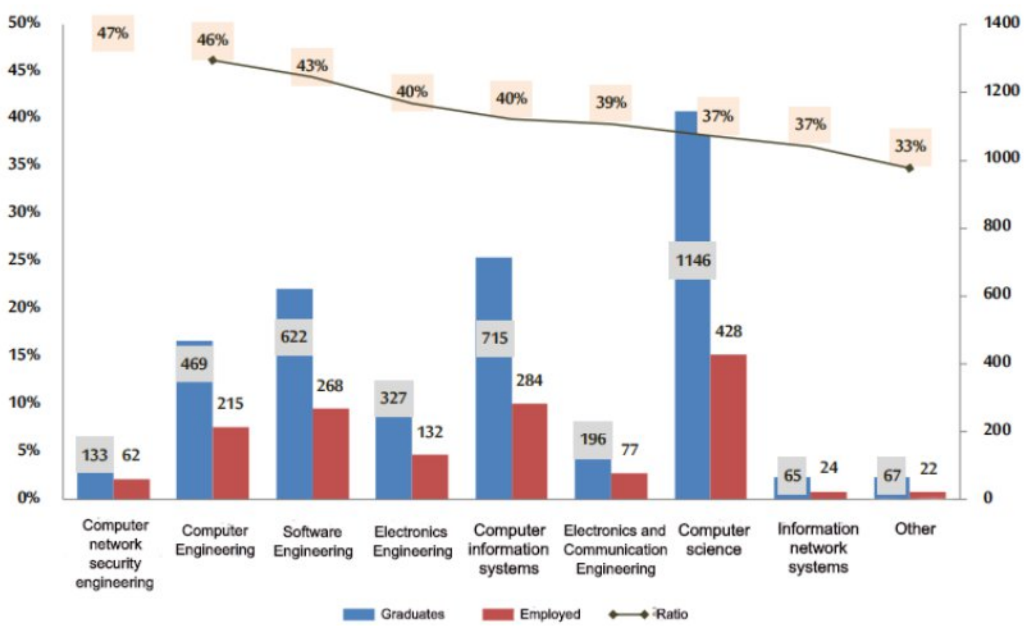

Figure 4. 3 Numbers of employed graduates according to their specializations for year 2018

Figure 4.3 shows that specialization in computer network security engineering has the highest employment rate with 47% of total graduates for the year 2018.

As shown in the above statistics, the demand for male graduates in the IT market at Jordan is higher than that for female graduates. Also, the statistical analysis indicates the need of some specializations in the IT field, such as computer network security engineering, computer engineering, and software engineering. As a result of previous statistics, the attributes that may influence the employability of graduates in the IT market in Jordan are: gender, specialization, and university. In this research, our training sample data includes these attributes, alongside other selected attributes.

4.3 The proposed research framework:

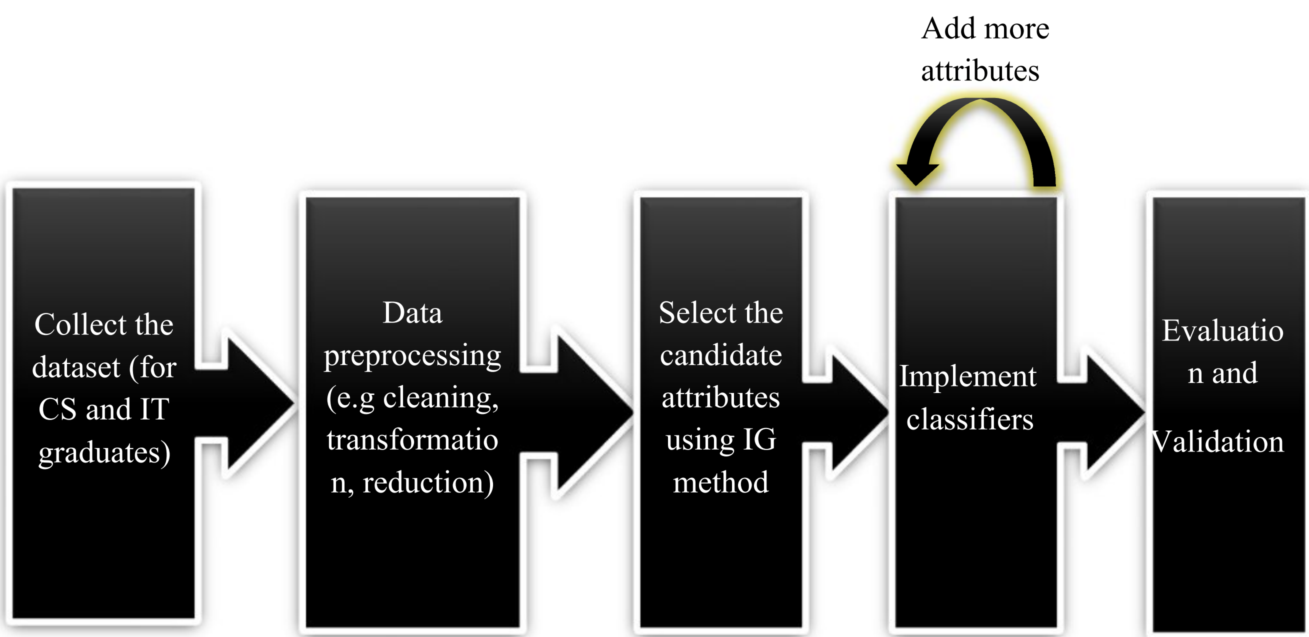

Figure 4. 4 Proposed research framework

4.3.1 Data collection:

The dataset used in this thesis is obtained from three Jordanian universities (Al Balqaa university, Al Zaytonah university, and Phladelphia university), the data collected from the tracer study carried out by career guide units of the same universities. The data was collected for the CS and IT graduates, mainly IT colleges in Jordan comprising various majors such as computer science, computer information system, computer network and software engineering. Recent new majors have been introduced in the universities such as cyber security and AI. In this thesis we study the employability of the information technology majors, thus we collected the data for 1095 IT graduates of three majors (CS, CIS, and software engineering), where the attributes that selected to build the prediction model for CS and IT graduates in this thesis that are not suitable for other majors like business administration or accounting because some attributes like programming skills not applicable to other majors. The graduates' data are distributed according to Table 4.1. The total graduates that studied in this thesis from Balqa Applied University, Philadelphia and Alzaytoneh is 560, 221 and 314 respectively. The table shows there are no graduates in CIS from Philadelphia University. The data was collected for the graduates during the years 2015-2019.

As mentioned previously, there are many definitions of employability. In all definitions the main idea is the ability of graduates to get jobs shortly after graduation. However, most definitions of employability concentrate in their core on one or more of the following elements:

1. Job type: refers to the graduates' ability to get satisfying employment that meets their skills.
2. Time: denotes the required time to get a job after graduation, or get a job without training.
3. Attributes on recruitment: the desired attributes of the graduates that denote the ability to do well or to learn quickly.
4. Lifelong learning: the readiness of the employee for further development and adaptation to the new technology. This notion is critical in the IT market.
5. Employability skills: the acquiring of basic skills determined by the employer which mean the candidate can meet the requirements of the job.

In this thesis 22 attributes that have been selected based on the different research that has been performed to define and determine the main elements of employability.

The dataset attributes are categorized into:

- Demographic attributes: consists of gender, province, social strata, number of applications, time to work, work status and age attributes.
- Soft skills attributes: consists of interpersonal skills, team work skills and talent.
- Technical/hard skills attributes: consists of programming skills, mathematical skills, English skills and number of technical certificates.

- Academic attributes: consists of university name, major, degree, high school grade, GPA, completion period, type of studying, experience.

Attribute	Value	Description
Gender	{male, female}	Takes two categorical values, male or female
University	{bal, ph, zay}	University of graduation
Major	{cs, cis, se}	Program (computer science, computer information system and software engineering)
Province	{Amman, Zarqa, Irbid,}	Province of graduate
Age	{20-25, 26-30, 31-40, >40}	The age divided into four intervals
Programming_ skills	{high, mid, low}	The programming and technical skill divided into low, moderate or high
Educational_ degree	{BSc, master. H.diploma}	Level of certificate of graduate (bachelor, high diploma and master)
Tech _ certificate	Numeric	Total number of computer-related certificates
Time to work	{0-6,6-12,>12}	The time taken to get job after graduation in month
Interpersonal skills	Numeric	Communication skill
High school grade	{60-70,>70-80,>80-90,>90-100}	Grade point average at high school

Social strata	{low, mid, high}	Social strata mostly depends on family income
GPA	{60-68, 68-76, 76-84, 84-100}	GPA at university divided into four categories based on Jordanian system
Type of degree or mode of studying	{evening, regular}	Two types, evening and morning studies
Completion period	{3,4,5,6,7}	Number of years taken to complete the degree
Math skills	{low, mid, high}	Depends on the marks in math courses
English _ skills	{weak, good, v. good, excellent}	Level of English command
Team work _ skills	{bad, low, mid, good, v. good}	Team work skills of graduate
Status	{employed, unemployed, other}	The work status
Experience	Numeric	Years of experience
Talent or hobbies	{art, sport…}	If the graduate has any talent, such as drawing, or other
No. of application	{0-5,5-10,11-15,>15}	Number of applications submitted by the graduate to companies in the field when applying for the job

Table 4. 1 List of dataset attributes

4.3.2 Data preprocessing:

Data preprocessing consists of data cleaning and transformation to make the dataset ready for the data mining algorithm.

1. Data Cleaning:

The collected data may contain inappropriate values or missed values. A cleaning process for the noisy values and filling in the missing data is mandatory.

(a). Missing Data:

This case emerges when the data is missed during the tracer study fulfillment. The missing data can be manipulated in different ways such as:

(1) Remove the whole instance. (2) Fill in the missing values.

The missing values can be filled manually by an expert or automatically by computing the mean of the attribute. In this thesis, because we have a large dataset we simply ignored all instances with missing values. We use the neuro-fuzzy classification technique (ANFIS) to build the classifier incrementally based on specific attribute that selected using IG method which does not need a big dataset.

(b). Noisy Data:

The data are described as noisy if they are corrupted, or distorted, or has a low Signal-to-Noise Ratio. This case arises during data entry error, faulty data collection or any corruption. The noisy data can be handled by (1) binning (2) clustering (3) regression. In this thesis we used the binning method and clustering, where we sorted and partitioned its equal bins then used the bin median to handle the noisy data, as well as we using clustering by dividing data into clusters then detecting and removing the outlier.

2. Data Transformation:

The data must be converted into suitable forms that can be used in any data mining technique. This process is called data transformation. The data transformation consists of the following methods:

1. Normalization process:

This is applied to re-write the value of the data in a manner that reduces redundancy and dependency.

2. Attribute Selection:

   It is done to extract features from the current set attributes to enhance the classification process. In this research we spitted some attributes based on the given attributes from the tracer study, like the "social strata" attributes which were created from "family income" attributes.

3. Discretization process:

   This is applied to substitute the values of numeric attributes by categorical values such as the continuous value of the GPA replaced by interval. Discrete values have critical roles in data mining and knowledge discovery. They are about intervals of numbers which are more concise to represent and specify, and easier to use and comprehend as they are closer to a knowledge-level representation than continuous values. Some of the attributes have been transformed into different categories, such as a GPA attribute. The range is classed into five parts: <60, >60-68, >68-76, >76-84, >84-100. Meanwhile, the age attribute is also being transformed into four ranges: 20-25, 26-30, 31-40, >40. The continuous attributes were transformed into nominal in preparing the data for classification. For example, the "current status" has been changed to nominal from previous numeric code values.

4. Generalization process:

   In this step, the attributes are generalized from lower levels to upper levels in hierarchy. In this thesis a lot of attributes from the tracer study were ignored in the final dataset and replaced with the general attributes, like "street name" replaced by "city".

The data source was transferred to Excel sheets. These files were prepared and converted to (.csv) format in order to be compatible with the data mining software such as WEKA and MATLAB that are used in building the model.

University / Major	Computer science	Computer information system	Software engineering	total
Balqa Applied university	356	71	133	560
Philadelphia university	144	-	77	221
Alzaytonah university	150	69	95	314
Total	656	140	299	1095

Table 4. 2 Collected data from three Jordanian universities

4.3.3 Implement ANFIS:

As discussed previously, there are two types that result from combining neural network and fuzzy logic: (1) Neuro-Fuzzy System (NFS) (2) Fuzzy Neural Network (FNN), and we mentioned the prevalent technique among researchers is NFS; also the most popular algorithm in NFS is the Adaptive Neuro-Fuzzy Inference System (ANFIS). In this thesis ANFIS will be used as a classification technique with modification on the attribute selection process where the IG selection method is used to determine the best attributes that will be suitable for the prediction model, the NFS is unsuitable for large number of attribute due to computational time problem also the attribute added incrementally to the ANFIS to study the effects each attribute on the accuracy of the produced model.

A lot of attributes in the dataset that is used in this thesis contain crisp values that make the prediction process inaccurate; the input attributes in this thesis are converted into fuzzy inputs by the fuzzification process. Fuzzy logic in ANFIS is used to define the weights *w*, from fuzzy sets, in NN. The *w* is computed as a membership function for each input.

ANFIS is an adaptive technique between Sagano's fuzzy inference system and the concept of the neural network approach. The adaptions between the two approaches consist of a set of functions which are represented as a graphical network. The network is constructed from a set of nodes that are distributed into five layers.

To demonstrate how ANFIS works, let us take two inputs variables A1 and A2, and one output variable O1. This inputs variable represents linguistic values such as high, mid and low, and the output value is always crisp value.

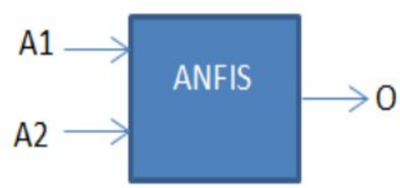

Figure 4. 5 ANFIS concept demonstration

As shown in Figure (4.5), the two inputs are represented using the suitable membership function distribution (such as triangle, Gaussian or trapezoidal). In ANFIS the output O is not represented in any membership function distribution, because it uses Takagi and Sugeno's approach which is unlike the Mamdani approach where the output is represented using membership function distribution. In ANFIS the output is expressed as a function of input parameters:

4.1 $O = f(A_1, A_2)$

The output function can be considering as a linear or nonlinear function. The linear function of the input parameter is used most of the time.

We can express the output function in linear wise like the equation 4.2.

4.2 $O_i=p_iA_1+q_iA_2+r_i$

Where i=1,2, 3…., n

In the equation 4.2, A1 and A2 represents the input parameters, also pi , qi and ri represent the coefficients of the equation. These coefficients update after each iteration of ANFIS using least square error techniques with backpropagation or using any natural inspired techniques like Genetic Algorithm, to get the best value from them. The notation i refers to the number of rules which generate from the linguistic variable of the input parameters.

The ANFIS algorithm comprises five layers as shown in Figure (4.6). Where the linguistic variable is received as input then five steps are done to give crisp value, which is almost similar to the consequence of the rule (i.e. the class employed, unemployed or other). The rules are evaluated according to the training dataset. The layers are:

5. Layer 1: the input is the linguistic values and the output is the computed membership function for each linguistic value.
6. Layer 2: the inputs are the values of the membership function of a particular rule and the output W is the result of applying any t-norm (such as prod. or min) on the output of the previous layer.
7. Layer 3: the input is all the w values of the previous layer and the output is the normalized w value for a specific rule.
8. Layer 4: the input is the input parameters (A1 and A2) and the normalized values that are generated from the rules, and the outputs are

   $$W_1O_1,w_2O_2,\ldots\ldots,W_nO_n$$

   where $\boldsymbol{O_i=p_iA_1+q_iA_2+r_i}$
9. Layer 5: the output of this layer is the result of the summation process on the values that were obtained from layer four.

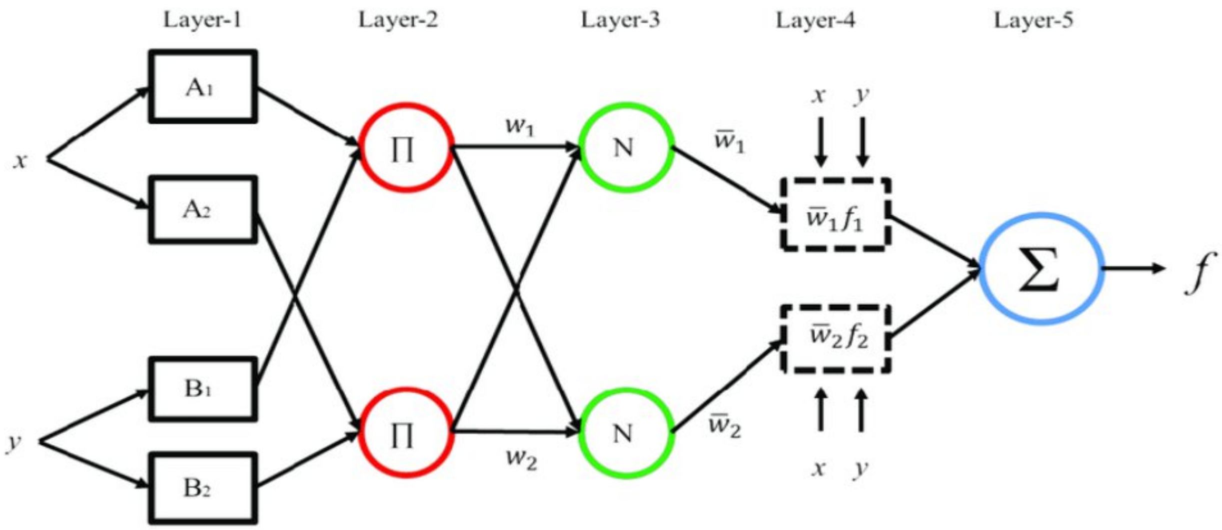

Figure 4. 6 Adaptive Neuro-Fuzzy Inference System

In this thesis the number of input parameters over 20, but will show here ANFIS architecture for two inputs parameter to simplified the presentation.

In this thesis we used incremental techniques to produce the classifier. We start by three parameters as input to ANFIS, then produce the classifier using the training dataset, then find the accuracy of the classifier using testing dataset. The number of input parameter increased incrementally by one. Our strategy was to build a set of classifiers using different number of input parameters. The computation time and accuracy of each produced classifier recorded and compared with another classification technique that considered as a classification benchmark. In this thesis we try to build classifier with best number of input parameters and the most suitable attributes. The final result that we try to get is the best classifier with better attributes with considers the computational time that needed by ANFIS to produce the classifier.

Figure 4.7 shows the network with two input values x and y, where each input has two linguistic values, A1 , A2 and B1, B2, respectively.

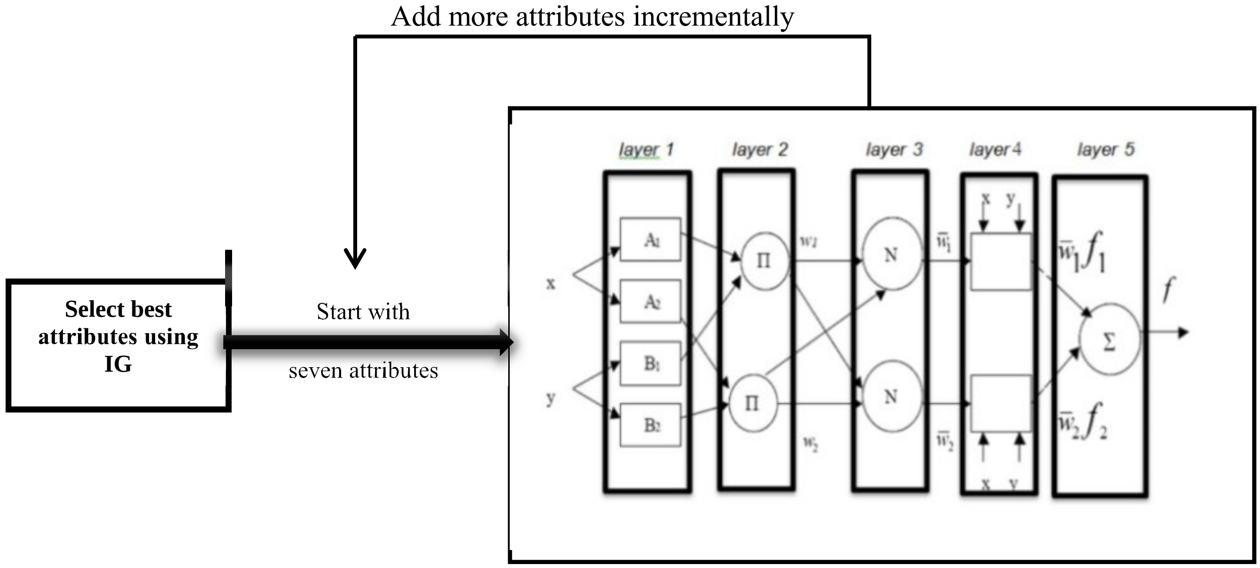

Figure 4. 7 ANFIS architecture

Here the ANFIS layers will be demonstrated in more detail and the customization of it for this research is shown clearly:

Layer 1: The fuzzification process is applied in this layer where each node produces the membership degree for each input label. The linguistic input value A_{1i} is the input to node *i,* and the membership function μ(A) is the output $O_{i,1}$of node i:

4.3 $O1,i= \mu_{Ai}(x)$

The membership function for the linguistic value A might be used any type of parameterized μ function (such as triangle, sigmoid and trapezoidal). In this thesis we apply Gaussian distribution to make the result more accurate. The membership function μ(X):

4.4 ???? ? $\frac{?}{?? ?\frac{???}{?}?^{??}}$

where *a*, *b*, *c* is the parameter set. The bell shapes change according to these values. Parameters in that layer are called *premise parameters*.

Layer 2: In this layer the nodes manipulate each rule using any T-norm such as min or prod operator. In this thesis we use product operation.

4.5 ? ? ? ? ? ?? ? ? ? ? ?? ?

Similarly, we compute the $w_2, w_3, \ldots, w_n$.

Layer 3: In this layer the normalized firing strength for each rule is calculated by each node. The result is a normalized firing strength.

4.6 ??? ? $\frac{?\ ?}{?\ ?\ ?\ ?\ ?}$

Similarly, we compute the w_2. So the output $O_{3,l}$ is the normalized firing strength .

$O_{3,1}$=???

Layer 4: The output of the nodes in this layer is produced from the multiply of normalized firing strength for each corresponding fired rule.

4.7 ? ? ?? ? ??? ??

where

4.8 ?? ? ? ? ? ? ? ? ? ? ? ? ??

Parameters p,q and r in this layer are called *consequent parameters*.

Layer 5: The sum of all outputs generated from the nodes in layer 4 is the output of this layer in this layer a single node that sums the overall output all nodes in layer 4. The output $O_{6,1}$ is

4.9 ? ? ?? ? ??? ?? ? ??? ??

The performance of ANFIS mainly depends of the membership distribution function used, as well as the coefficients of the consequence function p, q and r.

The ANIFS algorithm can be tuning or training or in other word, the coefficients of the function can be optimized using any optimization tool like Genetic algorithm or backpropagation algorithm. In this thesis we use Branch and bound algorithm for tuning purposes.

4.4 Chapter Summary

In this chapter, the methodology of this thesis has been demonstrated. In this thesis the ANFIS algorithm is implemented on the training dataset that was collected from the tracer studies conducted in three Jordanian universities and it is compared with a set of DM classifiers. The ANFIS algorithm is a hybrid algorithm that combines the fuzzy logic inference system and artificial neural network techniques. ANFIS is an algorithm that can handle the imprecise input and result in crisp output. ANFIS achieves superior results in different fields because it deals with fuzzy inputs, unlike others classification techniques that require crisp inputs, thus in this thesis we apply ANFIS on an educational dataset to predict the employability status after graduation of the universities' students. The number of attributes in the collected dataset is big to use with ANFIS, so in this thesis we use the incremental approach of the number of attributes in order to build the classifier (model). The incremental approach is performed by increasing the number of attributes by one, then build the classifier, then perform the testing process using the 10-fold cross validation. ANFIS is performed 50 times to build each classifier. The computation time is a very important factor so the experts opinion and other attribute selection techniques are considered to select the best attributes in this thesis.

5 Chapter 5. RESULTS, ANALYSIS, and DISCUSSION

In this chapter we demonstrate the result of applying the mentioned steps in our methodology using the dataset that has been described in the data collection phase. The classification task in the current research study as mentioned previously is to construct a graduate employability model in order to know the employment status (employed, unemployed or other) for graduates according to the dataset collected from the tracer study. This chapter describes how we carry out several experimental studies to test the accuracy of the ANFIS classifier. The testing process includes comparing the results of building the employability classifier of ANFIS against the results of most popular classification techniques, such as decision tree, SVM, Naïve Bayes, and Multilayer Perceptron (MLP). We will use several measures to evaluate the performance such as KAPPA, RMSE and accuracy measures, and also, we will evaluate the efficiency by measuring the execution time of building the classifiers.

This chapter is divided into two parts. Part (5.1) illustrates the techniques and measures that we will use in evaluating the performance of classifiers. This part includes five sections and they are the stages of applying our experiments. Part (5.2) discusses the final outcomes of our experiments.

5.1 Performance evaluation

There are two stages in the classification task, consisting of training and testing phases, of which the training phase was illustrated previously. The testing step is illustrated by determining the testing dataset for estimating the accuracy of the prediction. As was discussed previously, there are many popular testing strategies, but the most commonly used methods are the following four methods, that are used in WEKA and MATLAB:

- Training set: this method is based on selecting a testing dataset from the training dataset randomly, or we can use whole training dataset as a testing dataset after creating the model. This method almost always causes overfitting problems and makes the accuracy result unreliable, so this method rarely is rarely used for testing purposes.

- Supplied test set: this method is based on determining the training dataset and testing dataset separately. This method almost always gives reliable results about the accuracy of the prediction model because the testing data is not the same as the training data.

- Percentage split: in this method the dataset is partitioned into two portions in which the first part is for training and the other for testing. Researchers adopt different percentages according the type of dataset and application; some of them use 70% for training, 30% for testing, and others use 50%, 50%.

- Cross-validation: in the cross-validation method, the dataset is divided into a number of portions that have equal size. One of the portions is used for testing the classifier, while the remaining portions are dedicated for training the classifier. The process is re-implemented k times (the folds); each time a different partition is used for testing (validation) and others for training. The results produced from the K testing process are combined to compute the final estimation. A 10-fold class validation is commonly used to get the perfect for measuring error and reduce the overfitting issues (Faber and Rajkó, 2007). It has been widely applied on numerous datasets with different classification algorithms. The following are some advantages and disadvantages of cross-validation technique:

- **Advantages**:
  1- It provides us with the ability to use and test all of the available dataset, instead of using part for training and another part for testing. So we are able to make predictions on all of our data.
  2- Using cross-validation gives us more metrics and important conclusions both about our algorithm and our data.
  3- Cross-validation is very useful in multi-layer models such as neural network classifiers. We can train those different layers because we use the back-propagation technique. Each layer calculates its error and passes it back to the previous layer.
  4- Cross-validation helps in tuning the optimal value of parameters to raise the efficiency of the algorithm.
- **Disadvantages**:
  1- Using cross-validation leads to a long training time; with cross-validation we have to train our model on multiple training sets, which needs more time.

2- Cross-validation doesn’t really perform well with sequential data.
3- Cross-validation needs expensive computation in terms of processing power required.

Generalizability is an ability of our model, after the training process, to recognize new data and make reliable predictions. A model’s ability to generalize is key to the success of a classification model. One of the main advantages of neural networks is their ability to generalize. This means that a trained model could classify data from the same class as the training data that it has not seen before. In all classification jobs, we build our models on the available data and expect they will perfectly generalize to new data. To reach the best generalizability point, our dataset should be split into three parts:

- The training set is used to train the ANFIS system. The error of this dataset is reduced during training.
- The validation set is used to find the performance of a NN on patterns that are not trained during learning.
- A test set for determining the overall performance of a NN.

Training should stop because the best generalizability point has reached. Regularization technique performs minor modifications to the classification algorithm such that the model generalizes better. This enhances the model’s performance on the unseen data as well. In our experiments, we will use an early stopping technique which is a cross-validation strategy where we hold one part of the training set as the validation set. Then during the experiments when we find that the performance on the validation set is getting worse, we stop the training on the model.

In this thesis the 10-fold cross validation process is implemented to evaluate the accuracy of the model that was built using the neuro-fuzzy inference system (ANFIS). The ANFIS algorithm was applied 50 times for training the neural network architecture in order to reach the accurate model.
After building the above classifier, we needed to evaluate the performance of this model. We used the same data set to train several classifiers, such as decision tree, support vector machine SVM, Naive Bayes, and multilayer perceptron (MLP), in order to compare them with our classifier. All classifiers were trained and tested using a 10-fold cross-validation method to avoid overfitting. Since we used 10-fold cross technique, all the observation of our data sample was used for training and testing. The accuracy of the classifier was

specified by comparing the predicted class labels with the original consequence in the testing dataset. In this study we used several measures to evaluate our classification model. These measures include confusion matrix to compute the accuracy, Precision, Recall, False-Positive rate and recall. RMSE is another important measure used to test the efficiency of our model. Kappa measure is also used. Furthermore, as mentioned previously we adopted a technique based on attributes, an incremental approach to evaluate both the computational cost and efficiency of our classification model. We measured the computational cost by calculating the execution time for each classifier.

Our attributes' incremental approach included conducting several experiments by gradually increasing the selected attributes to be implemented in the classification models. We divided performance testing into phases according to the number of attributes.

We have used attributes selection methods such as information gaining technique to choose attributes having large value and to eliminate the less impotence attributes. But in other hand, we have decided to adopt another approach to apply all 22 attributes in our experiments to find if there are useful dependencies between those attributes and also to find another good weighted attributes. Our approach will start applying 4 attributes and then increasing the number of attributes gradually to reach 22 attributes. After adding each attribute, we will analyse the results and study the performance and accuracy of our model when adding the new attribute. This approach will give us more understanding of the impact of each individual attribute. By applying this approach, we will not only relay on attributes selection method to measure the importance of every single attribute. We will start our experiments by applying the most importance seven attributes using information gaining methods as follows:

5.1.1 Applying Seven Attributes Using Selection Methods:

After ranking the attributes based on their degree of importance using attributes selection technique such as information gaining technique, we have only selected seven attributes in this stage. These selected attributes have been applied on our classification and also have been applied on several classifiers, such as; decision tree, SVM, Naive Bayes, and multilayer perceptron (MLP).

First, we start evaluating these selected attributes by creating a confusion matrix for each classifier as follows. The comparison is demonstrated in the following tables, in which the confusion matrix is represented for each classifier. The most seven attributes according to selection are:

1- University.
2- Math_Skills.
3- Major.
4- Gender.
5- GPA.
6- English_skills.
7- Programming_skils.

Class	**Employed = yes**	**Employed = no**	**Total**
Employed = yes	282	18	300
Employed = no	23	377	400
Total	305	395	700

Table 5. 1 Confusion matrix of ANFIS classifier using the most seven ranking attributes

Accuracy for ANFIS classifier = 659/700 = 94%

Class	**Employed = yes**	**Employed = no**	**Total**
Employed = yes	268	32	300
Employed = no	45	355	400
Total	313	387	700

Table 5. 2 Confusion matrix of Decision Tree classifier using the most seven ranking attributes

Accuracy for Decision Tree = 623/700 = 89%

Class	**Employed = yes**	**Employed = no**	**Total**
Employed = yes	257	43	300
Employed = no	39	361	400
Total	296	404	700

Table 5. 3 Confusion matrix of SVM classifier using the most seven ranking attributes

Accuracy for SVM classifier = 618/700 = 88%

Class	**Employed = yes**	**Employed = no**	**Total**
Employed = yes	234	66	300
Employed = no	68	332	400
Total	302	398	700

Table 5. 4 Confusion matrix of Naive Bayes classifier using the most seven ranking attributes

Accuracy for Naïve Bayes classifier = 566/700 = 80%

Class	**Employed = yes**	**Employed = no**	**Total**
Employed = yes	276	24	300

Employed = no	37	363	400
Total	313	387	700

Table 5. 5 Confusion matrix of MLP classifier using the most seven ranking attributes

Accuracy for MLP classifier = 639/700 = 91%

The above tables show again that the accuracy of ANFIS classifier has gained the highest accuracy of 94%. The MLP classifier comes in second place with an accuracy of 91 %. While decision tree remains at third place and has got an accuracy of 89 %, Fourth place goes for SVM classifier with an accuracy of 88 %. Naive Bayes has got the lowest accuracy with an accuracy of 80%. As shown in Figure 5.1, ANFIS and MPL classifiers prove their efficiency in achieving the best accuracy over the other classifiers.

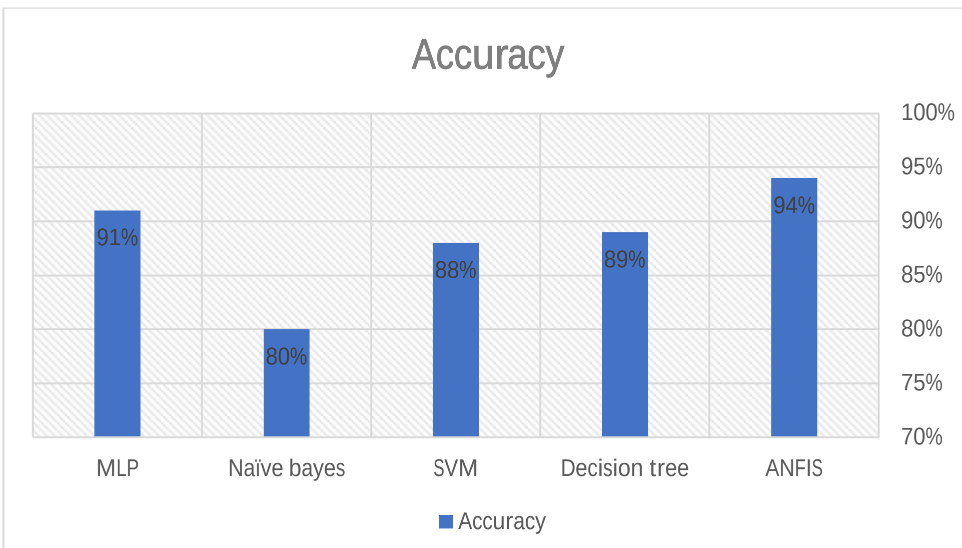

Figure 5. 1 Efficiency comparison of classifiers with the most seven ranking attributes

Also, from the above confusion matrix, we can derive: Recall, False Positive-Rate, Precision, and F-score values to conduct another performance comparison between classifiers. The measurements of the classifiers' performance can be determined according to statistical measures such as Precision, Recall, False Positive Rate, and F-score. The high values for previous measures indicates high performance, and vice versa.

Classifier	Class label	Recall (%)	False-Positive rate (%)	Prec. (%)	F-score (%)
ANFIS	Employed	98.4	1.4	98.8	99.5
	Not-employed	98.1	2.8	98.4	98.4
Decision tree	Employed	95.6	3.6	96.1	98.6
	Not-employed	94.3	6.5	95.5	96.4
SVM	Employed	93.6	7.2	95.1	95.3
	Not-employed	92.7	4.5	93.4	94.3
Naïve Bayes	Employed	90.2	3.8	91.4	92.7
	Not-employed	89.3	5.7	90.3	92.5
MLP	Employed	96.8	3.1	97.5	97.8
	Not-employed	96.2	3.4	96.4	97.5

Table 5. 6 Detailed accuracy by each class with the most seven ranking attributes.

Table 5.6 shows that the ANFIS classifier achieved the highest values for Recall, Precision, Recall, F-score; and lowest value for False-Positive rate. These values have scored in predicting both classes (Employed, Not-employed). On the other hand, lowest values for Recall, Precision, Recall, F-score; and highest value for False-Positive rate, which indicates again the superiority of ANFIS in predicting efficiency.

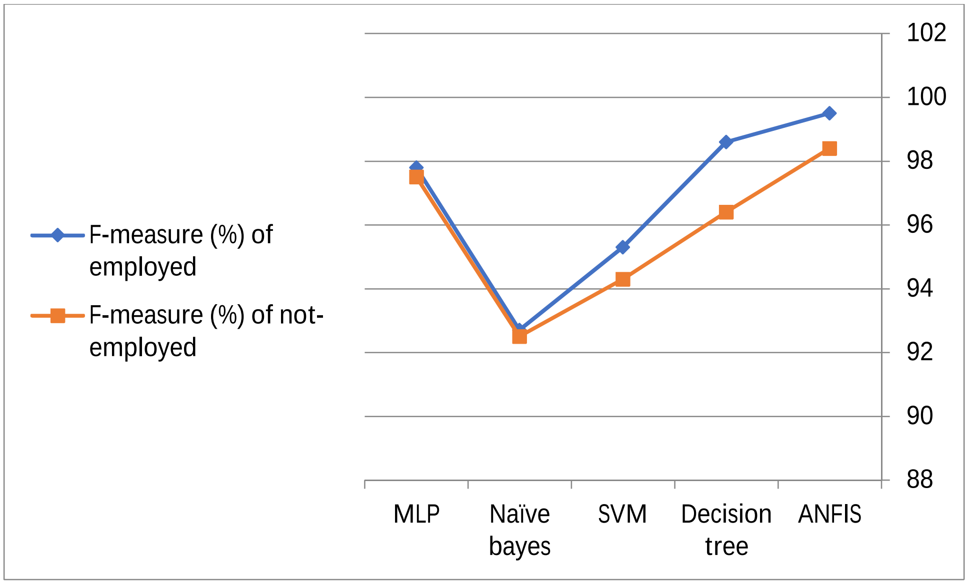

Figure 5. 2 F-score of employed and not employed classes for each classifier with the most seven ranking attributes

As shown in Figure 5.2, F-score values of the employed class for all classifiers are higher than the F-score values of the not-employed class. Thus, all classifiers have gained a higher prediction ratio for the employed class than prediction ratio for the not-employed class. Also note from the above figure that the F-score values are almost close to each other.

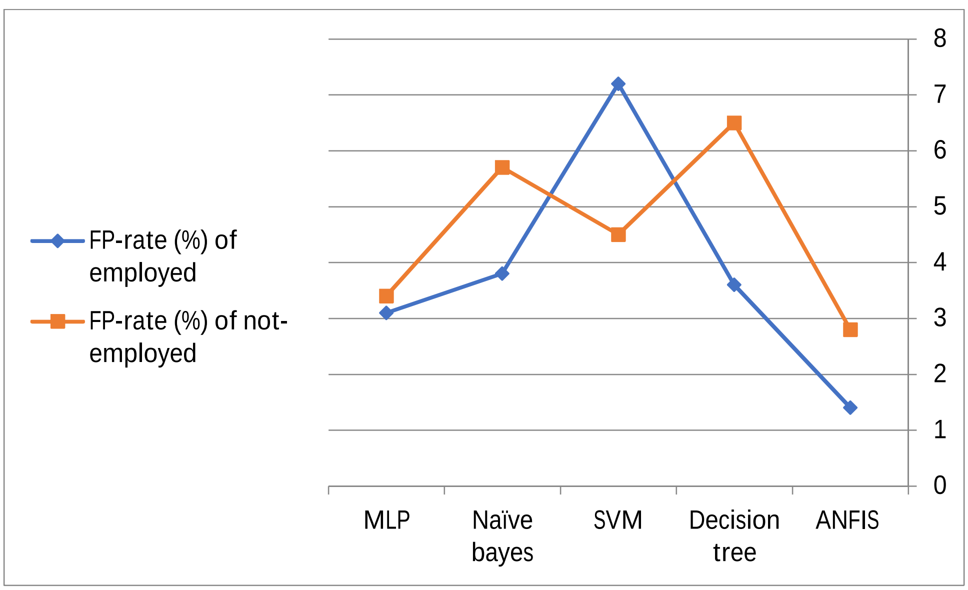

Figure 5. 3 False-Positive rate of employed and not employed classe for each classifier with the most seven ranking attributes.

As shown in Figure 5.3 the False-Positive rate values of the employed class for all classifier are lower than False-Positive rate values of the not-employed class, this indicates great ratio of prediction for the employed class when applying seven selected attributes.

To complete the evaluation process of the above classifiers, the performance of classifier determined according two statistical measures that are the Root Mean Square Error and Kappa measures. In our experiment, we applied a well-known measure which is the Root Mean Square Error (RMSE) where the value of its must be in minimum level to say the classifier achieve good performance. Kappa statistic is used to compare between two raters that are in this thesis the classified data and the testing data

Classifier	RMSE	Kappa statistic
ANFIS	0.1489	0.9364
Decision tree	0.1912	0.8847
SVM	0.2045	0.8713
Naïve Bayes	0.2251	0.8437
MLP	0.1634	0.9172

Table 5. 7 RMSE and Kappa statistic values for each classifier applying the most seven ranking attributes

As shown in Table 5.7 ANFIS classifier again proves its highest efficiency of performance, and this because it has got the lowest RMSE value of 0.1489 and a kappa statistic value of 0.9364; followed by MLP which achieved 0.1634 of RMSE value and a kappa statistic value of 0.9172; decision tree classifier comes out in third place with an RMSE value of 0.1912 and a kappa statistic value of 0.8847; followed by SVM classifier that has got 0.2045 of RMSE value and with a kappa statistic value of 0.8713; Naïve Bayes classifier stands last with the highest RMSE value (0.22451) and the lowest kappa statistic value (0.9172). Figure 5.4 shows an efficiency comparison of classifiers according to RMSE and Kappa statistic values applying seven selected attributes.

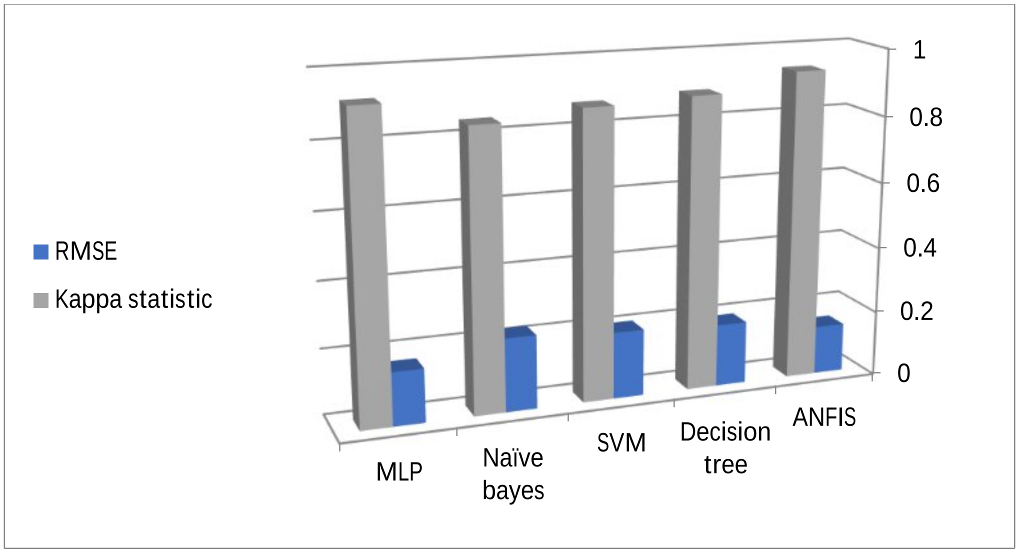

Figure 5. 4 An efficiency comparison of classifiers according to RMSE and Kappa statistic values with applying the most seven ranking attributes

Table 5.8 shows the execution time for each classifier, which indicates the computational cost in building classification models implementing the seven selected attributes.

Classifier	Execution Time (Secs)
ANFIS	5.53
Decision tree	2.91
SVM	3.12
Naïve Bayes	3.96
MLP	5.32

Table 5. 8 Execution time applying the most seven ranking attributes

As shown in table 5.8 ANFIS classifier has got the highest value of execution time with 5.53s, followed by MLP classifier with 5.32s. This result proves that both of MLP and ANFIS classifiers have time complexity problem when applying the seven selected attributes. Decision tree has gained the lowest value of execution time of 2.91s, which indicate again its best time efficiency over the other classifiers when applying the seven selected attributes. SVM comes in the second place with value of 3.12s. The third place goes for Naïve Bayes with time execution of 3.96s.

Now, let us start conducting the second stage of our experiments, which is applying all 22 attributes by increasing the number of selected attributes gradually. We will start applying eleven attributes as follows:

5.1.2 Applying Eleven Attributes:

In this stage, we increased the number of attributes to eleven. We added "English_skills" attribute to the selected attributes list.

After applying the classification models with the below attributes, we created a confusion matrix for each classification model. The following tables (Table 5.9 – Table 5.10) represent confusion matrixes for each classifier with implementing eleven attributes. The selected attributes in this stage are:

1- University.
2- Province.
3- Age.
4- Gender.
5- GPA.
6- Social strata.
7- Math_Skills.
8- Time to work.
9- Type_ of _ degree.
10- Educational_ degree.
11- English_skills.

Class	**Employed = yes**	**Employed = no**	**Total**

Employed = yes	245	55	300
Employed = no	50	350	400
Total	295	405	700

Table 5. 9 Confusion matrix of ANFIS eleven attibutes classifier

Accuracy for ANFIS classifier = 595 / 700 = 85 %

Class	**Employed = yes**	**Employed = no**	**Total**
Employed = yes	223	77	300
Employed = no	70	330	400
Total	283	407	700

Table 5. 10 Confusion matrix of Decision Tree eleven attibutes classifier

Accuracy for Decision Tree = 553/ 700 = 79 %

Class	**Employed = yes**	**Employed = no**	**Total**
Employed = yes	210	90	300
Employed = no	81	319	400

Total	291	409	700

Table 5. 11 Confusion matrix of SVM classifier with eleven attributes

Accuracy for SVM classifier = 529 / 700 = 75 %

Class	**Employed = yes**	**Employed = no**	**Total**
Employed = yes	187	113	300
Employed = no	82	308	400
Total	267	421	700

Table 5. 12 Confusion matrix of Naive Bayes classifier with eleven attibutes

Accuracy for Naïve Bayes classifier = 495/ 700 = 69.8 %

Class	**Employed = yes**	**Employed = no**	**Total**
Employed = yes	236	64	300
Employed = no	63	337	400
Total	299	408	700

Table 5. 13 Confusion matrix of MLP classifier with eleven attibutes

Accuracy for MLP classifier = 573 / 700 = 81 %

The above tables show that the accuracy of the ANFIS classifier has increased when "English_skills" added to selected attributes list, and it has achieved the highest accuracy with value of 85%. The MLP classifier comes in second place with an accuracy of 81 %. Decision tree came in third place and has got an accuracy of 79 %. Fourth place goes to SVM classifier with an accuracy of 74 %. Naive Bayes has got the lowest accuracy with an accuracy of 69.8%. As shown (Figure 5.5), ANFIS and MLP classifiers keep their superiority in achieving the best accuracy over the other classifiers. Based on the previous accuracy enhancement, we found that "English_skills" attribute has an effective role in improving the accuracy of the prediction.

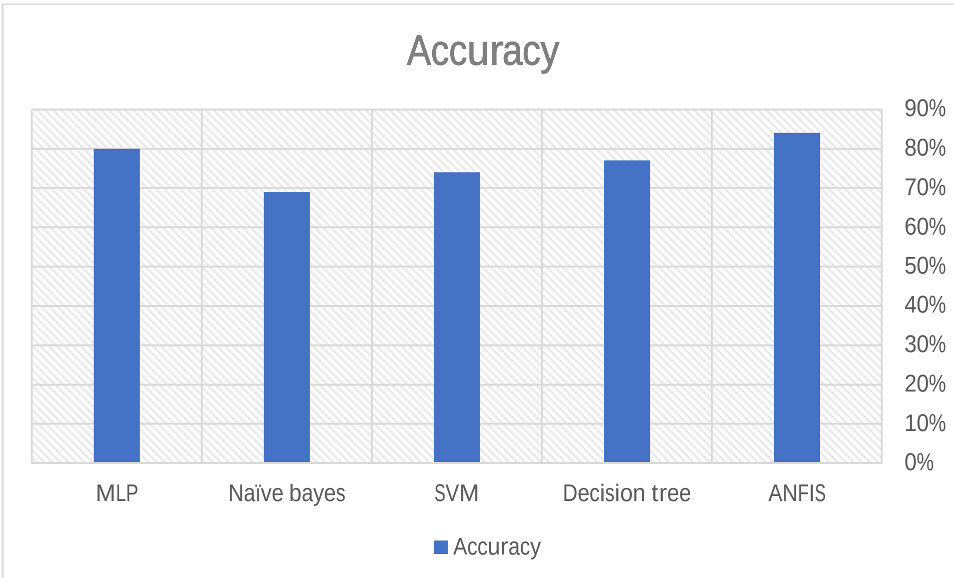

Figure 5. 5 Efficiency comparison of classifiers with eleven attributes

Table (5.14) shows Recall, False-Positive rate, Precision, and F-score values for applying the prediction classifiers with eleven attributes. The measure values in the table below are for both classes "Employed" and "Not-employed".

Classifier	**Class label**	**Recall (%)**	**False-Positive rate (%)**	**Prec. (%)**	**F-score (%)**
ANFIS	Employed	86.4	1.6	87.2	87.9
	Not-employed	86.7	2.2	85.7	87.5
	Employed	82.4	2.8	82.8	82.5

Decision tree	Not-employed	81.6	3.1	81.7	82.6
SVM	Employed	78.4	4.2	77.7	77.8
	Not-employed	78.1	4.4	78.2	76.7
Naïve Bayes	Employed	74.2	5.2	73.7	74.2
	Not-employed	73.7	4.1	73.4	72.9
MLP	Employed	84.2	2.1	84.7	83.9
	Not-employed	83.9	2.3	84.2	83.8

Table 5. 14 **Detailed accuracy by each class for eleven attributes classifiers.**

As shown in Table (5.14), after applying eleven attributes to all prediction classifiers, the ANFIS classifier gained the highest values for Recall, Precision, Recall, F-score; and lowest value for False-Positive rate. These predicting values for both classes (Employed, Not-employed) has enhanced when added "English_skills" attribute to selected attributes list.

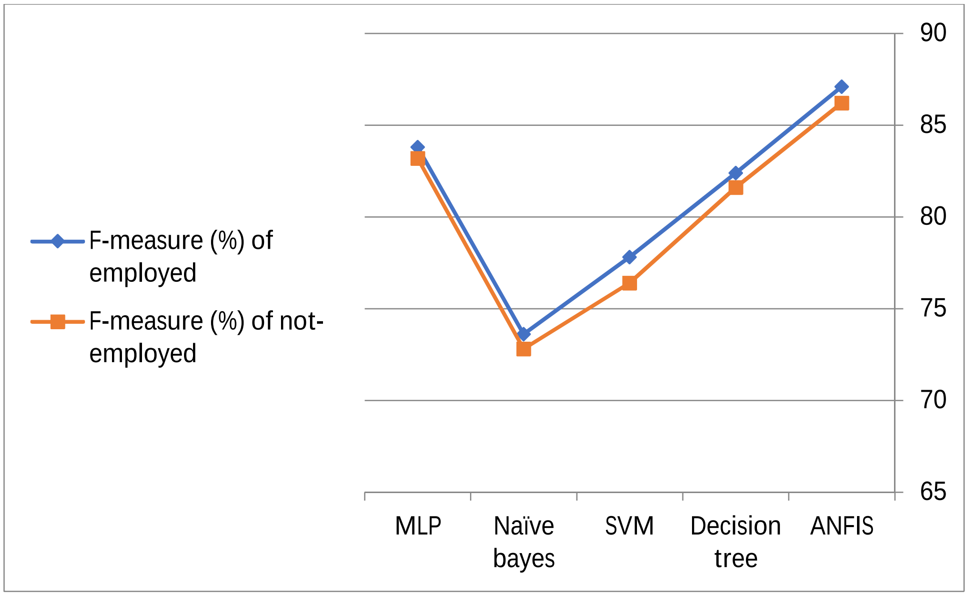

Figure 5. 6 F-score of employed and not employed classes for each classifier with applying eleven attributes

As shown in Figure (5.6) the prediction ratio enhanced when adding "English_skills" to the selected attributes list, which indicates good weight of this new added attribute.

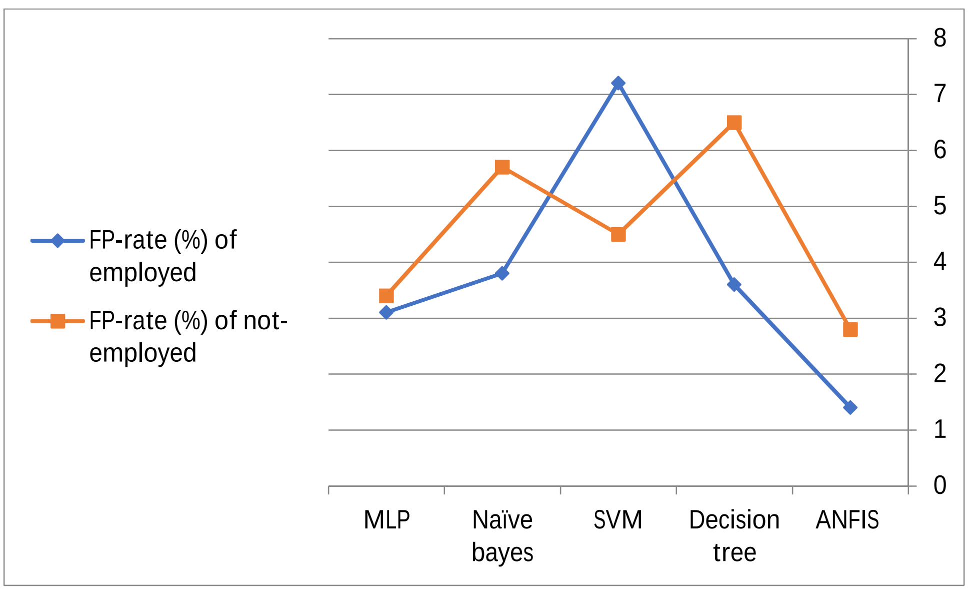

Figure 5. 7 False-Positive rate of employed and not employed classes for each classifier with eleven attributes

As shown in Figure (5.7) the False-Positive rate values of the employed class for all classifier are lower than False-Positive rate values of not-employed class. This indicates again, the excellent ratio of prediction for the “employed” class when applying eleven attributes. Furthermore, the previous result proves that the accuracy of the ANFIS classifier is the best over other classifiers. Still, the continuous in error reduction indicates better efficiency when increasing the number of attributes used in building the classification.

Table (5.15) shows the values of RMSE and Kappa measures to estimate the efficiency of the prediction classifier when applying eleven attributes including "English_skills".

Classifier	**RMSE**	**Kappa statistic**
ANFIS	0.2003	0.8746
Decision tree	0.2665	0.8237
SVM	0.2978	0.7832
Naïve Bayes	0.3675	0.7564
MLP	0.2163	0.8489

Table 5. 15 RMSE and Kappa statistic values for each classifier applying eleven attributes

As shown in Table (5.15), Kappa and RMSE measure values have improved when adding "English_skills" attribute. Figure (5.8) shows an efficiency comparison of classifiers according to RMSE and Kappa statistic values when applying eleven attributes.

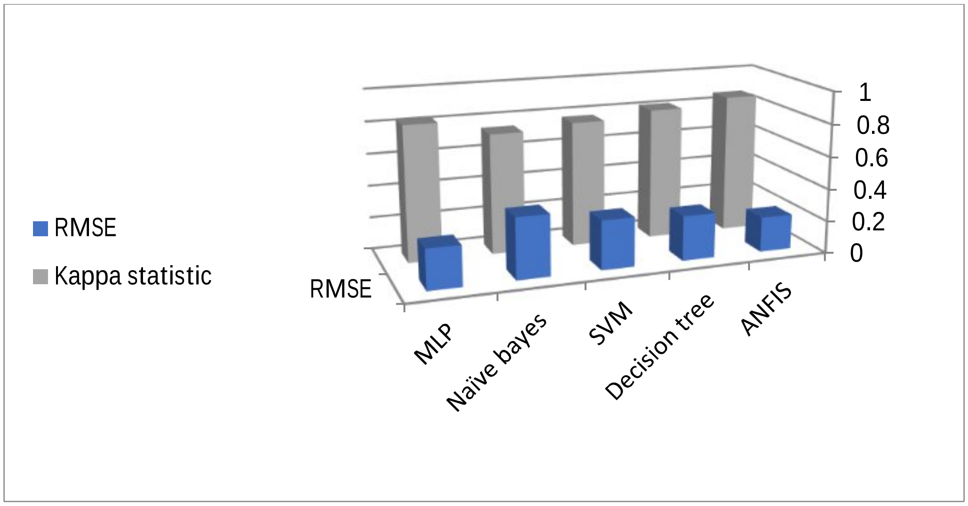

Figure 5. 8 An efficiency comparison of classifiers according to RMSE and Kappa statistic values when applying eleven attributes

As shown in Figure (5.8), the efficiency performance of all classifiers has enhanced the same according to RMSE and Kappa measures. This confirms the effective of "English_skills" attribute in increasing the accuracy ratio.

Table (5.16) shows the execution time for each classifier, which indicates the computational cost in building classification models implementing eleven attributes.

Classifier	Execution Time (Secs)
ANFIS	5.57
Decision tree	2.95
SVM	3.16
Naïve Bayes	4.00
MLP	6.36

Table 5. 16 Execution time for eleven attributes classifiers

As shown from Table (5.16), execution time for implementing ANFIS classifier when using eleven attributes has increased. As shown in the above table, MLP classifier has got the highest value of execution time with 6.36s, followed by ANFIS with 5.57s. The values of execution time for all classifiers have also increased, but the increase in time was acceptable. Decision tree has gained the lowest value of execution time of 2.95s, which indicates again its best time efficiency over the other classifiers when increasing the number of attributes. SVM comes in the second place with a value of 3.16s. The third place goes to Naïve Bayes with time execution of 4.00s.

5.1.3 Applying Fifteen Attributes:

In this stage, we increased the number of attributes to fifteen. We added "Talent" attribute to the selected attributes list.

Tables (Table 5.97 – Table 5.101) represent confusion matrixes for each prediction classifier using fifteen attributes. The selected attributes in this stage are:

1- University.
2- Province.
3- Age.
4- Gender.
5- GPA.
6- Social strata.
7- Math_Skills.
8- Time to work.
9- Type_ of _ degree.
10- Educational_ degree.
11- English_skills.
12- Completion _ period.
13- Interpersonal_skills.
14- High_school_grade.
15- Talent.

Class	Employed = yes	Employed = no	Total
Employed = yes	255	45	300
Employed = no	40	360	400
Total	295	405	700

Table 5. 17 Confusion matrix of ANFIS fifteen attibutes classifier

Accuracy for ANFIS classifier = 605 / 700 = 86 %

Class	Employed = yes	Employed = no	Total
Employed = yes	233	67	300
Employed = no	60	340	400
Total	283	407	700

Table 5. 18 Confusion matrix of Decision Tree fifteen attibutes classifier

Accuracy for Decision Tree = 563/ 700 = 80 %

Class	Employed = yes	Employed = no	Total
Employed = yes	220	80	300

Employed = no	71	329	400
Total	291	409	700

Table 5. 19 Confusion matrix of SVM classifier with fifteen attributes

Accuracy for SVM classifier = 539 / 700 = 77 %

Class	**Employed = yes**	**Employed = no**	**Total**
Employed = yes	197	103	300
Employed = no	72	318	400
Total	267	421	700

Table 5. 20 Confusion matrix of Naive Bayes classifier with fifteen attibutes

Accuracy for Naïve Bayes classifier = 505/ 700 = 72 %

Class	**Employed = yes**	**Employed = no**	**Total**
Employed = yes	246	54	300
Employed = no	53	347	400
Total	299	408	700

Table 5. 21 Confusion matrix of MLP classifier with fifteen attibutes

Accuracy for MLP classifier = 583 / 700 = 83 %

The above tables show that the accuracy ratio of applying the ANFIS classifier has not improved when "Talent" attribute added to selected attributes list, and it remains in the first place with accuracy ratio of 86%. The accuracy ratio of MLP classifier not affected with the same accuracy ratio of 83 %. Decision tree came in third place and has got an accuracy of 82 %. Fourth place goes to SVM classifier with an accuracy of 77 %. Naive Bayes has got the lowest accuracy with an accuracy of 72%. As shown (Figure 5. 9), ANFIS and MLP classifiers still achieved the highest accuracy over the other classifiers.

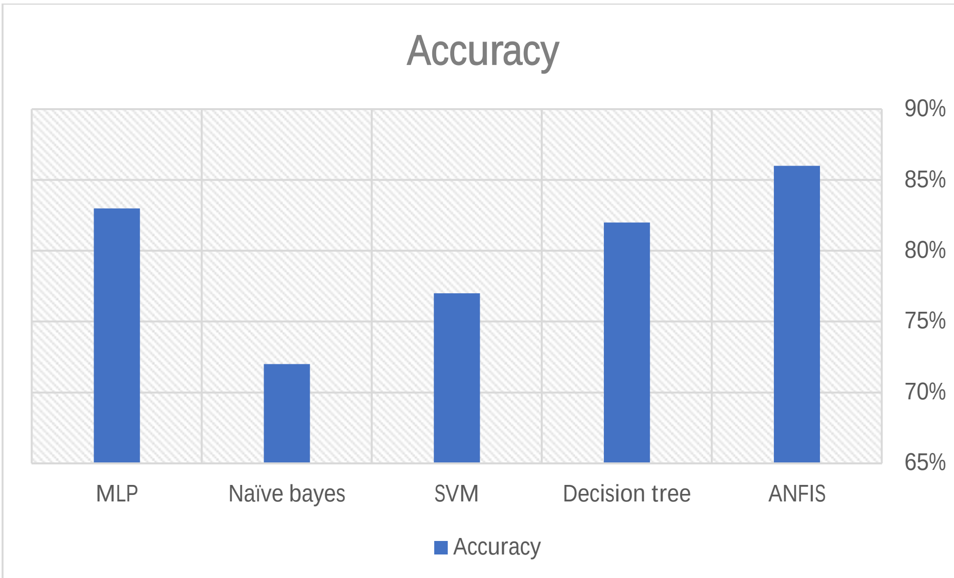

Figure 5. 9 Efficiency comparison of classifiers with fifteen attributes

Table (5.22) shows Recall, False-Positive rate, Precision, and F-score values for building the classification model with fifteen attributes. The measure values in the table below are for both classes "Employed" and "Not-employed".

Classifier	**Class label**	**Recall (%)**	**False-Positive rate (%)**	**Prec. (%)**	**F-score (%)**
ANFIS	Employed	87.3	1.4	87.9	88.7
	Not-employed	87.4	2.2	86.4	87.8
	Employed	83.4	2.6	83.7	83.3

Decision tree	Not-employed	82.8	2.9	82.5	83.6
SVM	Employed	79.4	4.2	78.6	78.5
	Not-employed	78.7	4.1	78.8	77.6
Naïve Bayes	Employed	75.8	5.1	74.5	74.8
	Not-employed	74.6	3.8	73.5	73.7
MLP	Employed	85.1	2.0	85.3	84.8
	Not-employed	84.7	2.1	85.4	84.5

Table 5. 22 Detailed accuracy by each class for fifteen attributes classifiers.

As shown in Table (5.22), after building prediction classifiers using fifteen attributes, the ANFIS classifier gained the highest values for Recall, Precision, Recall, F-score; and less value for False-Positive rate, which is the same result in previous stages. The accuracy ratio remains the same when added "Talent" attribute to selected attributes list.

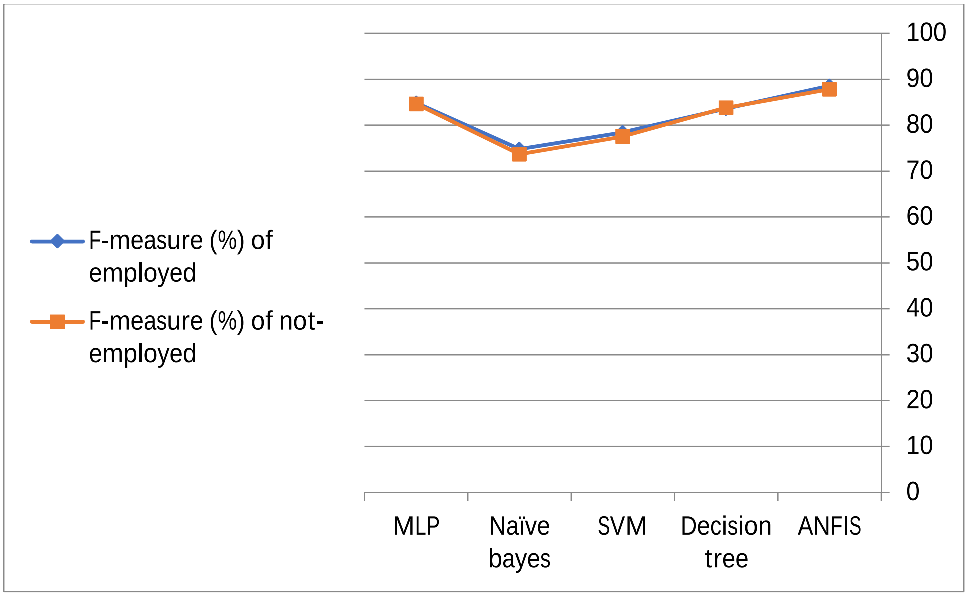

Figure 5. 10 F-score of employed and not employed classes for each classifier with applying fifteen attributes

As shown in Figure (5.10) the prediction ratio not enhanced when "Talent" attribute added to the selected attributes list, which shows less weight of this attribute.

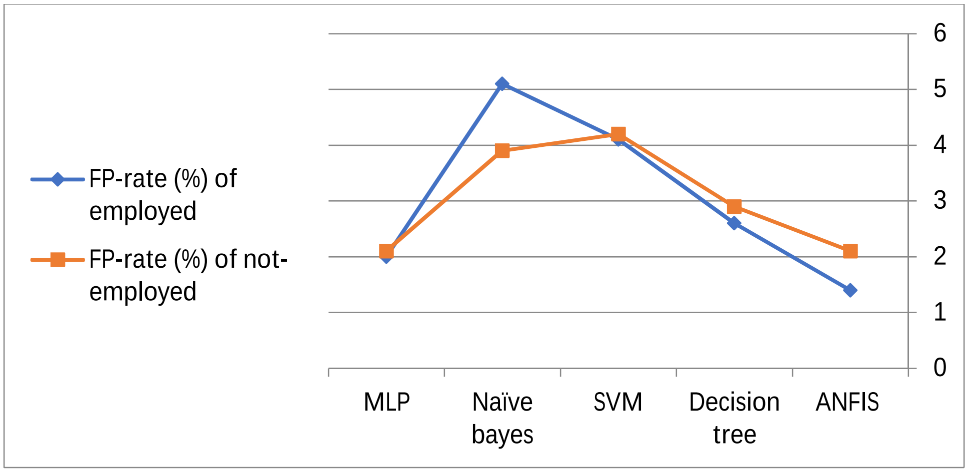

Figure 5. 11 False-Positive rate of employed and not employed classes for each classifier with fifteen attributes

As shown in Figure (5.11) the False-Positive rate values of the employed class for all classifier are lower than False-Positive rate values of not-employed class. This indicates also, the excellent ratio of prediction for the "employed" class when applying fifteen attributes. On the other hand, the previous result proves that the accuracy of the ANFIS classifier is the best over other classifiers using our dataset.

Table (5.23) shows the values of RMSE and Kappa measures to measure the efficiency of the prediction classifier when using fifteen attributes including "Talent".

Classifier	RMSE	Kappa statistic
ANFIS	0.1953	0.8859
Decision tree	0.2432	0.8358
SVM	0.2834	0.7969
Naïve Bayes	0.3441	0.7686
MLP	0.2011	0.8577

Table 5. 23 RMSE and Kappa statistic values for each classifier applying fifteen attributes

As shown in Table (5.23), Kappa and RMSE measure values have not enhanced when adding "Talent" attribute. Figure (5.12) shows an efficiency comparison of classifiers according to RMSE and Kappa statistic values when applying fifteen attributes.

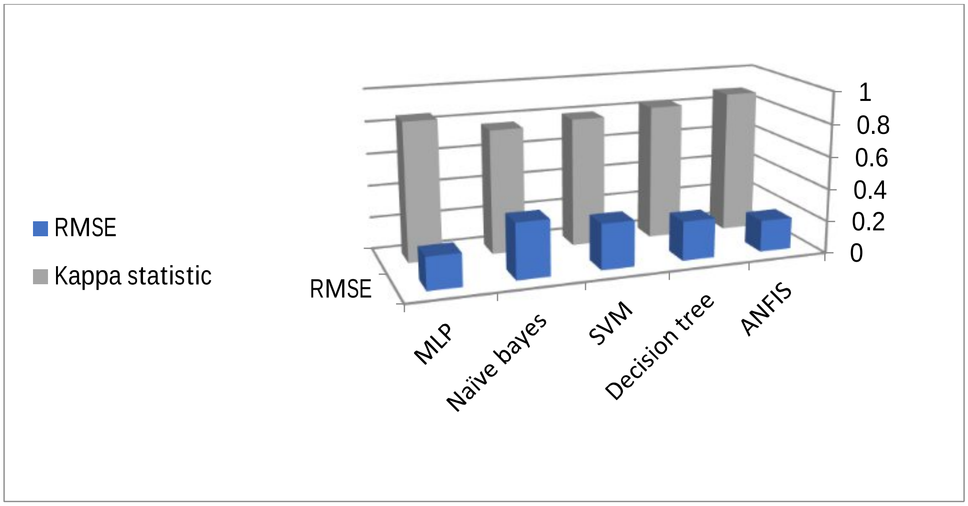

Figure 5. 12 An efficiency comparison of classifiers according to RMSE and Kappa statistic values when applying fifteen attributes

As shown in Figure (5.12), the efficiency performance of all classifiers has not improved according to RMSE and Kappa measures.

Table (5.24) shows the execution time for each classifier, which indicates the computational cost in building classification models implementing fifteen attributes.

Classifier	**Execution Time (Secs)**
ANFIS	8.46
Decision tree	4.74
SVM	4.63
Naïve Bayes	6.18
MLP	8.84

Table 5. 24 Execution time for fifteen attributes classifiers

As shown from Table (5.24), execution time for implementing ANFIS classifier when using fifteen attributes has increased. As shown in the above table, MLP classifier has got the highest value of execution time with 8.84s, then ANFIS with 8.46s. The values of execution time for all classifiers have also increased, but the rate of increase of time ratio is the same for all. Decision tree has got the less value of execution time of

4.74s, which shows again its best time efficiency over the other classifiers when increasing the number of attributes. SVM comes in the second place with a value of 4.63s. The third place goes to Naïve Bayes with time execution of 6.18s.

5.1.4 Applying Nineteen Attributes:

Now, we will increase the number of attributes to nineteen. We included "Team work skills" attribute to the selected attributes list.

Tables (Table 5.25 – Table 5.26) represent confusion matrixes for each prediction classifier with applying nineteen attributes. The selected attributes in this stage are:

1- University.
2- Province.
3- Age.
4- Gender.
5- GPA.
6- Social strata.
7- Math Skills.
8- Time to work.
9- Type of degree.
10- Educational degree.
11- English skills.
12- Completion period.
13- Interpersonal skills.
14- High school grade.
15- Talent.
16- Programming skills.
17- Tech certificate.
18- Experience.
19- Team work skills.

Class	Employed = yes	Employed = no	Total
Employed = yes	275	25	300
Employed = no	20	380	400
Total	295	405	700

Table 5. 25 Confusion matrix of ANFIS nineteen attibutes classifier

Accuracy for ANFIS classifier = 625 / 700 = 89 %

Class	Employed = yes	Employed = no	Total
Employed = yes	253	47	300
Employed = no	40	360	400
Total	283	407	700

Table 5. 26 Confusion matrix of Decision Tree nineteen attibutes classifier

Accuracy for Decision Tree = 583/ 700 = 83 %

Class	Employed = yes	Employed = no	Total
Employed = yes	240	60	300

Employed = no	51	349	400
Total	291	409	700

Table 5. 27 Confusion matrix of SVM classifier with nineteen attributes

Accuracy for SVM classifier = 559 / 700 = 79 %

Class	**Employed = yes**	**Employed = no**	**Total**
Employed = yes	217	83	300
Employed = no	52	338	400
Total	267	421	700

Table 5. 28 Confusion matrix of Naive Bayes classifier with nineteen attibutes

Accuracy for Naïve Bayes classifier = 525/ 700 = 75 %

Class	**Employed = yes**	**Employed = no**	**Total**
Employed = yes	266	34	300
Employed = no	33	367	400
Total	299	408	700

Table 5. 29 Confusion matrix of MLP classifier with nineteen attibutes

Accuracy for MLP classifier = 603 / 700 = 86 %

The tables above show that the accuracy ratio of applying the ANFIS classifier has not improved when "Team work skills" attribute added to selected attributes list, and it remains in the first place with accuracy ratio of 89%, which is the same ratio over the previous stage. The accuracy ratio of MLP classifier not affected with the same accuracy ratio of 86 %. Decision tree came in third place and has got an accuracy of 83 %. Fourth place goes to SVM classifier with an accuracy of 79 %. Naive Bayes has got the lowest accuracy with an accuracy of 75%. As shown (Figure 5.65), ANFIS and MLP classifiers keep the highest accuracy ratio over the other classifiers.

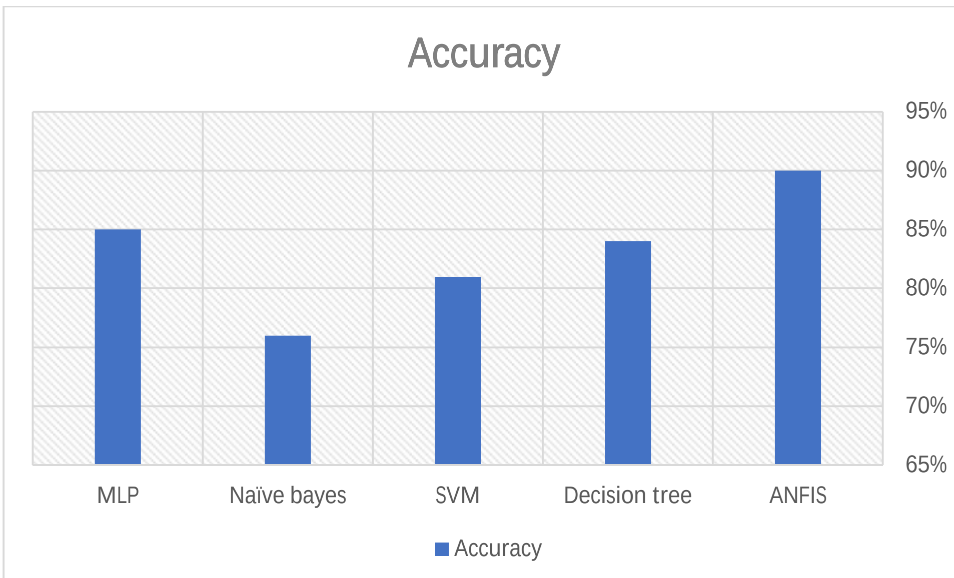

Figure 5. 13 Efficiency comparison of classifiers with nineteen attributes

Table (5.30) shows Recall, False-Positive rate, Precision, and F-score values for applying the classification model with nineteen attributes. The measure values in the following table are for both categories; "Employed" and "Not-employed".

Classifier	Class label	Recall (%)	False-Positive rate (%)	Prec. (%)	F-score (%)
ANFIS	Employed	89.3	1.2	88.7	89.9
	Not-employed	89.8	2.0	87.5	89.4

Decision tree	Employed	85.6	2.3	85.4	84.8
	Not-employed	84.4	2.5	84.3	85.1
SVM	Employed	81.3	4.0	80.6	80.4
	Not-employed	80.7	3.9	81.4	80.3
Naïve Bayes	Employed	77.5	4.9	76.8	76.6
	Not-employed	76.3	4.6	75.2	75.6
MLP	Employed	87.6	1.8	87.2	86.4
	Not-employed	86.3	2.0	87.5	86.4

Table 5. 30 Detailed accuracy by each class for nineteen attributes classifiers.

As shown in Table (5.30), after building prediction classifiers using nineteen attributes, the ANFIS classifier has obtained the largest values for Recall, Precision, Recall, F-score; and less value for False-Positive rate. The accuracy ratio has not improved when added "Team_work_skills" attribute to selected attributes list.

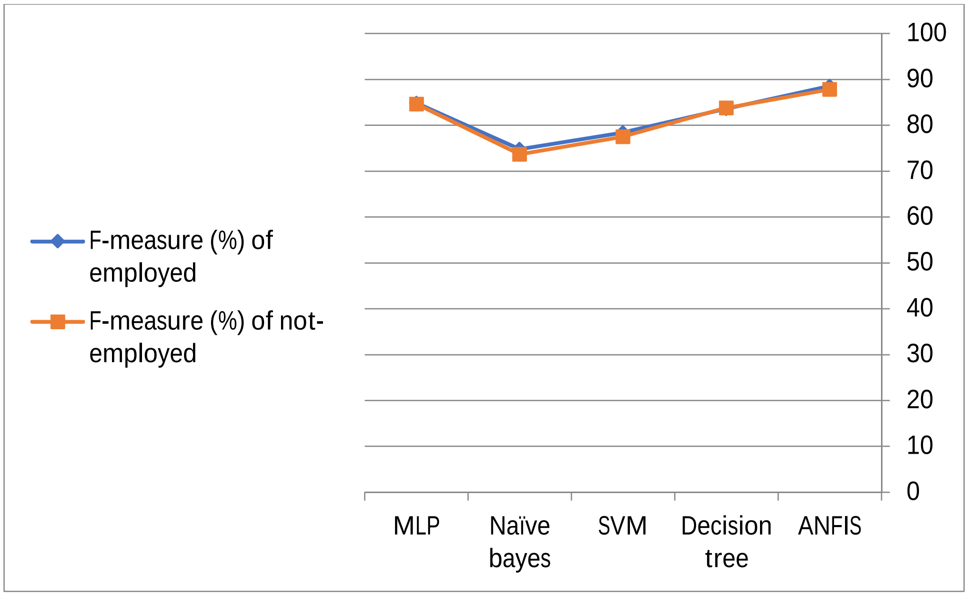

Figure 5. 14 F-score of employed and not employed classes for each classifier with applying nineteen attributes

As shown in Figure (5.14) the prediction ratio has enhanced when "Team work skills" attribute added to the selected attributes list, which shows less weight of this attribute.

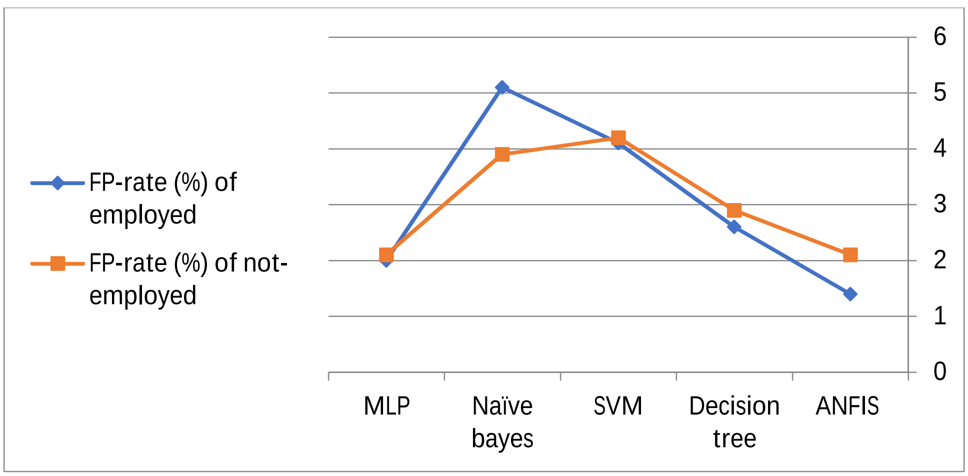

Figure 5. 15 False-Positive rate of employed and not employed classes for each classifier with nineteen attributes

As shown in Figure (5.15) the False-Positive rate values of the employed class for all classifier are lower than False-Positive rate values of not-employed class. This shows also, the excellent ratio of prediction for the "employed" class when applying nineteen attributes.

Table (5.31) shows the values of RMSE and Kappa measures to measure the efficiency of the prediction classifier when using nineteen attributes including "Team work skills" attribute.

Classifier	RMSE	Kappa statistic
ANFIS	0.1838	0.8963
Decision tree	0.2341	0.8465
SVM	0.2748	0.8054
Naïve Bayes	0.3303	0.7779
MLP	0.1965	0.8684

Table 5. 31 **RMSE and Kappa statistic values for each classifier applying nineteen attributes**

As shown in Table (5.31), Kappa and RMSE measure values have not really enhanced when adding "Team work skills" attribute. Figure (5.16) shows an efficiency comparison of classifiers according to RMSE and Kappa statistic values when applying nineteen attributes.

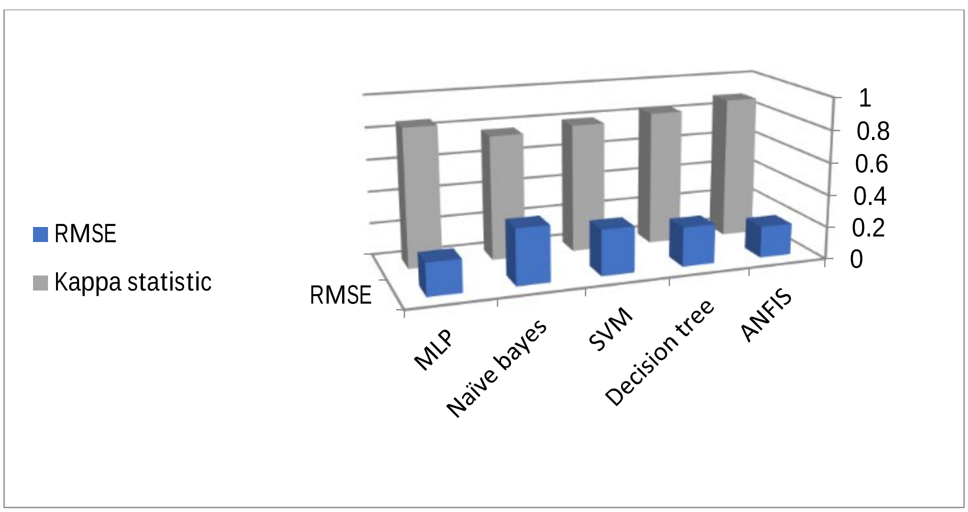

Figure 5. 16An efficiency comparison of classifiers according to RMSE and Kappa statistic values when applying nineteen attributes

As shown in Figure (5.16), the efficiency performance of all classifiers has not improved based on RMSE and Kappa measures.

Table (5.32) shows the execution time for each classifier, which shows the computational cost in building classification models using nineteen attributes.

Classifier	**Execution Time (Secs)**
ANFIS	12.67
Decision tree	5.45
SVM	5.64
Naïve Bayes	8.14
MLP	13.86

Table 5. 32 Execution time for nineteen attributes classifiers

As shown from Table (5.32), execution time for applying ANFIS classifier when using nineteen attributes has increased. As shown in the above table, MLP classifier as usual has got the highest value of execution time with 13.86s, which is considered long time in building this model, then ANFIS with 12.67s. The values of execution time for all classifiers have also increased. Decision tree has got the less value of execution time of 5.45s, which proves its best time efficiency over the other classifiers when increasing the number of attributes. SVM comes in the second place with a value of 5.64s. The third place goes to Naïve Bayes with time execution of 8.14s.

5.1.5 Applying Twenty-Two Attributes:

Finally, we reach final stage which is increasing the number of attributes to twenty-two. We included "No of application" attribute to the selected attributes list.

Tables (Table 5.153 – Table 5.157) represent confusion matrixes for each prediction classifier with applying twenty-two attributes. The selected attributes in this stage are:

1- University.
2- Province.

3- Age.
4- Gender.
5- GPA.
6- Social strata.
7- Math Skills.
8- Time to work.
9- Type of degree.
10- Educational degree.
11- English skills.
12- Completion period.
13- Interpersonal skills.
14- High school grade.
15- Talent.
16- Programming skills.
17- Tech certificate.
18- Experience.
19- Team work skills.
20- Status.
21- Major.
22- No of application.

Class	**Employed = yes**	**Employed = no**	**Total**
Employed = yes	290	10	300
Employed = no	5	395	400
Total	295	405	700

Table 5. 33 Confusion matrix of ANFIS twenty two attibutes classifier

Accuracy for ANFIS classifier = 640 / 700 = 91 %

Class	**Employed = yes**	**Employed = no**	**Total**
Employed = yes	295	5	300
Employed = no	2	398	400
Total	297	403	700

Table 5. 34 Confusion matrix of Decision Tree twenty two attibutes classifier

Accuracy for Decision Tree = 650/ 700 = 92 %

Class	**Employed = yes**	**Employed = no**	**Total**
Employed = yes	255	45	300
Employed = no	36	364	400
Total	291	409	700

Table 5. 35 Confusion matrix of SVM classifier with twenty two attributes

Accuracy for SVM classifier = 574 / 700 = 82 %

Class	**Employed = yes**	**Employed = no**	**Total**

Employed = yes	237	83	300
Employed = no	32	358	400
Total	267	421	700

Table 5. 36 Confusion matrix of Naive Bayes classifier with twenty two attibutes

Accuracy for Naïve Bayes classifier = 545/ 700 = 77 %

Class	**Employed = yes**	**Employed = no**	**Total**
Employed = yes	278	22	300
Employed = no	21	379	400
Total	299	408	700

Table 5. 37 **Confusion matrix of MLP classifier with twenty two attibutes**

Accuracy for MLP classifier = 615 / 700 = 87 %

As shown in the tables above that important change has occurred, which is Decision tree classifier beats both ANFIS and MLP classifiers with accuracy ratio of 92% when applying 22 attributes including "No of application" attribute. The ANFIS classifier comes in second place with accuracy ratio of 91. The accuracy ratio of MLP classifier not affected with the same accuracy ratio of 87 %. Fourth place goes to SVM classifier with an accuracy of 82 %. Naive Bayes has got the lowest accuracy with an accuracy of 77%. As shown (Figure 5.17), Decision tree and ANFIS classifiers have gained the highest accuracy ratio over the other classifiers.

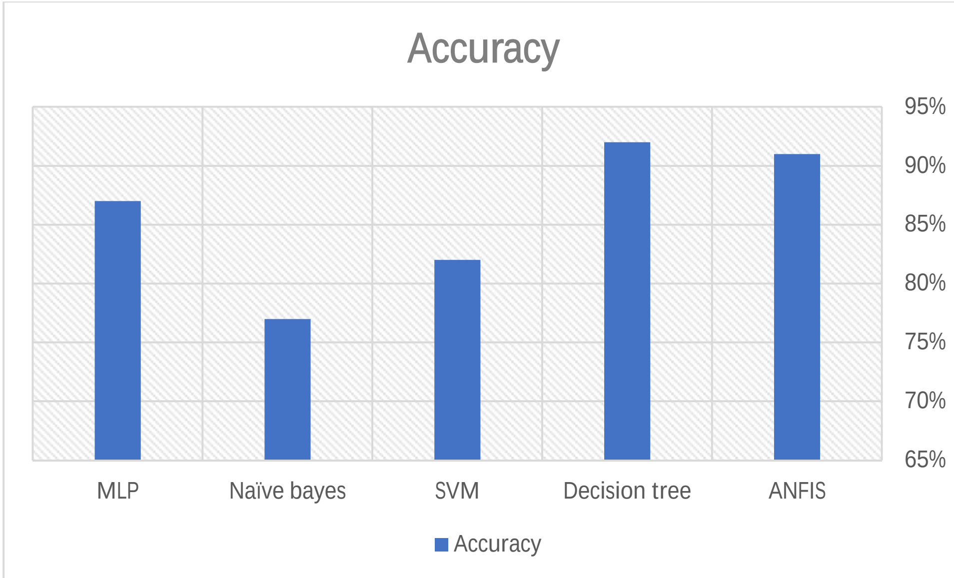

Figure 5. 17 Efficiency comparison of classifiers with twenty two attributes

Table 5.38 shows Recall, False-Positive rate, Precision, and F-score values for applying the classification model with twenty-two attributes. The measure values in the following table are for both classes; "Employed" and "Not-employed".

Classifier	Class label	Recall (%)	False-Positive rate (%)	Prec. (%)	F-score (%)
ANFIS	Employed	91.2	1.1	89.6	91.3
	Not-employed	90.9	1.8	89.3	90.1
Decision tree	Employed	92.4	0.9	92.3	92.6
	Not-employed	91.7	1.0	90.1	91.3
SVM	Employed	82.5	3.9	81.6	82.5
	Not-employed	81.7	3.8	82.5	81.4

Naïve Bayes	Employed	78.5	4.1	77.7	77.5
	Not-employed	77.4	4.2	76.1	76.5
MLP	Employed	88.6	1.7	88.1	88.6
	Not-employed	87.8	2.0	87.9	87.3

Table 5. 38 Detailed accuracy by each class for twenty two attributes classifiers.

As shown in Table (5.38), after building prediction classifiers using twenty-two attributes, the Decision tree classifier has obtained the largest values for Recall, Precision, Recall, F-score; and less value for False-Positive rate. The accuracy ratio has improved when added "No_of_application" attribute to selected attributes list.

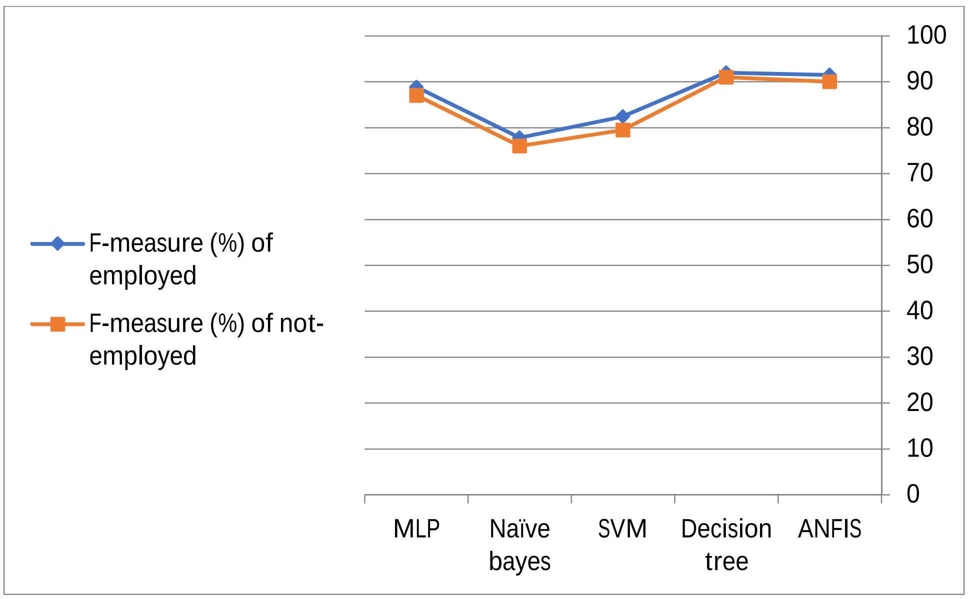

Figure 5. 18 F-score of employed and not employed classes for each classifier with applying twenty two attributes

As shown in Figure (5.18) the prediction ratio has enhanced when "No of application" attribute added to the selected attributes list, which indicates good weight of this attribute.

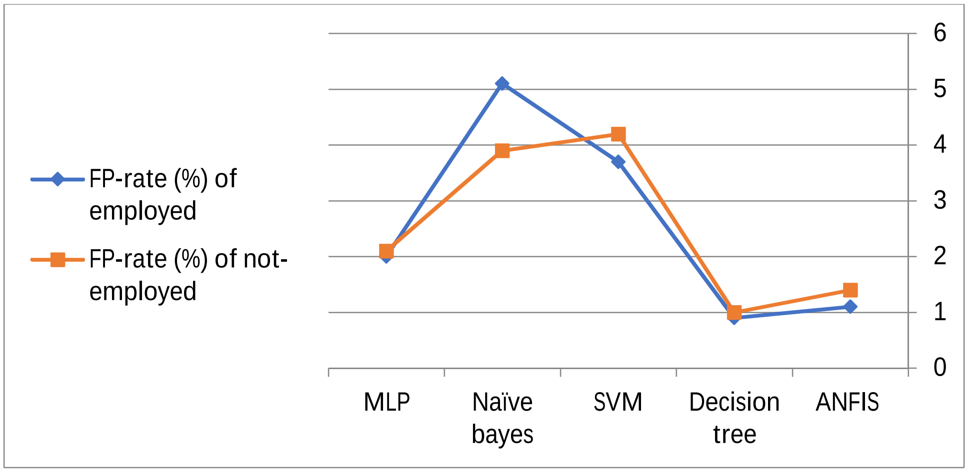

Figure 5. 19 False-Positive rate of employed and not employed classes for each classifier with twenty two attributes

As shown in Figure (5.19) the False-Positive rate values of the employed class for all classifier are lower than False-Positive rate values of not-employed class. This shows again, the excellent ratio of prediction for the “employed” class when applying twenty-two attributes.

Table (5.39) shows the values of RMSE and Kappa measures to measure the efficiency of the prediction classifier when using twenty-two attributes including "No of application" attribute.

Classifier	**RMSE**	**Kappa statistic**
ANFIS	0.1746	0.9176
Decision tree	0.1638	0.9235
SVM	0.2538	0.8264
Naïve Bayes	0.3125	0.7848
MLP	0.1813	0.8843

Table 5. 39 RMSE and Kappa statistic values for each classifier applying twenty two attributes

As shown in Table (5.39), Kappa and RMSE measure values have enhanced when adding "No of application" attribute. Figure (5.20) shows an efficiency comparison of classifiers according to RMSE and Kappa statistic values when applying twenty-two attributes. Also, the above shows that Decision tree has defeated the ANFIS classifier for the first time.

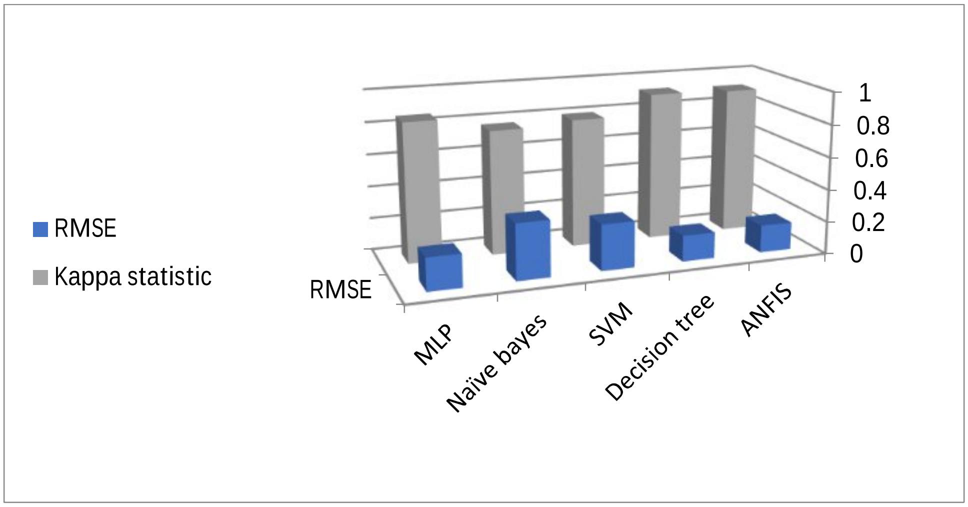

Figure 5. 20 An efficiency comparison of classifiers according to RMSE and Kappa statistic values when applying twenty two attributes

As shown in Figure (5.20), the efficiency performance of all classifiers has improved with preference to Decision tree classifier.

Table (5.40) shows the execution time for each classifier, which shows the computational cost in building classification models using twenty-two attributes.

Classifier	Execution Time (Secs)
ANFIS	17.48
Decision tree	7.16
SVM	7.59
Naïve Bayes	10.48
MLP	18.74

Table 5. 40 Execution time for twenty two attributes classifiers

As shown from Table 5.40, execution time for applying ANFIS classifier when using twenty-two attributes has increased significantly. As shown in the above table, MLP classifier as usual has got the highest value of execution time with 18.74s, which is considered very long time in execution the building task for this classifier, then ANFIS with 17.48 s, which also long time. The values of execution time for all classifiers have also increased. Decision tree has got the less value of execution time of 7.16s, which proves its best time efficiency over the other classifiers when increasing the number of attributes. SVM comes in the second place with a value of 7.59s. The third place goes to Naïve Bayes with time execution of 10.48s.

5.2 Final discussion:

In this section we will discuss the final outcome of conducting the previous experiments. Various numbers of attributes have been used to build and train several classification models. The previous results show that our approach in applying the most seven ranked attributes for each classifier has achieved better accuracy ratio than applying all attributes, which indicates that there are a lot of useless applied attributes. Also, the results of applying all attributes proves that ANFIS classifier unsuitable for applying a large number of attributes, because the more attributes applied the more time is needed to build the model, that is, the number of attributes is directly proportional to the time of implementation. On the other hand, the decision tree classifier proves its efficiency in applying large number of attributes due to the short time it needs in carrying out the classification task and the high accuracy it has achieved in all stages.

During conducting the experiments and increasing the number of attributes gradually, we have found the following high weight attributes:

1- Gender.
2- GPA.
3- Math skills.
4- English Skills.
5- Interpersonal skills.
6- Programming skills.
7- Experience.
8- Status.
9- No of application.

As mentioned previously, we have used attributes selection method to eliminate the attributes and to choose high ranking attributes, these are:

1- University.
2- Math Skills.
3- Major.
4- Gender.
5- GPA.
6- English skills.
7- Programming skills.

As noted from the above two lists, that in our approach we have only seven high weighted attributes, but in the first list there are nine attributes. In our approach, we don't have some of these attributes, such as; Interpersonal skills, Programming skills, Experience, Status, and No of application. And we have five attributes out of seven that exist in the first list, which gives a high rate of convergence between the two approaches.

Table (5.41) illustrates the final outcomes of building ANFIS, decision tress, SVM, Naïve Bayes, and MLP classifiers. The table separates the values of accuracy, error ratio, Kappa, and execution time based on the two different approaches that we have used; we called applying all 22 attributes “Approach no.1” and applying high seven attributes “Approach no. 2”

Classifier	Approach no.	Accuracy (%)	Error ratio	KAPPA	Execution time (Secs)
ANFIS	1	91	0.1746	0.9176	2.40
	2	94	0.1489	0.9364	17.48
Decision Tree	1	92	0.1638	0.9235	1.80
	2	89	0.1912	0.8847	7.16
SVM	1	82	0.2538	0.8264	2.05

	2	88	0.2045	0.8713	7.59
Naïve Bayes	1	77	0.3125	0.7848	2.30
	2	80	0.2251	0.8437	10.48
MLP	1	87	0.1813	0.8843	2.70
	2	91	0.1634	0.9172	18.74

Table 5. 41 Final outcome of applying two different approaches for all classifiers.

As we can note from the above table, the highest accuracy ratio in approach no.1 goes for Decision tree, and it also has got the lowest error ratio with shortest time for execution the classification task. This indicates high efficiency and accuracy for this classifier when applying 22 attributes. On the other hand, ANFIS classifier has got the highest accuracy ratio in "Approach no. 2" and also has achieved highest accuracy over all stages with 94%. As shown in Table (5.41) ANFIS and MLP classifier have the longest time in execution the classification job, which indicates problem in handling large number attributes in both classifiers and increasing the execution time is directly proportional to increasing the number of attributes. The best way to use ANFIS classifier is to eliminate the large number of attributes using attributes selection method to avoid the complexity problem.

5.3 Chapter Summary:

This chapter details how we conducted several experiments to evaluate and test the performance of the ANFIS classifier. The testing process included comparing the result of building an employability prediction model of ANFIS with the results of several classifiers, such as decision tree, SVM, Naïve Bayes, and MLP. In the testing phase, we adopted an incremental approach based on increasing the number of applied attributes gradually. We started applying the classifiers with only three attributes, and increased the number of attributes by one in each testing phase. In the last testing phase, we applied seven attributes. We used several measures to evaluate the performance such as Kappa, RMSE and accuracy measures, and also, we evaluated the efficiency by measuring the execution time of building the classifiers. The results of applying classification models based on various numbers of attributes showed that an increasing number of attributes is directly proportional to increased accuracy and Kappa values, as well as increasing the number of attributes is inversely proportional to RMSE values. Based on these results, increasing the number of attributes used for building the classifiers will enhance the performance of these classifiers. Furthermore, we concluded with a sudden change in the values of execution time for the ANFIS classifier. We have noticed from the results that the execution time of building the ANFIS classifier is increased significantly when the number of attributes is greater than five. This indicates that ANFIS has a complexity problem when applying a large number of attributes.

6 Chapter 6. FINAL PREDICTION OUTCOMES

Introduction

Classification issues are considered as one of the most important data mining topics. Various assessment measures can be used to determine the quality of the data mining technique in the classification, extracting the optimal classifier from the huge dataset that achieves the trade-off between the accuracy of the classifier and the complexity of the classification algorithm.

In this chapter we will discuss the best model (classifier) based on the experimental studies that were conducted in Chapter five of the current research study. In this chapter we will determine the best model, mainly based on the accuracy and computational time as the most important factors for evaluating the classification techniques and the classifiers. Finally, we demonstrate the best classifier based on specific measures.

This chapter is organized as follows: in section 6.1, we will show the best extracted classifier after comparing ANFIS, decision tree, SVM, MLP, and Naïve Bayes on our dataset based on a specific number of attributes. In section 6.2 we will discuss the attributes that most affect the classifier to make it achieve the best performance against other classifiers. In section 6.3 we will compare the best attributes that were determined by our experiments with the human experts' opinions about the best characteristics for graduates to get a job quickly.

6.1 The optimal classifier:

In this thesis, we studied 22 attributes that were collected from the tracer studies of three Jordanian universities. The attributes selected depend on the opinion of experts and related studies of employability. In this thesis, we used some attributes according to attributes selection methods, such as information gaining techniques to choose the attributes with the highest weights. The result of applying an information gaining technique on 22 attributes was seven attributes. The selected attributes are: Gender, Major, GPA, University, Math skills, English skills, and Programming skills.

On the other hand, we adopted an incremental approach to gradually increasing the number of attributes to be implemented in order to generate the classification models, and by applying this approach, we have found nine high weighted attributes these are; Gender, GPA, Math_skills, English_Skills, Interpersonal_skills, Programming skills, Experience, Status, and No of application. We divided our experiments into several phases. In the first phase we applied seven attributes, and increased the number of attributes in each phase by one attribute until reaching seven attributes in the last phase.

Number of attributes	Accuracy (%)	RMSE	KAPPA	Execution Secs time
T	94	0.1489	0.9364	5.53
NN	84	0.2003	0.8746	5.57
NR	86	0.1953	0.8859	8.46
N?	90	0.1838	0.8963	12.67
OO	91	0.1746	0.9176	17.48

Table 6. 1 Number of attributes Accuracy (%) RMSE, Kappa, Execution time (Secs)

As shown in Table 6.1, the best classifier was produced when the number of attributes was seven followed with the 22, with dramatic enhancement. Moreover, the execution time of building the ANFIS classifier was very high.

After using the information gaining technique on the 22 attributes to select the most important attributes and applying the ANFIS incrementally to build the classifier, we started with seven attributes to build the classifier using ANFIS. In the last stage of our experiment, we used all 22 attributes, and we encountered the complexity problem: the computer was entering an infinite loop due to the large number of attributes.

6.2 The best /attributes determination:

After applying the ANFIS algorithm to extract the classifier from the dataset by using the highest ranked attributes, we found that selected attributes significantly affect the accuracy of the classifier during the testing phase. In this section we will discuss the most important attributes through their influences on accuracy; these attributes may help the decision makers in the educational institutions in enhancing their instructional strategies and curriculum, that basically will enhance the outcomes of these institutions and solve the employability problem. These attributes are:

1. Gender: during the experimental studies that were performed using ANFIS algorithm on our dataset, we found the “gender” attribute to be very important to enhance the accuracy of the classifier. This attribute plays an important role in building robust classifiers. When applying this attribute, the accuracy of the classifier was enhanced from 69% to 71%. This means the accuracy is enhanced with a ratio of 2%.
2. GPA: the “GPA” attribute came out in the third place of the most affecting attributes on accuracy. Considering this attribute in the selected attributes has shown good enhancement on the performance of the classifier, which indicates a good role for the “GPA” attribute when building employability classifiers. The accuracy has been enhanced from 71% to 75%, with an enhanced ration of 4%.
3. English skills: in this thesis we studied various skills of the graduates, and after applying the information gain technique different skills were removed from the candidates’ attributes that will be used to construct the classifier. The English language skill is one of the remaining skills which achieved a significant effect on the accuracy of the constructed classifier. When adding this attribute to the selected attributes list, the accuracy is enhanced from 75% to 84% with an enhanced ratio of 9%, which indicates a very important influencing factor.
4. Programming skills: one of the attributes that achieved the best accuracy was the “programming skills” attribute. This attribute has proved its superiority over the other attributes in terms of accuracy. When using this attribute to build the classifier, significant enhancement has occurred. The accuracy has enhanced from 84% to 94% with achieving the highest enhanced ratio of 10%, therefore, the programming skills must be taken into consideration by the students and universities as well. The graduate must have a high level of programming skills to get a job.

6.3 Comparing our classifier with other opinions:

In Chapter five we handled 22 attributes that were selected to build the dataset. The processing on the attributes was performed by three steps: 1- applying the information gain technique to select the most important attributes from 22 attributes; this step was performed to minimize the number of attributes to be suitable as input data to ANFIS algorithm, where if we use the 22 attributes as input data, the system will hang. 2- using the most 7 ranked attributes to build the classifier, and studying the effect of the attributes. 3- Applying incremental approach to implement all 22 attributes and determining the high weight attributes.

In section 6.2 we demonstrated the most influential attributes on the accuracy of the classifier. These results were compared with the experts' opinions to estimate our classifier in terms of the most affecting attributes in building classifiers.

We have interviewed the most important Jordanian employers in the IT field to find out their opinions about the employment factors. The group of experts consisted of recruitment manager, talent hunt recruitment consultants, IT managers, team leaders and senior software engineers. In order to achieve the interviewing task, we firstly looked for well-known and stable Jordanian companies specialized in software development and IT fields such as Amazon (was known previously as Souq), mawdoo3.com, opensooq and dot.jo. the main question was addressed to all of the experts was about which factors affect their choice for hiring a person in CS and IT domains. This main question is divided into some sub-questions about each attribute separately. In this case, it became easier to compare the predictive model outcomes with the experts' opinions. Moreover, giving them the chance to share their experience of hiring fresh graduates from the same domains. The experts pointed out that the most important factors that enhance the chances of employment are:

1- Gender: The experts take the gender attribute into consideration as an employment factor due to these reasons; the CS and IT fields, in general, requires long working hours and sometimes working at night. Due to some cultural restrictions in Jordan, female employees prefer working for short hours and during the day. However, we noticed that this perception is not based on scientific studies. Furthermore, the essential factor restricting employment opportunities for the female is the employer's perception that women are costlier and less productive than male employees. This perception is directly related to women's role in childbearing and baby rising and is emphasized by regulation that places the costs of maternity leave, nursing breaks, and child care directly on the employer. Therefore, female's childbearing and family duties not only limit their availability for work but also discourage employers from hiring them. For these reasons, Jordanian employers prefer to choose male candidates rather than females. This extracted attribute (gender) has met our information-gaining technique in selecting the gender attribute.

2- Communication skills: human experts consider communication skills in order to determine the preferred candidates. Most of the experts confirm that the communication skills factor affects their decision in picking employees. According to the experts, the systems and programs that IT companies develop depend on specific requirements. Understanding the requirements is more important than the coding because otherwise you are coding for nothing. So, experts focus on listening and open-mindedness skills as part of communication skills to understand the requirements. Anyway, in our experiments and based on information-gaining techniques that we have used, the communication skills attribute has been excluded. On the other hand, the majority of employers want a candidate with strong written communication skills. Written communication came in second place of the experts' opinions. That's because being a good writer is about more than writing clear writing. According to the experts' opinions; Clear writing is a sign of having a clear thinking, Good writers know how to interact with teams, they make things easy to be recognized, and they know what to skip. Teamwork and collaboration skills come in third place. According to the experts, by collaborating with team members, the institution or company will have developed and success. By the employee interaction with his colleagues, he/she may reach a better conclusion or idea than he would has on his own. Leadership Skills come in fourth place. The expert considered leadership skills are critical for any executive, management, or supervisory position, and they're the skills needed to generate a vision, inspire people to believe in that vision, and see-through its execution. Figure 6.1 shows the percentage of importance ratio of the communication skills according to the experts' opinions.

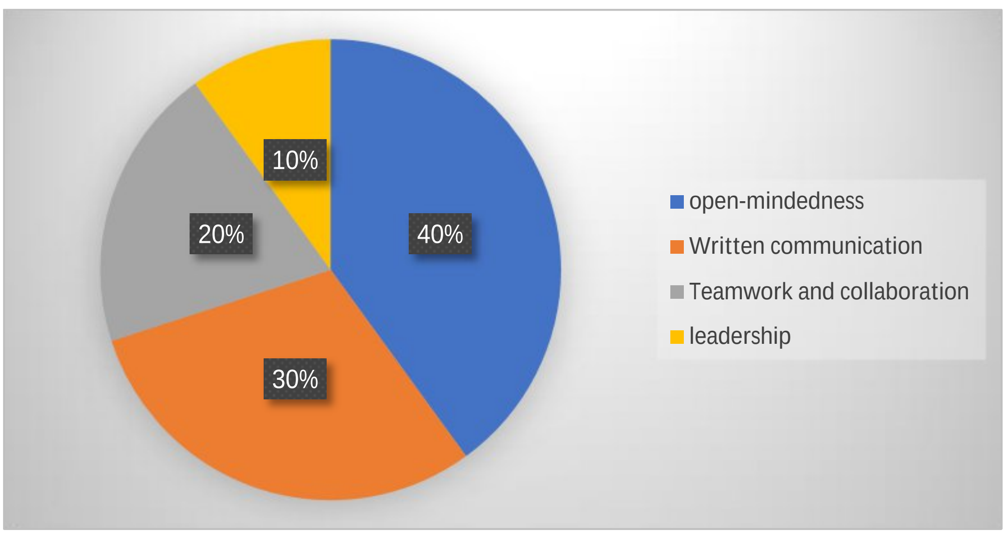

Figure 6. 3 percentage of importance ratios of the communication skills according to the experts' opinions

3- Programming skills: the experts emphasized the importance of this factor and considered it the basis for recruiting candidates. Also, the experts confirm that specific programming skills give a high chance of employment rather than other programming skills. According to the experts, mastering one or more of these skills: IOS, android, java, PHP, Python, JavaScript, Angular, and ReactJS, significantly affects the decision to accept the candidate or not. As mentioned in the previous section, the best accuracy of applying the ANFIS classifier in implementing the highest seven ranked attributes occurred when using the "Programming Skills" attribute. This indicates that our results are consistent with expert opinion. According to the collected data from the experts, JavaScript programing language comes in first place, JavaScript is a well-known language between developers who dream to work on server-side and client-side programming. It is compatible with several other programming languages. JavaScript is an ultimate hit in the IT domain and required a lot in IT companies. Python programing language comes in second place, it continues to be one of the best programming languages each programmer should learn, this language is easy-to-learning and offers a clean and well-structured code, making it influential enough to build a decent web application, the experts consider this language is very helpful to achieve data analyst tasks which is very required topic recently. Java programming language comes in third place; Java is a practical choice for developing Android apps as it can be applied to create highly functional programs and platforms. The experts confirmed that Java has a good level of security. Moreover, it is easier to learn Java in comparison to languages such as C and C++. Figure 6.2 shows the most required programming language in the Jordanian market according to the data collected from the experts.

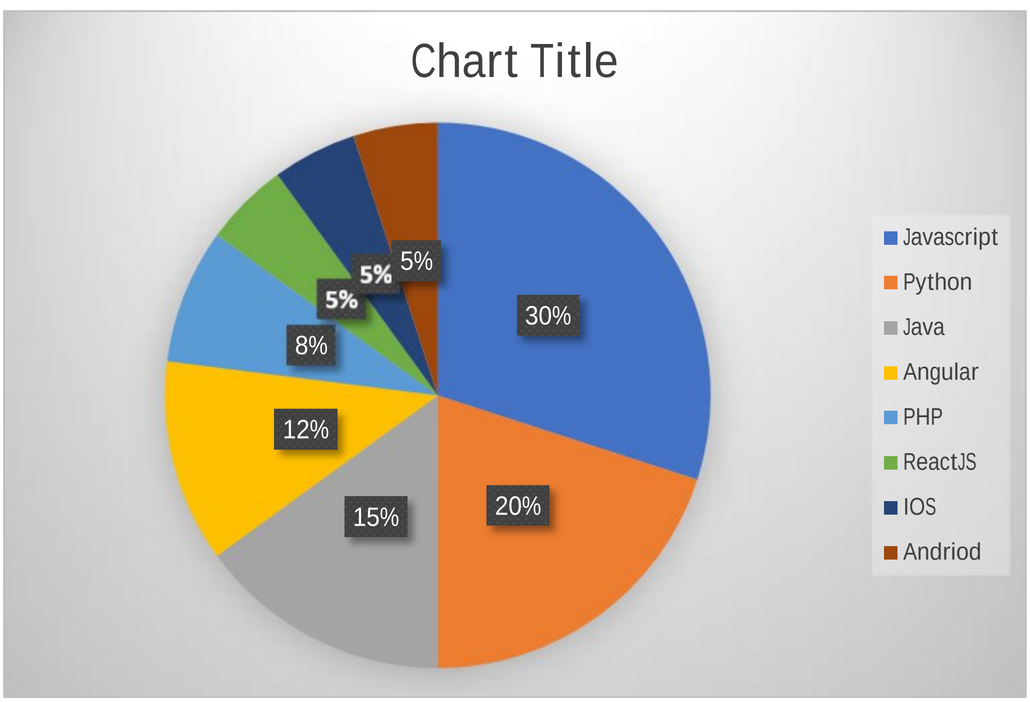

Figure 6. 4 The ratio of the most required programming languages in Jordanian market according to the experts' opinions

As mentioned in the previous section, the best accuracy of applying the ANFIS classifier in our experiments occurred when using the “Programming Skills” attribute. This indicates that our results are consistent with expert opinion.

6.4 Chapter Summary

In this chapter we discuss the best classifier of the overall classifiers that were extracted in Chapter five. In this thesis we take more into consideration the accuracy of the classifier, thus the best classifier from our point of view is the classifier that contains the highest seven ranked attributes according to the attributes selection method. The computational time to extract the best classifier was very long when the number of attributes is seven; also the computational time is extremely long when the dataset is very big. In section two of this chapter, we discussed the best attribute that affects the accuracy of the classifier. We found Gender, GPA, Programming skills and English skills to be the most important attributes that the universities should be interested in to make their students more likely to get a job. In section three we compared the results that we found in section two with the experts’ opinions and we noted compatibility between the experts’ opinions and our results.

7 Chapter 7. CONCLUSION

This chapter provides the answers of the research questions depending on the proposed framework of the research problem, discusses the conclusion from the experiments, and goes on to note the study limitations, contributions, and recommendations for further studies in the future.

This research study provides a multi-faceted domain for statistical analysis and data analytics in order to consider different attributes affecting CS and IT graduates future employability in Jordan. Neuro-fuzzy technique is used to classify CS and IT graduates in Jordan according to their employment status. Additionally, there is a critical need to find a relation between what is supplied by the ministries of high education (graduates) and what is needed by the labour market (employees). In order to answer the research questions, this research study tried to figure out the most effective factors affecting the future of employability for graduate students from both of Computer Science and Information Technology majors in the Middle East by applying the following steps:

- To better understanding the current status of IT employment rates in Jordan, we contacted the Ministry of Digital Economy and Entrepreneurship and the Ministry of Labour in Jordan, to collect useful data. The collected data indicated that about 40% of total graduates can got job, male students have high chance to get job against female. The data indicated that most of CS, IT graduates are employed in the private sectors with more than 90%, and less than 10% of total graduates are employed in the public sector. The computer science graduates got the highest number of jobs, and the lowest number goes to information network systems specialization with less than 1% of total graduates. Also, the collected data showed that the demand for male graduates in the IT market of Jordan is more than for female graduates. And also, the statistics indicated the superiority of some specializations in the IT field, such as computer network security engineering, computer engineering, and software engineering.
- After that, the testing process included comparing the results of building an employability prediction model of ANFIS with the results of several classifiers, such as decision tree, SVM, Naïve Bayes, and MLP. In the testing phase, we adopted an incremental approach based on increasing the number of applied attributes gradually. We started by applying the classifiers with only three attributes, and increased the number of attributes by one in each testing phase. In the last

testing phase, we applied seven attributes. The experimental studies showed that the ANFIS algorithm as a classification technique achieved the best prediction for unseen instance based on several evaluation measure such as Kappa, RMSE and accuracy measures, but unfortunately, the computational time for implemented ANFIS to build the classifier become very high when the number of attributes is greater than five. This indicates that ANFIS has a complexity problem when applying a large number of attributes.

- After applying the ANFIS algorithm to extract the classifier from the dataset by using an incremental approach, we found that certain attributes significantly affect the accuracy of the classifier during the testing phase. In this thesis, we found the most important attributes that effect the accuracy of the classifier; these factor most affect the IT student's employability future. These attributes are:
  1- Gender: during the experimental studies that were performed using ANFIS algorithm on our dataset, we found "gender" attribute to be very important for enhancing the accuracy of the classifier. The gender of the graduate in Jordan play important role in employability issue.
  2- GPA: this attribute is very important affecting the graduate's employment. The "GPA" attribute came out in the third place of the most affecting attributes on accuracy of the classifier during experiments. Student that achieved high GPA has good chance to get job.
  3- English skills: our experiments indicated that the English language skills is a significant factor that helps the CS and IT graduates to get job in their fields. The English skills attribute achieved a significant effect on the accuracy of the constructed classifier. When adding this attribute to the selected attributes list, the accuracy is enhanced with a ratio of 9%, which indicates a very important influencing factor.
  4- Programming skills: this attribute is considered as the most important factor that affect CS, IT student's employability future. In this thesis this attribute has proved its superiority over the other attributes in terms of accuracy of the classifier when it takes in consideration to build the classifier. The accuracy achieving the highest enhanced ratio of 10%, therefore, the programming skills must be taken into consideration by the students and universities as well. The graduate must have a high level of programming skills to get a job.
- All in all, in this thesis the ANFIS algorithm is implemented on the training dataset that was collected from the tracer studies conducted in three Jordanian universities. The training dataset is used for training

and testing purposes. The testing process carried out based on 10-fold cross validation, where the dataset divided into 10 portions, one of them used for testing and the rest are used for training. This process is performed ten times with different portion each time. The final classifier that is built from ANFIS consist of 7 attributes as antecedents and one as consequent which is the employability status. the results indicated Gender, GPA, Programming skills and English skills are the most important factors affecting the CS and IT student's employability future, thus the universities should be interested in providing their students with the required skills to get a job. When we compared our results with the experts' opinions about the most important factors that affect the CS and IT student's employability we noted compatibility between the experts' opinions and our results. According to the results, a list of recommendations is provided to the ministry of high education as followed:

- The admission process should follow a certain standard in accepting allocated percentage from both males and females in the majors of CS and IT.
- English language skills should be taken into considerations in accepting newly registered students in both of CS and IT majors.
- Decision makers and curriculum developers should enhance the courses given to CS and IT students with more programming and soft skills.

7.1 Contributions:

In this thesis the employability issue has been studied, which represents a very important problem for Middle East countries, where the number of graduates from higher institutions is very large, but in contrast the job opportunities are too low. Actually, the higher educational institutions realize the importance of using data mining to improve their curriculums, teaching strategies and employability issues. In this thesis we use a modern data mining technique called neuro-fuzzy inference system algorithm (ANFIS) to study the real problem facing the Middle East countries, which is the employability problem. This problem was studied using various classification algorithms such as SVM, K-nearest, Naïve based and Neural Networks.

In this thesis the problem was studied using the ANFIS which is an algorithm that combines fuzzy inference system and neural networks, which to the best of our knowledge has never been implemented on the employability problem. The experimental study was performed using the dataset collected through tracer studies carried out by three public and private universities in Jordan. This provided a major foundation for the current research study which is combining statistical methods that have not been combined before with analytics methods. This combination will ease the understanding of the statistical relationships between data sets.

The experimental study was implemented on the dataset containing 22 attributes, which is a huge number of attributes when you apply ANFIS, thus the experiments were incrementally carried out with different numbers of attributes, as we started with three attributes and gradually increased attributes (i.e. 4, 5, … 22). The experiments showed that the accuracy increased with increasing the number of elements, but the computational time increased significantly. The ANFIS was compared with various datamining classification techniques, which are SVM, Decision Tree, Naïve based and Multilayer Perceptron in order to classify unseen instances. The experimental results showed the ANFIS achieved high accuracy against four classification algorithms (SVM, Naïve based, Decision Tree and Multilayer Perceptron). ANFIS achieved 94% accuracy on the graduates' dataset but the complexity of ANFIS was very high when a large number of attributes was used.

7.2 Limitations:

1. Complexity: ANFIS uses a neural networks structure, where there are fixed and adapted nodes. The adapted nodes modify after each iteration of ANFIS. The membership functions are modified using gradient descent classifier and the computed consequent is modified using the least square error (LSE). The modifications of the nodes increase the complexity of the implementation, that makes the computational time when we input a high number of attributes and rules very high. In this thesis we carried out the experiments using various numbers of input attributes to reach the tradeoff between the accuracy and computational time.
2. Number of parameters: in this thesis an incremental strategy was used on the dataset during experiments. The incremental strategy was used because when we used the whole attributes of the dataset as input for ANFIS, we faced problems in implementation (i.e. space out of memory, the computer ran a long time to create the model). Thus a number of attributes were selected each running time then the created model was tested. The number of attributes was increased each time.

3. Dataset size: the dataset size and the number of rules is the most important factor that affects the computation time for creating the model from ANFIS. The number of rules that extract from the huge dataset is very big, thus the computation time to create the model expands dramatically; also, the interoperability problem emerges, for the end user to be able to understand the classifier.
4. Collecting data: in this thesis the dataset was collected from the tracer study conducted by three universities in Jordan. The tracer study was performed in different ways and the components of them differ, hence the processing of them in order to build a unified dataset takes a lot of time and effort. Some of the tracer studies performed well while others were weaker, therefore in this thesis we relied on only three tracer studies in order to build our dataset.
5. Accuracy: in this thesis we implement ANFIS with grid partitioning using MATLAB. We use this technique because we want to achieve high accuracy during the testing phase, and unfortunately this method produces a large number of rules which make it hardly understood by the end user. Hence, if we need to increase the accuracy we will lose the interoperability. During our experimental results we try to make tradeoff between the accuracy and interoperability.
6. Membership function: there is a distribution of various membership functions (e.g. Triangular, Trapezoidal and Gaussian). In this thesis Gaussian membership function has been used in order achieve better accuracy; unfortunately, the accuracy is achieved at the expense of computation time. Other membership function distributions such as triangular are simple but the accuracy is low in comparison with Gaussian.
7. Other important limitations that we encountered are interpretability and explainability. The main idea of interpretability is to make the internals of a system able to be understood by humans, and it can ensure confidence in our proposed model. Explanations help us to determine whether the predictions are unbiased. Unfortunately, ANFIS is considered a black-box model and we can't determine how and where the decision comes from in order to judge the prediction. Interpretability and explainability play important roles to ensure effective interactions between humans and classification systems but this is not the case in our proposed system. ANFIS and NN systems generally just output the decision itself without explaining, and it is more convenient for humans to trust a system that interprets its decision.

7.3 Implications for Future Work:

1. Determine the best attributes: the number of attributes greatly affect the performance of ANFIS. Many research studies have touched on this issue; some of them aggregate the attributes into a small number of attributes, which reduces the computational time but reduces the accuracy of prediction dramatically as well. Another group of research studies determined the best attribute heuristically. The information gain and Gini index can be used to select the best attributes to build the model using a small number of attributes, also the human expert can help to select the suitable attributes. A lot of techniques can be studied in the future to reduce the number of inputs by selecting the best attributes.
2. Sampling: sampling is a statistical process in which a number of instances are taken from large population. The sampling might be random sampling or systematic sampling. As the size of the dataset and the number of rules has an effect on the efficiency of the ANFIS algorithm, the sampling can be used to reduce the size of the rule by taking a stratified random set which may give good accuracy and reduce the computational time.
3. Reduce the computational time: time is required to build the model using ANFIS, especially when the number of attributes is large, or is long. A lot of research studies are performed to reduce the running time to produce the model, some of which focus on the membership distribution function by selecting the simple function or by reducing the modification process, and other research studies focus on pruning the insignificant rules to reduce the number of rules that will be processed, then reducing the time. Time reduction is a very important topic that needs a lot of research studies to solve it.
4. Study different features: modification of the computed consequent and the membership distribution function, carried out by LSE and gradient descent, respectively. A lot of research studies are performed to modify the adaptable node in ANFIS, such as using Genetic algorithm and Bee Colony, but this topic needs further research to enhance the modification process of the adaptable node in ANFIS.

8 REFERENCES:

Aamodt, P. O. and Havnes, A. (2008) 'Factors affecting professional job mastery: Quality of study or work experience?', *Quality in Higher Education*. doi: 10.1080/13538320802507539.

Abadi, M. *et al.* (2016) 'TensorFlow: A system for large-scale machine learning', in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2016*.

Abas, M. C. and Imam, O. A. (2016) 'Graduates' Competence on Employability Skills and Job Performance', *International Journal of Evaluation and Research in Education (IJERE)*. doi: 10.11591/ijere.v5i2.4530.

Abdallah, S. *et al.* (2014) 'DNVA: A Tool for Visualizing and Analyzing Multi-agent Learning in Networks', in *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*. doi: 10.1109/ICTAI.2014.67.

Abu, A. (2016) 'Educational Data Mining & Students' Performance Prediction', *International Journal of Advanced Computer Science and Applications*. doi: 10.14569/ijacsa.2016.070531.

Agrawal, R. and Srikant, R. (1994) 'Fast Algorithms for Mining Association Rules', in *Proc. of 20th International Conference on Very Large Data Bases, {VLDB'94}*.

Ahmed, A. B. E. D. and Elaraby, I. S. (2014) 'Data Mining: A prediction for performance improvement using classification', *World Journal of Computer Application and Technology*. doi: 10.13189/wjcat.2014.020203.

Ajith, P., Tejaswi, B. and Sai, M. S. S. (2013) 'Rule Mining Framework for Students Performance Evaluation', *International Journal of Soft Computing and Engineering (IJSCE)*.

Al-Shalabi, R., Kanaan, G. and Gharaibeh, M. H. (2006) 'Arabic Text Categorization Using kNN Algorithm', *Proceedings of The 4th International Multiconference on Computer Science and Information Technology*.

Albawi, S., Mohammed, T. A. and Al-Zawi, S. (2018) 'Understanding of a convolutional neural network', in *Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017*. doi: 10.1109/ICEngTechnol.2017.8308186.

Altaher, A. and BaRukab, O. (2017) 'Prediction of Student's Academic Performance Based on Adaptive Neuro-Fuzzy Inference', *International Journal of Computer Science and Network Security*.

Amin, R. K., Indwiarti and Sibaroni, Y. (2015) 'Implementation of decision tree using C4.5 algorithm in decision making of loan application by debtor (Case study: Bank pasar of Yogyakarta Special Region)', in *2015 3rd International Conference on Information and Communication Technology, ICoICT 2015*. doi: 10.1109/ICoICT.2015.7231400.

Amrin Amrin (2017) 'Data Mining Dengan Algoritma Apriori untuk Penentuan Aturan Asosiasi Pola Pembelian Pupuk', *Paradigma*. doi: https://doi.org/10.31294/p.v19i1.1836.

Anand, S. S., Bell, D. A. and Hughes, J. G. (1996) 'EDM: A general framework for Data Mining based on Evidence Theory', *Data and Knowledge Engineering*. doi: 10.1016/0169-023X(95)00038-T.

Anderson, D. *et al.* (2009) 'Modeling human activity from voxel person using fuzzy logic', *IEEE Transactions on Fuzzy Systems*. doi: 10.1109/TFUZZ.2008.2004498.

Araque, F., Roldán, C. and Salguero, A. (2009) 'Computers & Education Factors influencing university drop out rates', *Computers & Education*. doi: 10.1016/j.compedu.2009.03.013.

Arel, I., Rose, D. and Karnowski, T. (2010) 'Deep machine learning-A new frontier in artificial intelligence research', *IEEE Computational Intelligence Magazine*. doi: 10.1109/MCI.2010.938364.

Arnould, T. *et al.* (1995) 'Algorithms for fuzzy inference and tuning in the fuzzy inference software FINEST', in *IEEE International Conference on Fuzzy Systems*. doi: 10.1109/fuzzy.1995.409811.

Atashi, A. *et al.* (2017) 'Breast Cancer Risk Assessment Using adaptive neuro-fuzzy inference system (ANFIS) and Subtractive Clustering Algorithm', *Multidisciplinary Cancer Investigation*. doi: 10.21859/mci-01029.

Atsumi et al., 2007 (2007) 'Methods used to assess implant stability: current status.', *The International journal of oral & maxillofacial implants*.

Ayodele, T. O. (2010) 'Types of Machine Learning Algorithms', in *Types of Machine Learning Algorithms, New Advances in Machine Learning*. doi: 10.5772/9385.

Badr, A., Din, E. and Elaraby, I. S. (2014) 'Data Mining : A prediction for Student ' s Performance Using Classification Method', *World Journal of Computer Application and Technology*. doi: 10.13189/wjcat.2014.020203.

Bardenet, R. *et al.* (2013) 'Collabourative hyperparameter tuning', in *30th International Conference on Machine Learning, ICML 2013*.

Belaout, A. *et al.* (2017) 'Multi-class neuro-fuzzy classifier for photovoltaic array faults diagnosis', in *2017 5th International Conference on Electrical Engineering - Boumerdes, ICEE-B 2017*. doi: 10.1109/ICEE-B.2017.8192007.

Berland, M., Baker, R. S. and Blikstein, P. (2014) 'Educational data mining and learning analytics: Applications to constructionist research', *Technology, Knowledge and Learning*. doi: 10.1007/s10758-014-9223-7.

Bezuidenhout, A. and Jeppesen, S. (2011) 'Between market, state and society: Labour codes of conduct in the southern African garment industry', *Development Southern Africa*. doi: 10.1080/0376835X.2011.623923.

Bharambe, Y. *et al.* (2017) 'Assessing employability of students using data mining techniques', in *2017 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2017*. doi: 10.1109/ICACCI.2017.8126157.

Bhaskaran, S., Lu, K. and Al Aali, M. (2016) ‘A data mining approach for investigating students’ completion rates’, in *Higher Education in the Twenty-First Century II - Papers From The Two-Day Workshop, 2014*. doi: 10.1201/b19335-11.

Biau, G. and Scornet, E. (2016) ‘A random forest guided tour’, *Test*. doi: 10.1007/s11749-016-0481-7.

Bielza, C. and Larrañaga, P. (2014) ‘Discrete bayesian network classifiers: A survey’, *ACM Computing Surveys*. doi: 10.1145/2576868.

Bolón-Canedo, V. and Alonso-Betanzos, A. (2018) ‘Software tools’, in *Intelligent Systems Reference Library*. doi: 10.1007/978-3-319-90080-3_9.

Bovo, A. *et al.* (2013) ‘Clustering moodle data as a tool for profiling students’, in *2013 2nd International Conference on E-Learning and E-Technologies in Education, ICEEE 2013*. doi: 10.1109/ICeLeTE.2013.6644359.

Boyacioglu, M. A. and Avci, D. (2010) ‘An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: The case of the Istanbul stock exchange’, *Expert Systems with Applications*. doi: 10.1016/j.eswa.2010.04.045.

Boyer, K. E. (2010) *Structural and Dialogue Act Modeling in Task-Oriented Tutorial Dialogue*, ProQuest Dissertations and Theses.

Bradford, J. P. *et al.* (1998) ‘Pruning decision trees with misclassification costs’, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. doi: 10.1007/bfb0026682.

Branine, M. and Avramenko, A. (2015) ‘A Comparative Analysis of Graduate Employment Prospects in European Labour Markets: A Study of Graduate Recruitment in Four Countries’, *Higher Education Quarterly*. doi: 10.1111/hequ.12076.

Bridgstock, R. (2009) ‘The graduate attributes we’ve overlooked: Enhancing graduate employability through career management skills’, *Higher Education Research and Development*. doi: 10.1080/07294360802444347.

BROWN, P., HESKETH, A. and WILIAMS, S. (2013) ‘Employability in a Knowledge-driven Economy’, *Journal of Education and Work*. doi: 10.1080/1363908032000070648.

Cairns, A. H. *et al.* (2015) ‘Process Mining in the Education Domain’, *International Journal on Advances in Intelligent Systems*.

Cao, L. (2017) ‘Data science: A comprehensive overview’, *ACM Computing Surveys*. doi: 10.1145/3076253.

Castro, F. *et al.* (2005) ‘Finding relevant features to characterize student behavior on an e-learning system’, in *Proceedings of the 2005 International Conference on Frontiers in Education: Computer Science and Computer Engineering, FECS’05*.

Çaydaş, U., Hasçalik, A. and Ekici, S. (2009) 'An adaptive neuro-fuzzy inference system (ANFIS) model for wire-EDM', *Expert Systems with Applications*. doi: 10.1016/j.eswa.2008.07.019.

Chan, Y. H., Correa, C. D. and Ma, K. L. (2013) 'The generalized sensitivity scatterplot', *IEEE Transactions on Visualization and Computer Graphics*. doi: 10.1109/TVCG.2013.20.

Chattopadhyay, S. *et al.* (2020) 'Towards effective discovery of natural communities in complex networks and implications in e-commerce', *Electronic Commerce Research*. doi: 10.1007/s10660-019-09395-y.

Chen, G. D. *et al.* (2000) 'Discovering decision knowledge from web log portfolio for managing classroom processes by applying decision tree and data cube technology', *Journal of Educational Computing Research*. doi: 10.2190/5JNM-B6HP-YC58-PM5Y.

Chen, G. H. and Shah, D. (2018) 'Explaining the success of nearest neighbor methods in prediction', *Foundations and Trends in Machine Learning*. doi: 10.1561/2200000064.

Chen, J., Li, Q. and Jia, W. (2005) 'Automatically generating an E-textbook on the Web', *World Wide Web*. doi: 10.1007/s11280-005-1319-5.

Cheng, J. and Greiner, R. (2001) 'Learning bayesian belief network classifiers: Algorithms and system', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. doi: 10.1007/3-540-45153-6_14.

Chu, J., Lee, T. H. and Ullah, A. (2020) 'Component-wise AdaBoost algorithms for high-dimensional binary classification and class probability prediction', in *Handbook of Statistics*. doi: 10.1016/bs.host.2018.10.003.

Computational Intelligence: Soft Computing and Fuzzy-Neuro Integration with Applications (1998) *Computational Intelligence: Soft Computing and Fuzzy-Neuro Integration with Applications*. doi: 10.1007/978-3-642-58930-0.

Curtis, E. M. I. (2015) *Firm behavior, environmental externalities and public policy.*, *Dissertation Abstracts International Section A: Humanities and Social Sciences*.

Cyphers, B. (2017) 'Differential privacy , part 3 : Extraordinary claims require extraordinary scrutiny', *Access Now*.

Dawson, C. and Dawson, C. (2019) 'Educational data mining', in *A–Z of Digital Research Methods*. doi: 10.4324/9781351044677-18.

Deci, E. L. and Ryan, R. M. (2015) 'Self-Determination Theory', in *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*. doi: 10.1016/B978-0-08-097086-8.26036-4.

Dejaeger, K., Verbraken, T. and Baesens, B. (2013) 'Toward comprehensible software fault prediction models using bayesian network classifiers', *IEEE Transactions on Software Engineering*. doi: 10.1109/TSE.2012.20.

Do, Q. H. and Chen, J. F. (2013) 'A comparative study of hierarchical anfis and ann in predicting student academic performance', *WSEAS Transactions on Information Science and Applications*.

Van Dongen, B. F. *et al.* (2005) 'The ProM framework: A new era in process mining tool support', in *Lecture Notes in Computer Science*. doi: 10.1007/11494744_25.

Dorfman, K. D. and Daoutidis, P. (2018) 'MATLAB "Tutorial"', in *Numerical Methods with Chemical Engineering Applications*. doi: 10.1017/9781316471425.010.

Dutt, A., Ismail, M. A. and Herawan, T. (2017) 'A Systematic Review on Educational Data Mining', *IEEE Access*. doi: 10.1109/ACCESS.2017.2654247.

Ellis, T. J. and Dringus, L. P. (2005) 'Evaluating threaded discussion forum activity: Faculty and student perspectives on categories of activity', in *Proceedings - Frontiers in Education Conference, FIE*. doi: 10.1109/fie.2005.1611921.

El-Temtamy, O., Kathleen O'Neill, K. and Midraj, S. (2016) 'Undergraduate employability training and employment: A UAE study', *Higher Education, Skills and Work-based Learning*. doi: 10.1108/HESWBL-02-2015-0006.

Ezrachi, A. and Stucke, M. E. (2017) 'Artificial intelligence & collusion: When computers inhibit competition', *University of Illinois Law Review*.

Faber, N. M. and Rajkó, R. (2007) 'How to avoid over-fitting in multivariate calibration-The conventional validation approach and an alternative', Analytica Chimica Acta. doi: 10.1016/j.aca.2007.05.030.

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996) 'From data mining to knowledge discovery in databases', *AI Magazine*.

Fernández-Delgado, M. *et al.* (2014) 'Do we need hundreds of classifiers to solve real world classification problems?', *Journal of Machine Learning Research*. doi: 10.1117/1.JRS.11.015020.

Fransen, K. *et al.* (2019) 'The relationship between transport disadvantage and employability: Predicting long-term unemployment based on job seekers' access to suitable job openings in Flanders, Belgium', *Transportation Research Part A: Policy and Practice*. doi: 10.1016/j.tra.2018.01.023.

García-Martínez, C., Rodriguez, F. J. and Lozano, M. (2018) 'Genetic algorithms', in *Handbook of Heuristics*. doi: 10.1007/978-3-319-07124-4_28.

García, E. *et al.* (2011) 'A collabourative educational association rule mining tool', *Internet and Higher Education*. doi: 10.1016/j.iheduc.2010.07.006.

Ghesu, F. C. *et al.* (2019) 'Multi-Scale Deep Reinforcement Learning for Real-Time 3D-Landmark Detection in CT Scans', *IEEE Transactions on Pattern Analysis and Machine Intelligence*. doi: 10.1109/TPAMI.2017.2782687.

Ghosh, A., Uma Shankar, B. and Meher, S. K. (2009) 'A novel approach to neuro-fuzzy classification', *Neural Networks*. doi: 10.1016/j.neunet.2008.09.011.

Ghosh, S. *et al.* (2014) ‘A novel Neuro-fuzzy classification technique for data mining’, *Egyptian Informatics Journal*. doi: 10.1016/j.eij.2014.08.001.

Grobelnik, M., Mladenic, D. and Jermol, M. (2002) ‘Exploiting text mining in publishing and education’, in *ICML-2002 Workshop on Data Mining Lessons Learned*.

de Guzman, A. B. and Choi, K. O. (2013) ‘The relations of employability skills to career adaptability among technical school students’, *Journal of Vocational Behavior*. doi: 10.1016/j.jvb.2013.01.009.

Hall, M. *et al.* (2009) ‘The WEKA data mining software’, *ACM SIGKDD Explorations Newsletter*. doi: 10.1145/1656274.1656278.

Hamdan, H. and Garibaldi, J. (2013) ‘An Exploration of the Adaptive Neuro-Fuzzy Inference System (ANFIS) in Modelling Survival’, *School of Computer Science* .

Hammouda, K. and Kamel, M. (2008) ‘Distributed collabourative Web document clustering using cluster keyphrase summaries’, *Information Fusion*. doi: 10.1016/j.inffus.2006.12.001.

Han, Y., & Lee, J. (2016). U. P. D. for M.-C. M. M. N. I. C. L. https://doi. org/10. 1109/LCOMM. 2016. 257169. *et al.* (2016) ‘Performance Analysis of Physical Layer Security Over Generalized-<inline-formula> <img src="/images/tex/523.gif" alt="K"> </inline-formula> Fading Channels Using a Mixture Gamma Distribution’, *IEEE Communications Letters*. doi: 10.1109/LCOMM.2015.2504580.

Han, J. *et al.* (2007) ‘Frequent pattern mining: Current status and future directions’, *Data Mining and Knowledge Discovery*. doi: 10.1007/s10618-006-0059-1.

Han, J., Kamber, M. and Pei, J. (2011) *Data Transformation by Normalization*, *Data Mining: Concepts and Techniques*. doi: 10.1016/B978-0-12-381479-1.00001-0.

HarvinderChauhan and AnuChauhan (2014) ‘Implementation of the Apriori algorithm for association rule mining’, *Compusoft*.

Hassan, D. J., 2011/2020.
https://www.ilo.org/dyn/youthpol/es/equest.fileutils.docHandle?p_uploaded_file_id=171. [Online]
Available at: https://www.ilo.org/dyn/youthpol/es/equest.fileutils.docHandle?p_uploaded_file_id=171
[Accessed Friday April 2020].

Helens-Hart, R. (2015) *Employability and empowerment: Discursive constructions of career planning*, *ProQuest Dissertations and Theses*.

Heller, T. *et al.* (2011) ‘Self-determination across the life span: Issues and gaps’, *Exceptionality*. doi: 10.1080/09362835.2011.537228.

Henderson, A. (2013) ‘Top 5: Countries by highest inflation rate’, *Nomad Capitalist*.

Hernández-Blanco, A. *et al.* (2019) ‘A Systematic Review of Deep Learning Approaches to Educational Data Mining’, *Complexity*. doi: 10.1155/2019/1306039.

Hernández-Orallo, J., Flach, P. and Ferri, C. (2013) ‘ROC curves in cost space’, *Machine Learning*. doi: 10.1007/s10994-013-5328-9.

Hesterberg, T. (2011) ‘Bootstrap’, *Wiley Interdisciplinary Reviews: Computational Statistics*. doi: 10.1002/wics.182.

Hillage, J. and Pollard, E. (1998) ‘Employability: Developing a Framework for Policy Analysis, Department for Education and Employment’, *Institute for Employment Studies*.

Hjorth, J. S. U. and Hjorth, J. S. U. (2018) ‘Cross validation’, in *Computer Intensive Statistical Methods*. doi: 10.1201/9781315140056-3.

Hoffmann, F. (2004) ‘Combining boosting and evolutionary algorithms for learning of fuzzy classification rules’, in *Fuzzy Sets and Systems*. doi: 10.1016/S0165-0114(03)00113-1.

HSSINA, B. *et al.* (2014) ‘A comparative study of decision tree ID3 and C4.5’, *International Journal of Advanced Computer Science and Applications*. doi: 10.14569/specialissue.2014.040203.

Huang, M. L. *et al.* (2012) ‘Usage of case-based reasoning, neural network and adaptive neuro-fuzzy inference system classification techniques in breast cancer dataset classification diagnosis’, *Journal of Medical Systems*. doi: 10.1007/s10916-010-9485-0.

Husain, M. Y. *et al.* (2010) ‘Importance of employability skills from employers’ perspective’, in *Procedia - Social and Behavioral Sciences*. doi: 10.1016/j.sbspro.2010.10.059.

Hutson, M. (2018) ‘Artificial intelligence faces reproducibility crisis’, *Science*. doi: 10.1126/science.359.6377.725.

Jang, J. S. R. (1992) ‘Self-Learning Fuzzy Controllers Based on Temporal Back Propagation’, *IEEE Transactions on Neural Networks*. doi: 10.1109/72.159060.

Jang, J. S. R. (1993) ‘ANFIS: Adaptive-Network-Based Fuzzy Inference System’, *IEEE Transactions on Systems, Man and Cybernetics*. doi: 10.1109/21.256541.

Jang, J. S. R. and Sun, C. T. (1995) ‘Neuro-Fuzzy Modeling and Control’, *Proceedings of the IEEE*. doi: 10.1109/5.364486.

Jantawan, B. and Tsai, C.-F. (2013) ‘The Development of Educational Quality Administration: a Case of Technical College in Southern Thailand’, *International Journal of Computer Science and Information Security*.

Kalaiselvi, C. and Nasira, G. M. (2014) ‘A new approach for diagnosis of diabetes and prediction of cancer using ANFIS’, in *Proceedings - 2014 World Congress on Computing and Communication Technologies, WCCCT 2014*. doi: 10.1109/WCCCT.2014.66.

Kar, S., Das, S. and Ghosh, P. K. (2014) ‘Applications of neuro fuzzy systems: A brief review and future outline’, *Applied Soft Computing Journal*. doi: 10.1016/j.asoc.2013.10.014.

Kaur, J. and Madan, N. (2015) ‘Association Rule Mining: A Survey’, *International Journal of Hybrid*

Information Technology. doi: 10.14257/ijhit.2015.8.7.22.

Kesavaraj, G. and Sukumaran, S. (2013) 'A study on classification techniques in data mining', in *2013 4th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2013*. doi: 10.1109/ICCCNT.2013.6726842.

Keviczky, L. *et al.* (2019) 'Introduction to MATLAB', *Advanced Textbooks in Control and Signal Processing*. doi: 10.1007/978-981-10-8321-1_1.

Kohavi, R. and Quinlan, R. (1999) 'Decision Tree Discovery', *in Handbook of Data Mining and Knowledge Discovery*. doi: 10.1111/j.1461-0248.2010.01557.x.

Kou, Y. and Lu, C.-T. (2016) 'Outlier Detection', in *Encyclopedia of GIS*. doi: 10.1007/978-3-319-23519-6_944-2.

Kumar, B. and Pal, S. (2011) 'Mining Educational Data to Analyze Students Performance', *International Journal of Advanced Computer Science and Applications*. doi: 10.14569/ijacsa.2011.020609.

Kundi, Y. M. *et al.* (2017) 'Affective Commitment as Mechanism behind Perceived Career Opportunity and Turnover Intentions with Conditional Effect of Organizational Prestige', *Journal of Managerial Sciences*.

Kuosa, K. *et al.* (2016) 'Interactive visualization tools to improve learning and teaching in online learning environments', *International Journal of Distance Education Technologies*. doi: 10.4018/IJDET.2016010101.

Laine, K., Leino, M. and Pulkkinen, P. (2015) 'Open Innovation Between Higher Education and Industry', *Journal of the Knowledge Economy*. doi: 10.1007/s13132-015-0259-2.

Landuyt, D. *et al.* (2013) 'A review of Bayesian belief networks in ecosystem service modelling', *Environmental Modelling and Software*. doi: 10.1016/j.envsoft.2013.03.011.

Larson, M. B. and Lockee, B. B. (2009) 'Preparing instructional designers for different career environments: A case study', *Educational Technology Research and Development*. doi: 10.1007/s11423-006-9031-4.

Lefebvre-Ulrikson, W. *et al.* (2016) 'Data Mining', in *Atom Probe Tomography: Put Theory Into Practice*. doi: 10.1016/B978-0-12-804647-0.00009-7.

Li, X., Wang, L. and Sung, E. (2008) 'AdaBoost with SVM-based component classifiers', *Engineering Applications of Artificial Intelligence*. doi: 10.1016/j.engappai.2007.07.001.

Lin, C. T., Lin, C. J. and Lee, C. S. G. (1995) 'Fuzzy adaptive learning control network with on-line neural learning', *Fuzzy Sets and Systems*. doi: 10.1016/0165-0114(94)00195-D.

Lin, R.-F., Korsakul, Nattawadee and Korsakul, Nattharuja (2012) 'Cultural Intelligence as Core Competency to Employability', *The Business & Management Review*.

Lingala, M. *et al.* (2014) 'Fuzzy logic color detection: Blue areas in melanoma dermoscopy images', *Computerized Medical Imaging and Graphics*. doi: 10.1016/j.compmedimag.2014.03.007.

Little, B. M. (2011) 'Employability for the workers - what does this mean?', *Education and Training*. doi: 10.1108/00400911111102360.

Liu, H. *et al.* (2019) 'A Fuzzy Approach to Text Classification with Two-Stage Training for Ambiguous Instances', *IEEE Transactions on Computational Social Systems*. doi: 10.1109/TCSS.2019.2892037.

Liu, X. M. *et al.* (2007) 'A sparse least squares Support Vector Machine classifier', *Moshi Shibie yu Rengong Zhineng/Pattern Recognition and Artificial Intelligence*. doi: 10.1109/ijcnn.2004.1379967.

Loh, W. Y. (2011) 'Classification and regression trees', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. doi: 10.1002/widm.8.

Mahmoudi, S., Lahijan, B. S. and Kanan, H. R. (2013) 'ANFIS-based wrapper model gene selection for cancer classification on microarray gene expression data', in *13th Iranian Conference on Fuzzy Systems, IFSC 2013*. doi: 10.1109/IFSC.2013.6675687.

Maitra, S., Madan, S. and Mahajan, P. (2018) 'An Adaptive Neural Fuzzy Inference System for prediction of student performance in Higher Education', in *Proceedings - IEEE 2018 International Conference on Advances in Computing, Communication Control and Networking, ICACCCN 2018*. doi: 10.1109/ICACCCN.2018.8748869.

Manogaran, G., Varatharajan, R. and Priyan, M. K. (2018) 'Hybrid Recommendation System for Heart Disease Diagnosis based on Multiple Kernel Learning with Adaptive Neuro-Fuzzy Inference System', *Multimedia Tools and Applications*. doi: 10.1007/s11042-017-5515-y.

Marchante, A. J. *et al.* (2003) 'Employer Needs and Graduate Skills: The Gap between Employer Expectations and Job Expectations of Sri Lankan University Graduates', *Journal of Hospitality & Tourism Research*. doi: l;klgl;

Martínez-Cerdá, J. F. *et al.* (2018) 'Opening the black-box in lifelong E-learning for employability: A framework for a Socio-Technical E-learning Employability System of Measurement (STELEM)', *Sustainability (Switzerland)*. doi: 10.3390/su10041014.

Martínez-Martínez, J. M. *et al.* (2016) 'A new visualization tool for data mining techniques', *Progress in Artificial Intelligence*. doi: 10.1007/s13748-015-0079-4.

Mazza, R. and Milani, C. (2005) 'Exploring Usage Analysis in Learning Systems : Gaining Insights From Visualisations', *Time*.

McQuaid, R. W. and Lindsay, C. (2005) 'The concept of employability', *Urban Studies*. doi: 10.1080/0042098042000316100.

Md Razak, M. I. *et al.* (2014) 'Factors Influencing Unemployment among Graduates in Malaysia – An Overview', *Journal of Economics and Sustainable Development*. doi: 10.1016/j.sbspro.2014.01.1269.

Mesiar, R. and Rybárik, J. (1998) 'Entropy of fuzzy partitions: A general model', *Fuzzy Sets and Systems*.

doi: 10.1016/S0165-0114(97)00024-9.

Mikolov, T., Joulin, A. and Baroni, M. (2018) ‘A roadmap towards machine intelligence’, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. doi: 10.1007/978-3-319-75477-2_2.

Minaei-Bidgoli, B. and Punch, W. F. (2003) ‘Using genetic algorithms for data mining optimization in an educational web-based system’, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. doi: 10.1007/3-540-45110-2_119.

‘Mining Association Rules in Students Assessment Data’ (2012) *International Journal of Computer Science Issues*.

Miranda, E. and Suñé, J. (2020) ‘Memristors for Neuromorphic Circuits and Artificial Intelligence Applications’, *Materials*. doi: 10.3390/ma13040938.

Mishra, A., Bansal, R. and Singh, S. N. (2017) ‘Educational data mining and learning analysis’, in *Proceedings of the 7th International Conference Confluence 2017 on Cloud Computing, Data Science and Engineering*. doi: 10.1109/CONFLUENCE.2017.7943201.

Mishra, S., Sahoo, S. and Mishra, B. K. (2018) ‘Neuro-Fuzzy Models and Applications’, in. doi: 10.4018/978-1-5225-5793-7.ch004.

Mishra, T., Kumar, D. and Gupta, S. (2016a) ‘Students’ employability prediction model through data mining’, *International Journal of Applied Engineering Research*. doi: 10.17485/ijst/2017/v10i24/110791.

Mishra, T., Kumar, D. and Gupta, S. (2016b) ‘Students’ employability prediction model through data mining’, *International Journal of Applied Engineering Research*.

Mishra, T., Kumar, D. and Gupta, S. (2017) ‘Students’ Performance and Employability Prediction through Data Mining: A Survey’, *Indian Journal of Science and Technology*. doi: 10.17485/ijst/2017/v10i24/110791.

Modu, B. *et al.* (2017) ‘Towards a predictive analytics-based intelligent malaria outbreakwarning system’, *Applied Sciences (Switzerland)*. doi: 10.3390/app7080836.

Mor, E. and Minguillón, J. (2004) ‘E-learning personalization based on itineraries and long-Term navigational behavior’, in *Proceedings of the 13th International World Wide Web Conference on Alternate Track, Papers and Posters, WWW Alt. 2004*. doi: 10.1145/1013367.1013427.

Mostow, J. and Beck, J. E. (2001) ‘Project LISTEN’s Reading Tutor: Interactive Event Description’, *Reading*.

Muehlenbrock, M. (2005) ‘Automatic Action Analysis in an Interactive Learning Environment’, in *Proceeding of the 12th international conference on artificial intelligence in education (AIED)*.

Nagpal, R., Mehrotra, D. and Bhatia, P. K. (2016) ‘Usability evaluation of website using combined weighted method: fuzzy AHP and entropy approach', *International Journal of Systems Assurance Engineering and Management*. doi: 10.1007/s13198-016-0462-y.

Nauck, D. and Kruse, R. (1997) 'A neuro-fuzzy method to learn fuzzy classification rules from data', *Fuzzy Sets and Systems*. doi: 10.1016/S0165-0114(97)00009-2.

Nauck, D. and Kruse, R. (1999) 'Neuro-fuzzy systems for function approximation', *Fuzzy Sets and Systems*. doi: 10.1016/S0165-0114(98)00169-9.

Nauta, A. *et al.* (2009) 'Understanding the factors that promote employability orientation: The impact of employability culture, career satisfaction, and role breadth self-efficacy', *Journal of Occupational and Organizational Psychology*. doi: 10.1348/096317908X320147.

Nickson, D. *et al.* (2012) 'Soft skills and employability: Evidence from UK retail', *Economic and Industrial Democracy*. doi: 10.1177/0143831X11427589.

Nilakant, K. and Mitrovic, A. (2005) 'Applications of Data Mining in Constraint-Based Intelligent Tutoring Systems', *Proceedings of the artificial intelligence in education, AIED*.

Oliver, B. and Jorre de St Jorre, T. (2018) 'Graduate attributes for 2020 and beyond: recommendations for Australian higher education providers', *Higher Education Research and Development*. doi: 10.1080/07294360.2018.1446415.

Osofisan, A., Adeyemo, O. and Oluwasusi, S. (2014) 'Empirical Study of Decision Tree and Artificial Neural Network Algorithm for Mining Educational Database', *African Journal of Computing & ICT ©*.

Othman, Z. *et al.* (2018) 'Classification techniques for predicting graduate employability', *International Journal on Advanced Science, Engineering and Information Technology*. doi: 10.18517/ijaseit.8.4-2.6832.

Ouguengay, Y. A., El Faddouli, N. E. and Bennani, S. (2015) 'A neuro-fuzzy inference system for the evaluation of reading/writing competencies acquisition in an e-learning environnement', *Journal of Theoretical and Applied Information Technology*.

Oyama, T. *et al.* (1995) 'FINEST: fuzzy inference environment software with tuning', in *IEEE International Conference on Fuzzy Systems*. doi: 10.1109/fuzzy.1995.410055.

Pääkkönen, P. *et al.* (2013) 'The Application of Data Mining to Build Classification Model for Predicting Graduate Employment', *International Journal of Computer Science and Information Security*. doi: 10.1016/j.bdr.2015.01.001.

Pahl, C. and Donnellan, D. (2002) 'Data Mining Technology for the Evaluation of Web-based Teaching and Learning Systems', in *Conference on E-Learning in Business, Government and Higher Education*.

Panda, M. (2018) 'Developing an Efficient Text Pre-Processing Method with Sparse Generative Naive Bayes for Text Mining', *International Journal of Modern Education and Computer Science*. doi: 10.5815/ijmecs.2018.09.02.

Panja, R. and Pal, N. R. (2018) 'MS-SVM: Minimally Spanned Support Vector Machine', *Applied Soft Computing Journal*. doi: 10.1016/j.asoc.2017.12.017.

Pavlov, Y. L. (2019) *Random forests*, *Random Forests*. doi: 10.1201/9780429469275-8.

Pechenizkiy, M. *et al.* (2011) *Proceedings of the International Conference on Educational Data Mining (EDM) (4th, Eindhoven, the Netherlands, July 6-8, 2011)*, *International Working Group on Educational Data Mining*.

Pegrum, M., Bartle, E. and Longnecker, N. (2015) 'Can creative podcasting promote deep learning? The use of podcasting for learning content in an undergraduate science unit', *British Journal of Educational Technology*. doi: 10.1111/bjet.12133.

Peña-Ayala, A. (2014) 'Educational data mining: A survey and a data mining-based analysis of recent works', *Expert Systems with Applications*. doi: 10.1016/j.eswa.2013.08.042.

Pérez Zorrilla, M. (2005) 'Evaluación de la comprensión lectora: dificultades y limitaciones', *Revista de educación*.

Piad, K. C. *et al.* (2016) 'Predicting IT employability using data mining techniques', in *2016 3rd International Conference on Digital Information Processing, Data Mining, and Wireless Communications, DIPDMWC 2016*. doi: 10.1109/DIPDMWC.2016.7529358.

Piad, K. C. (2018) 'Determining the dominant attributes of information technology graduates employability prediction using data mining classification techniques', *Journal of Theoretical and Applied Information Technology*.

Piironen, J. and Vehtari, A. (2017) 'Comparison of Bayesian predictive methods for model selection', *Statistics and Computing*. doi: 10.1007/s11222-016-9649-y.

Pool, L. D., Qualter, P. and Sewell, P. J. (2014a) 'Exploring the factor structure of the CareerEDGE employability development profile', *Education and Training*. doi: 10.1108/ET-01-2013-0009.

Pool, L. D., Qualter, P. and Sewell, P. J. (2014b) 'Exploring the factor structure of the CareerEDGE employability development profile', *Education and Training*. doi: 10.1108/ET-01-2013-0009.

'Proceedings of the International Conference on Soft Computing for Problem Solving, SocProS 2011' (2012) *Advances in Intelligent and Soft Computing*.

Psaromiligkos, Y. *et al.* (2011) 'Mining log data for the analysis of learners' behaviour in web-based learning management systems', *Operational Research*. doi: 10.1007/s12351-008-0032-4.

Pujianto, A., Kusrini and Sunyoto, A. (2018) 'Designing Decision Support System for Scholarship Prediction Using Adaptive Neuro Fuzzy Inference System Algorithm', in *Journal of Physics: Conference Series*. doi: 10.1088/1742-6596/1140/1/012049.

Rahman, N. A. A., Tan, K. L. and Lim, C. K. (2017) 'Predictive analysis and data mining among the employment of fresh graduate students in HEI', in *AIP Conference Proceedings*. doi:

10.1063/1.5005340.

Ramanathan, L., Geetha, A. and Khalid, M. (2015) 'Mining students' record to predict their performance in undergraduate degree', *International Journal of Applied Engineering Research*.

Rathore, R. K. and Jayanthi, J. (2017) 'Student Prediction System For Placement Training Using Fuzzy Inference System', *ICTACT Journal on Soft Computing*. doi: 10.21917/ijsc.2017.0199.

Richman, J. S. (2011) 'Multivariate neighborhood sample entropy: A method for data reduction and prediction of complex data', in *Methods in Enzymology*. doi: 10.1016/B978-0-12-381270-4.00013-5.

Roelofs, R. *et al.* (2019) 'A Meta-Analysis of Overfitting in Machine Learning', *NeurIPS*.

Romero, C. *et al.* (2009) 'Evolutionary algorithms for subgroup discovery in e-learning: A practical application using Moodle data', *Expert Systems with Applications*. doi: 10.1016/j.eswa.2007.11.026.

Romero, C. and Ventura, S. (2007) 'Educational data mining: A survey from 1995 to 2005', *Expert Systems with Applications*. doi: 10.1016/j.eswa.2006.04.005.

Romero, C. and Ventura, S. (2013) 'Data mining in education', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. doi: 10.1002/widm.1075.

Romero, C., Ventura, S. and García, E. (2008) 'Data mining in course management systems: Moodle case study and tutorial', *Computers and Education*. doi: 10.1016/j.compedu.2007.05.016.

Saber Iraji, M. *et al.* (2012) 'Students Classification With Adaptive Neuro Fuzzy', *International Journal of Modern Education and Computer Science*. doi: 10.5815/ijmecs.2012.07.06.

Safavian, S. R. and Landgrebe, D. (1991) 'A Survey of Decision Tree Classifier Methodology', *IEEE Transactions on Systems, Man and Cybernetics*. doi: 10.1109/21.97458.

Sapaat, M. A. *et al.* (2011) 'A classification-based graduates employability model for tracer study by MOHE', in *Communications in Computer and Information Science*. doi: 10.1007/978-3-642-22389-1_25.

Saturi, R., Dara, R. and Prem Chand, P. (2019) 'Extracting subset of relevant features for breast cancer to improve accuracy of classifier', *International Journal of Innovative Technology and Exploring Engineering*. doi: 10.35940/ijitee.K1507.0981119.

Schetinin, V. *et al.* (2007) 'Estimating classification uncertainty of bayesian decision tree technique on financial data', *Studies in Computational Intelligence*. doi: 10.1007/978-3-540-36247-0_6.

Schmidhuber, J. (2015) 'Deep Learning in neural networks: An overview', *Neural Networks*. doi: 10.1016/j.neunet.2014.09.003.

Schnell, B. and Rodríguez, N. (2017) 'Ivory tower vs. workplace reality: Employability and the T&I curriculum–balancing academic education and vocational requirements: a study from the employers' perspective', *Interpreter and Translator Trainer*. doi: 10.1080/1750399X.2017.1344920.

Sedkaoui, S. and Khelfaoui, M. (2019) 'Understand, develop and enhance the learning process with big data', *Information Discovery and Delivery*. doi: 10.1108/IDD-09-2018-0043.

Seigfried, T. (2017) *There's a long way to go in understanding the brain*, *Science News*.

Seising, R. (2007) 'Fuzzifications', *Studies in Fuzziness and Soft Computing*. doi: 10.1007/978-3-540-71795-9_6.

Sewell, P. and Dacre Pool, L. (2010) 'Moving from conceptual ambiguity to operational clarity', *Education + Training*. doi: 10.1108/00400911011017708.

Shana, Z. and Abdulla, S. (2015) 'Educational data mining: An intelligent system to predict student graduation AGPA', *International Review on Computers and Software*. doi: 10.15866/irecos.v10i6.6488.

Shukla, N. (2018) *Machine learning with Tensorflow*, *Manning*. doi: 10.1201/b17476.

Siddique, N. and Adeli, H. (2013) 'Neural Fuzzy Systems', in *Computational Intelligence*. doi: 10.1002/9781118534823.ch10.

Silva, D. R. and Vieira, M. T. P. (2002) 'Using Data Warehouse And Data Mining Resources For Ongoing Assessment Of Distance Learning', *Proceedings of IEEE International Conference on Advanced Learning*.

Sivasankar, E. and Rajesh, R. S. (2012) 'Design and development of a clinical decision support system for diagnosing appendicitis', in *2012 Computing, Communications and Applications Conference, ComComAp 2012*. doi: 10.1109/ComComAp.2012.6154864.

Spasić, I. *et al.* (2014) 'Text mining of cancer-related information: Review of current status and future directions', *International Journal of Medical Informatics*. doi: 10.1016/j.ijmedinf.2014.06.009.

Stern, H. S. (2015) 'Bayesian Statistics', in *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*. doi: 10.1016/B978-0-08-097086-8.42003-9.

Stott, T. (2007) 'Adding value to students in higher education: a 5-year analysis of student attainment of National Governing Body Awards in a UK Outdoor Education Degree Programme', *Journal of Adventure Education & Outdoor Learning*. doi: 10.1080/14729670701750680.

Sun, H. and Hu, X. (2017) 'Attribute selection for decision tree learning with class constraint', *Chemometrics and Intelligent Labouratory Systems*. doi: 10.1016/j.chemolab.2017.02.004.

Syeda Farha Shazmeen, S. F. S. (2013) 'Performance Evaluation of Different Data Mining Classification Algorithm and Predictive Analysis', *IOSR Journal of Computer Engineering*. doi: 10.9790/0661-1060106.

Tair, M. M. A. and El-halees, A. M. (2012) 'Mining Educational Data t o Improve Students ' Performance : A Case Study', *International Journal of Information and Communication technology Research*.

Tane, J., Schmitz, C. and Stumme, G. (2004) 'Semantic resource management for the web: An elearning application', in *Proceedings of the 13th International World Wide Web Conference on Alternate Track, Papers and Posters, WWW Alt. 2004*. doi: 10.1145/1013367.1013369.

Tang, J., Alelyani, S. and Liu, H. (2014) 'Feature selection for classification: A review', in *Data Classification: Algorithms and Applications*. doi: 10.1201/b17320.

Taylan, O. and Karagözoğlu, B. (2009) 'An adaptive neuro-fuzzy model for prediction of student's academic performance', *Computers and Industrial Engineering*. doi: 10.1016/j.cie.2009.01.019.

Thakar, P., Mehta, A. and Manisha (2017) 'A unified model of clustering and classification to improve students' employability prediction', *International Journal of Intelligent Systems and Applications*. doi: 10.5815/ijisa.2017.09.02.

Thijssen, J. G. L., Van Der Heijden, B. I. J. M. and Rocco, T. S. (2008) 'Toward the employability-link model: Current employment transition to future employment perspectives', *Human Resource Development Review*. doi: 10.1177/1534484308314955.

Ting, K. M. (2017) 'Confusion Matrix', in *Encyclopedia of Machine Learning and Data Mining*. doi: 10.1007/978-1-4899-7687-1_50.

Tiwari, A. K. (2017) 'Introduction to machine learning', *Ubiquitous Machine Learning and Its Applications*. doi: 10.4018/978-1-5225-2545-5.ch001.

Toivonen, H. (2017) 'Apriori Algorithm', in *Encyclopedia of Machine Learning and Data Mining*. doi: 10.1007/978-1-4899-7687-1_27.

Tran, T. T. (2015) 'Is graduate employability the "whole-of-higher-education-issue"?', *Journal of Education and Work*. doi: 10.1080/13639080.2014.900167.

Truong, T. T. H., Laura, R. S. and Shaw, K. (2018) 'The Importance of Developing Soft Skill Sets for the Employability of Business Graduates in Vietnam: A Field Study on Selected Business Employers', *Journal of Education and Culture Studies*. doi: 10.22158/jecs.v2n1p32.

Tuma, J. M., & Pratt, J. M. (1982). Clinical child psychology practice and training: A survey. \ldots of Clinical Child & Adolescent Psychology, 137(August 2012), 37–41. http://doi.org/10.1037/a0022390 *et al.* (2011) 'Self-organization of complex, intelligent systems: an action ontology for transdisciplinary integration', *Integral Review*.

137(August 2012), 37–41. http://doi.org/10.1037/a0022390 Tuma J. M. & Pratt J. M. (1982). Clinical child psychology practice and training: A survey. \ldots of Clinical Child & Adolescent Psychology *et al.* (2016) *Detecting diseases in medical prescriptions using data mining tools and combining techniques*, *Iranian Journal of Pharmaceutical Research*. doi: 10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C.

Tymon, A. (2013) 'The student perspective on employability', *Studies in Higher Education*. doi: 10.1080/03075079.2011.604408.

Ueno, M. (2004) 'Online outlier detection system for learning time data in e-learning and its evaluation', in *Proceedings of the Seventh IASTED International Conference on Computers and Advanced Technology in Education*.

Vanhercke, D. *et al.* (2014) ‘Defining perceived employability: A psychological approach’, *Personnel Review*. doi: 10.1108/PR-07-2012-0110.

Vieira, J., Dias, F. and Mota, A. (2004) ‘Neuro-fuzzy systems: a survey’, *… on Neural Networks and Applications, Udine …*.

Vishwakarma, D. K. and Kapoor, R. (2012) ‘Simple and intelligent system to recognize the expression of speech-disabled person’, in *4th International Conference on Intelligent Human Computer Interaction: Advancing Technology for Humanity, IHCI 2012*. doi: 10.1109/IHCI.2012.6481804.

Vysokov, N. V *et al.* (2018) ‘Proteolytically released Lasso/teneurin-2 induces axonal attraction by interacting with latrophilin-1 on axonal growth cones’, *eLife*. doi: 10.7554/elife.37935.

Wager, S. and Athey, S. (2018) ‘Estimation and Inference of Heterogeneous Treatment Effects using Random Forests’, *Journal of the American Statistical Association*. doi: 10.1080/01621459.2017.1319839.

Wald, R., Khoshgoftaar, T. M. and Fazelpour, A. (2013) ‘The use of balance-aware subsampling for bioinformatics datasets’, in *Proceedings of the 2013 IEEE 14th International Conference on Information Reuse and Integration, IEEE IRI 2013*. doi: 10.1109/IRI.2013.6642489.

Walia, N., Singh, H. and Sharma, A. (2015) ‘ANFIS: Adaptive Neuro-Fuzzy Inference System- A Survey’, *International Journal of Computer Applications*. doi: 10.5120/ijca2015905635.

Wang, F. (2002) ‘On Using Data-Mining Technology for Browsing Log file Analysis in Asynchronous Learning Environment’, in *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*.

Wang, Y.-F. and Tsai, C.-T. (Simon) (2014) ‘Employability of Hospitality Graduates: Student and Industry Perspectives’, *Journal of Hospitality & Tourism Education*. doi: 10.1080/10963758.2014.935221.

Watson, I. and Marir, F. (1994) ‘Case-Based Reasoning: A Review’, *The Knowledge Engineering Review*. doi: 10.1017/S0269888900007098.

Wedyan, S. (2014) ‘Review and Comparison of Associative Classification Data Mining Approaches’, *International Journal of Computer, Information, Systems and Control Engineering*.

Williams, A.-M. C. (2015) *Soft skills perceived by students and employers as relevant employability skills*, *ProQuest Dissertations and Theses*.

Winterton, J. and Haworth, N. (2013) ‘Employability’, in *The Transformation of Employment Relations in Europe: Institutions and Outcomes in the Age of Globalization*. doi: 10.4324/9780203743850.

Witten, Frank and Hall (2011) *Data Mining: Practical Machine Learning Tools and Techniques (Google* *eBook)*, *Complementary literature None*.

Wu, W. W. (2011) ‘Mining significant factors affecting the adoption of SaaS using the rough set approach’, *Journal of Systems and Software*. doi: 10.1016/j.jss.2010.11.890.

Wu, X. *et al.* (2014) ‘Data mining with big data’, *IEEE Transactions on Knowledge and Data Engineering*. doi: 10.1109/TKDE.2013.109.

Xu, J. J. (2014) ‘Knowledge discovery and data mining’, in *Computing Handbook, Third Edition: Information Systems and Information Technology*. doi: 10.1201/b16768.

Xu, S. (2018) ‘Bayesian Naïve Bayes classifiers to text classification’, *Journal of Information Science*. doi: 10.1177/0165551516677946.

Yao, Y. *et al.* (2009) ‘A self-constructing fuzzy neural network with new learning algorithm and its control application’, in *Proceedings of the 2009 International Conference on Artificial Intelligence, ICAI 2009*.

Yorke, M. (2005) ‘Employability in higher education: what it is – what it is not’, *Learning & Employability*. doi: 10.1002/ir.162.

Yorke, M. and Knight, P. T. (2006) ‘Curricula for economic and social gain’, *Higher Education*. doi: 10.1007/s10734-004-1704-5.

Yusof, N. *et al.* (2009) ‘Evaluation of student’s performance and learning efficiency based on ANFIS’, in *SoCPaR 2009 - Soft Computing and Pattern Recognition*. doi: 10.1109/SoCPaR.2009.95.

Yusof, N. *et al.* (2012) ‘A Concise Fuzzy Rule Base to Reason Student Performance Based on Rough-Fuzzy Approach’, in *Fuzzy Inference System - Theory and Applications*. doi: 10.5772/37773.

Zadeh, L. A. and Aliev, R. A. (2018) ‘Introduction to Fuzzy Logic Control’, in *Fuzzy Logic Theory and Applications*. doi: 10.1142/9789813238183_0012.

Zaïane, O. R. (2001) ‘Web Usage Mining for a better Web-Based Learning Environment’, *Proceedings of Conference on Advanced Technology for Education*.

Zarandi, M. H. F., Jouzdani, J. and Turksen, I. B. (2007) ‘Generalized reinforcement learning fuzzy control with vague states’, *Advances in Soft Computing*. doi: 10.1007/978-3-540-72432-2_81.

Zhang, L. *et al.* (2010) ‘Classification of polarimetric SAR image based on Support Vector Machine using Multiple-Component Scattering Model and texture features’, *Eurasip Journal on Advances in Signal Processing*. doi: 10.1155/2010/960831.

Zhang, L. (2018) ‘Artificial neural networks model design of Lorenz chaotic system for EEG pattern recognition and prediction’, in *2017 IEEE Life Sciences Conference, LSC 2017*. doi: 10.1109/LSC.2017.8268138.

Zhao, Y. (2013) *R and Data Mining*, *R and Data Mining*. doi: 10.1016/C2011-0-06686-3.

Zwane, F. N., Du Plessis, E. and Slabbert, E. (2017) ‘Learners’ and employers’ perceptions of vocational training in the South African tourism industry’, *African Journal for Physical, Health Education, Recreation and Dance*.

Zwane, F. N., Du Plessis, L. and Slabbert, E. (2014) ‘Analysing employers’ expectations of employee skills in the South African tourism industry’, *SA Journal of Human Resource Management*. doi: 10.4102/sajhrm.v12i1.550.

‘Applications of Data Mining in Higher Education’ (2012) International Journal of Computer Science Issues.

Ayodele, T. O. (2010) ‘Types of Machine Learning Algorithms’, New Advances in Machine Learning. doi: 10.5772/56672.

Bagdasarian, I. et al. (2019) ‘National system of qualifications as a mechanism for ensuring an integrated approach to the development of human capacity in the Yenisei Siberia’, in International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM. doi: 10.5593/sgem2019/5.4/S22.030.

C., H. (1998) ‘What is data science? Fundamental concepts and a heuristic exemple. in Data Science classification, and related methods’, Springer.

Chen, C. P. (2017) ‘Career self-determination theory’, in Psychology of Career Adaptability, Employability and Resilience. doi: 10.1007/978-3-319-66954-0_20.

Choi, K. S., Sun, H. and Heng, P. A. (2004) ‘An efficient and scalable deformable model for virtual reality-based medical applications’, Artificial Intelligence in Medicine. doi: 10.1016/j.artmed.2004.01.013.

Creasey, R. (2013) ‘Improving students’ employability’, Engineering Education. doi: 10.11120/ened.2013.00006.

Desmarais, M. C. and Pelczer, I. (2010) ‘On the faithfulness of simulated student performance data’, in Educational Data Mining 2010 - 3rd International Conference on Educational Data Mining.

Ehret, A. et al. (2015) ‘Application of neural networks with back-propagation to genome-enabled prediction of complex traits in Holstein-Friesian and German Fleckvieh cattle’, Genetics Selection Evolution. doi: 10.1186/s12711-015-0097-5.

Hawkins, D. M. (2004) ‘The Problem of Overfitting’, Journal of Chemical Information and Computer Sciences. doi: 10.1021/ci0342472.

Huang, D., Cabral, R. and Torre, F. Dela (2016) ‘Robust Regression’, IEEE Transactions on Pattern Analysis and Machine Intelligence. doi: 10.1109/TPAMI.2015.2448091.

Jang, H. et al. (2019) ‘An Introduction to Probabilistic Spiking Neural Networks: Probabilistic Models, Learning Rules, and Applications’, IEEE Signal Processing Magazine. doi: 10.1109/MSP.2019.2935234.

Jia, Y. et al. (2019) 'A tree-structured neural network model for household energy breakdown', in The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019. doi: 10.1145/3308558.3313405.

Köppen, M., Wolpert, D. H. and Macready, W. G. (2001) 'Remarks on a recent paper on the "no free lunch" theorems', IEEE Transactions on Evolutionary Computation. doi: 10.1109/4235.930318.

M., J. and A., C. (2016) 'Application of Data Mining Classification in Employee Performance Prediction', International Journal of Computer Applications. doi: 10.5120/ijca2016910883.

Neapolitan, R. E. and Neapolitan, R. E. (2018) 'Neural Networks and Deep Learning', in Artificial Intelligence. doi: 10.1201/b22400-15.

Omar, N. H. et al. (2012) 'Graduates' employability skills based on current job demand through electronic advertisement', Asian Social Science. doi: 10.5539/ass.v8n9p103.

Pimparkar, K. et al. (2019) 'Document Management using Artificial Neural Network', in Proceedings of the 4th International Conference on Communication and Electronics Systems, ICCES 2019. doi: 10.1109/ICCES45898.2019.9002062.

Polamuri, S. (2014) Difference between classification and regression in machine learning, Dataaspirant.

Ramlall, S. (2004) 'A Review of Employee Motivation Theories and their Implications for Employee Retention within Organizations', Journal of American Academy of Business.

Rasul, M. S. et al. (2013) 'Graduate Employability For Manufacturing Industry', Procedia - Social and Behavioral Sciences. doi: 10.1016/j.sbspro.2013.10.739.

Schaap, M. G. and Bouten, W. (1996) 'Modeling water retention curves of sandy soils using neural networks', Water Resources Research. doi: 10.1029/96WR02278.

Sokolova, M. and Lapalme, G. (2009) 'A systematic analysis of performance measures for classification tasks', Information Processing and Management. doi: 10.1016/j.ipm.2009.03.002.

Svozil, D., Kvasnička, V. and Pospíchal, J. (1997) 'Introduction to multi-layer feed-forward neural networks', in Chemometrics and Intelligent Laboratory Systems. doi: 10.1016/S0169-7439(97)00061-0.

Tiwari, A. K. (2017) 'Introduction to machine learning', Ubiquitous Machine Learning and Its Applications. doi: 10.4018/978-1-5225-2545-5.ch001.

Várkonyi-Kóczy, A. R., Tusor, B. and Bukor, J. (2014) 'Improving the model convergence properties of classifier feed-forward MLP neural networks', Studies in Fuzziness and Soft Computing. doi: 10.1007/978-3-319-06323-2_18.

Wolpert, D. H. (1995) No free lunch theorems for search, Technical Report SFI-TR-95-02-010. doi: 10.1145/1389095.1389254.

Kinash, S. et al. (2016) 'Discrepant stakeholder perspectives on graduate employability strategies', Higher Education Research and Development. doi: 10.1080/07294360.2016.1139555.

Maroosis, J. (2018) ‘Analytics’, in Liberal Arts of Management. doi: 10.4324/9781315630670-5.

Sheikh, N. (2013) ‘Using Analytics’, in Implementing Analytics. doi: 10.1016/b978-0-12-401696-5.00003-7.

Siemens, G. and Baker, R. S. J. D. (2012) ‘Learning analytics and educational data mining: Towards communication and collaboration’, in ACM International Conference Proceeding Series. doi: 10.1145/2330601.2330661.