Using Educational Data Mining Techniques in Predicting Grade-4 students’ performance in TIMSS International Assessments in the UAE

استخدام تقنيات استنباط البيانات التعليمية في تقدير أداء طلبة الصف الرابع في تقييمات TIMSS الدولية في دولة الإمارات العربية المتحدة

by

FATEN SHWEDEH

Dissertation submitted in fulfilment

of the requirements for the degree of

MSc INFORMATICS

at

The British University in Dubai

April 2018

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that the content of this research is the direct result of my own work and that any use made in it of published or unpublished copyright material falls within the limits permitted by international copyright conventions.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

Signature of the student

Abstract

Educational Data Mining (EDM) is the process of discovering information and relationships from educational data for better understanding of students' performance, and characteristics of their education providers. Classification is a Data Mining (DM) technique used for prediction. On the other hand, feature selection is the process of finding the best set of features that has the most impact on a specific target.

This dissertation provides an extensive descriptive and predictive analysis on Grade-4 student performance in the Trends in Mathematics and Science Study (TIMSS) in the United Arab Emirates (UAE). The main purpose is to bridge the gap between EDM and International Assessments in the Arab world by applying EDM to predict Grade-4 student levels in TIMSS assessments in the UAE. We examined different feature selection methods and classification algorithms to find the best prediction model with the highest accuracy. The study in this dissertation was expanded to delve deeper into Dubai's private schools data and discover the important features leading to improvements. In addition to building a prediction model to examine if a school will improve in the future TIMSS assessment cycles.

As a result, it was found that the Tree-based feature selection method associated with Decision Tree (DT) classifier built the most accurate prediction models on most TIMSS datasets. The main key factors influencing students' performance in science is discovered and presented.

To the best of our knowledge, this study is the first scientific analysis implementing EDM in the field of international assessments in the UAE. In addition to being the first scientific study that considers all TIMSS questionnaires database in EDM task.

ملخص

إن استنباط البيانات التعليمية (EDM) هي عملية اكتشاف المعلومات والعلاقات من البيانات التعليمية والتي توفر فهماً أفضل لأداء الطلبة وخصائص مزوديهم بالخدمات التعليمة. حيث تعتبر عملية التصنيف من إحدى تقنيات استخراج البيانات (DM) المستخدمة للتنبؤ، بينما تعد عملية "اختيار الميزات" طريقةً للعثور على أفضل مجموعة من الميزات التي تحقق تأثيراً أكبر بناءً على هدفٍ معينٍ.

ستقدم هذه الأطروحة تحليلاً وصفياً وتنبّؤياً واسعاً لمؤشرات مستويات الطلبة في الصف الرابع في دراسة اتجاهات الرياضيات والعلوم (TIMSS) في دولة الإمارات العربية المتحدة. وقد تم تطبيق عمليات استنباط البيانات التعليمية (EDM) للتنبؤ بمستويات طلبة الصف الرابع في تقييمات TIMSS، من خلال إختبار طرق مختلفة في "اختيار الميزات" وخوارزميات التصنيف للعثور على أفضل نموذج تنبؤ يتميز بأعلى قدر من الدقة. وتعمقت هذه الأطروحة في دراسة بيانات المدارس الخاصة بدبي لبناء نموذج للتنبؤ بمستوى تحسن المدرسة في دورات تقييمات TIMSS المستقبلية، بالإضافة إلى اكتشاف الميزات المهمة التي ستؤدي إلى تلك التحسينات.

كما أظهرت نتائج الدراسة أن طريقة اختيار الميزات بخاصية التفرّع (Tree-based) مع طريقة التصنيف بالتفرّع (Decision Tree) معاً يبنيان نماذج التنبؤ الأكثر دقة لمعظم بيانات TIMSS. وكما يقدمان قائمة بأهم الميزات الرئيسة المؤثؤة على نتائج الطلبة.

وعلى حد علمنا، فإن هذه الدراسة تعتبر أول تحليل علمي يستخدم تقنيات استنباط البيانات التعليمية (EDM) في مجال التقييمات الدولية في دولة الإمارات العربية المتحدة، بالإضافة إلى كونها الدراسة الأولى التي تشمل جميع قواعد البيانات الخاصة باستبيانات TIMSS بطريقة استنباط البيانات التعليمية (EDM).

Acknowledgement

First, I thank Allah who gave me the strength to continue my study and to do the best I can in all aspects of my life. A special thanks go to my supervisor, Dr. Sherief Abdullah, who did not hesitate in his support and guidance through the journey of this dissertation.

The biggest thanks to the great hearts I'm surrounded with. My beloved parents, who have been encouraging and supporting me throughout all stages of my life. My sister, who overwhelmed me with her care and love. My brothers, who made my journey full of joy. My sweet sisters in low, who were real sisters and close friends to me. My friends, who were there for me all the time. Last but not least, my colleagues who never hesitated in sharing information.

(Faten Shwedeh)

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Description
TIMSS	Trends in Mathematics and Science Study
PISA	Programme for International Student Assessment
DM	Data Mining
KDD	Knowledge Discovery in Databases
EDM	Educational Data Mining
IEA	International Association for the Evaluation of Educational Achievement
UAE	United Arab Emirates
DT	Decision Tree
CRISP-DM	Cross Industry Standard Process for Data Mining
KNN	K-Nearest-Neighbours
SVM	Support Vector Machine
UFS	Univariate Feature Selection
RFE	Recursive Feature Elimination
OECD	Organisation for Economic Co-operation and Development
NAEP	National Assessment of Educational Progress
UK	United Kingdom
US	United States
MOE	Ministry of Education

1. Chapter One: Introduction

1.1. Overview of International assessments

International assessments are unified learning examinations held in various countries which aim to measure the effectiveness of educational system and students' performance around the world. Each participating country is ranked in order based on the performance of their students' in assessments (Mullis 2012). International Assessments collect information about how students are performing in science and mathematics in different countries. In addition, a plethora of background information is derived via questionnaires covering different aspects of student's life (Martin et al. 2004).

Table 1 shows a comparison between the most common International Assessments in the educational field referring to the *International surveys: PISA, TIMSS, PIRLS* (2015), Hutchison & Schagen (2007), PISA 2015 Results in Focus (2016), and the *International Findings from PIRLS* (2016).

Dataset	TIMSS	PISA	PIRLS
Organization	IEA	OECD	IEA
Year of establishment	1995	2000	2001
First year of UAEs' participation	2007	2009	2011
Assessment cycle	Four years cycle	Three years cycle	Five years cycle
The last assessment	2015	2015	2016
Number of participated countries in the last cycle	57	72	50
Assesses	Mathematics Science	Reading Mathematics Science	Reading
Targeted students	Grade-4 and Grade-8	15 years old students	Grade-4
Associated questionnaires	Student questionnaire Home questionnaire Teachers questionnaire School questionnaire	Student questionnaire Home questionnaire School questionnaire	Student questionnaire Home questionnaire Teachers questionnaire School questionnaire

Table 1: Comparison between different international assessments

1.2. Overview of Educational Data Mining

Data mining and KDD are interchangeable terms used to describe the exploration and analysis of large data sets in order to discover valid, novel and potentially useful patterns of data. They have wide applications across many different industries including the medical and pharmaceutical industries (Padhy 2012). Implementing DM in the field of education is developed to become a separate research field named Educational Data Mining (EDM) with a growing research community. Romero & Ventura (2007) described EDM as an iterative cycle to discover valuable information and guide educators and policy makers (Figure. 1).

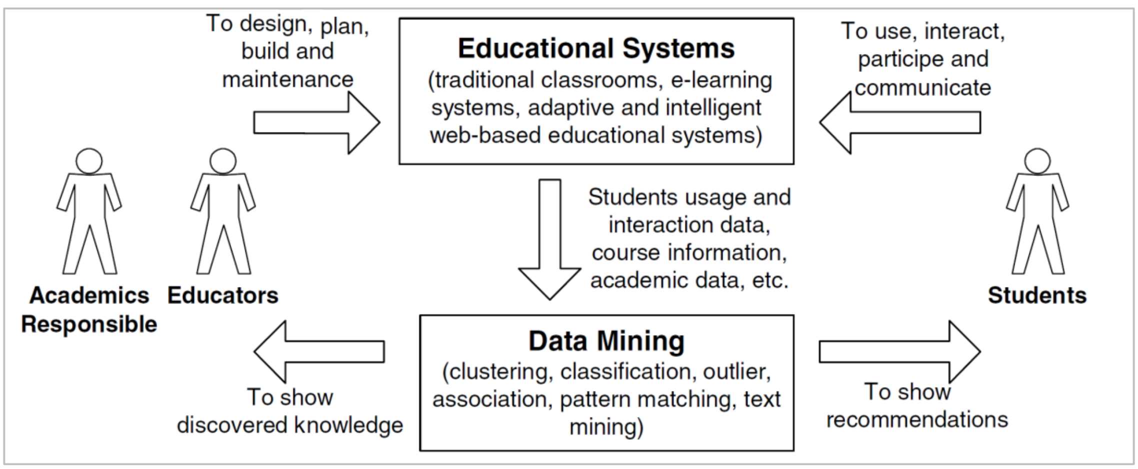


Figure 1: Cycle of applying DM in educational systems (Romero & Ventura 2007)

There are different types of DM tasks. Classification and clustering are commonly used in EDM studies and researches. Classification is a predictive task aiming to discover knowledge and build prediction models from historical data in order to predict the targeted data on new unseen data. Clustering is considered to be a descriptive DM task that restructures the data into clusters based on its similarity (Padhy 2012).

1.3. Motivation

Education is one of the most important sectors in any country. Focusing on understanding and improving educational systems lead to nations' growth. In 2010, His Highness Sheikh Mohammed bin Rashid Al Maktoum released the UAE's National Agenda that expects the UAE to be among the top fifteen countries in TIMSS international assessments by 2021 ("First-Rate Education System" 2018). Several ministries and authorities are working toward achieving this target. An essential step is to understand the current status and discover improvement directives. Using EDM on UAE students' data from the TIMSS database holds great potential to improve student educational achievements. The ability to predict students' future learning patterns and advance scientific knowledge about learning can assist educational institutes across the UAE in targeting factors that leads to improvement.

During the analysis in this dissertation, we found that students' performance in Dubai schools were much better than those achieved by other students within the Emirates. In addition to the presence of schools' inspections data conducted in Dubai's private schools which is published online ("Home | Knowledge and Human Development Authority" 2018). This was the inspiration to extend the study and analyze Dubai's private schools data, combining TIMSS data and inspections data.

The lack of previous studies in international assessments data using EDM was obvious during the literature review which was the source for the motivation to conduct this study.

1.4. Aims and Objectives

In the recent decade, students' data is increasing and there are many factors influencing student outcomes. International assessments provide a huge number of datasets about student' outcomes linked to their home, parents, teachers and principal's background data. Using the traditional method of statistical analysis is an assumption driven process in which the hypothesis is made and then tested based on the data (Romero & Ventura 2007). In contrast, data mining is discovery driven in which the whole data is analysed through machine learning algorithms and the hypothesis automatically extracted.

The main goals of this dissertation is to:

- Provide a comprehensive analysis about Grade-4 students' outcomes in TIMSS – science international assessments in the UAE
- Apply data mining techniques to build a prediction model which predicts students' level in the UAE based on selected features
- Discover the main features influencing student achievements in science assessments
- Apply data mining techniques to build a prediction model for Dubai's private schools to predict if a school will improve between the two TIMSS assessments cycles
- Discover the main features leading to improving TIMSS assessments scores in private schools in Dubai
- Bridge the gap between EDM and International Assessments in the Arab world

1.5. Uniqueness of the study

The contribution of the study conducted in this dissertation is listed below:

- First research in the UAE using EDM in international assessments
- First research linking teachers attributes from TIMSS with inspection results in Dubai's private schools
- First study to use feature selection methods in TIMSS assessment data
- First study that looks deeply into what happens in class rooms and links it to TIMMS scores in Dubai's private schools

1.6. Research Questions

The main purpose of this dissertation is to answer the following questions:

- Does the use of feature selection methods improves the accuracy of predicting Grade-4 students' performance in science?
- Which classification method has the best performance in predicting Grade-4 students' achievement in the UAE?
- What are the key factors impacting on Grade-4 students' science achievement in the UAE?
- What are the key factors affecting the improvement of Grade-4 students in science at Dubai private schools?

1.7. Research methodology

The research methodology conducted throughout this dissertation is a mixture of quantitative and qualitative methods. We started by conducting unstructured interviews with experts of international assessments and international assessments data managers in the UAE. This step was very useful to build the basic thought about international assessments conducted in the UAE. Then followed the quantitative research methodology in which the analysis and prediction model was built on. Cross Industry Standard Process for Data Mining (CRISP-DM) process described by Olson and Delen (2008) was followed in conducting a comprehensive Data Mining analysis in this dissertation. CRISP-DM process contains six steps for data mining task as presented in Figure 2.

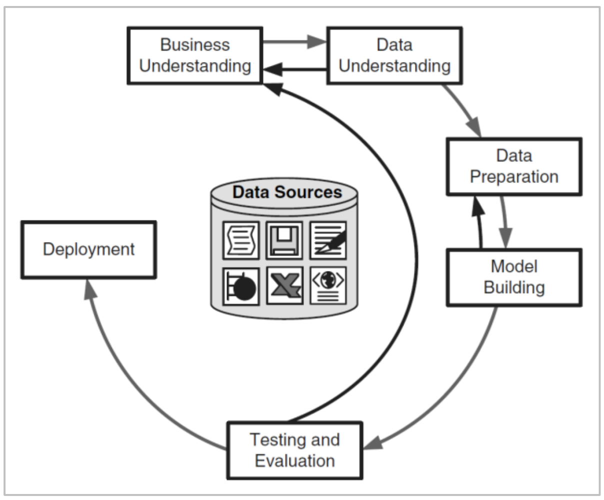


Figure 2: CRISP-DM process (Olson & Delen 2008)

1- Business understanding: is exploring the concerned domain of the study and defining the objectives of the study.
2- Data understanding: is the process of data collection, description and exploration.
3- Data preparation: is the process of cleaning the data and transforming it to the desired format.
4- Model building: contains two processes. First, visualizing the data and showing relationships as an initial analysis for greater data understanding. Second, building data mining model and using the data collected to train and test the model.
5- Testing and evaluation: is the process of evaluating the model and make sure it serves the objective of the data mining task.
6- Deployment: is a continues process of monitoring the changes, analysing and applying data mining tasks to insure it's compatibility with changes over the time.

Figure 3 shows the methodology we propose in this dissertation aiming to build an accurate predictive model using TIMSS 2015 Grade-4 data. First, we start with data pre-processing which prepares the data and forms it in a way to increase the accuracy of the prediction model.

Then, we applied three different feature selection techniques: 1) Filter feature selection method - Univariate Feature Selection (UFS) with SelectKBest (chi2) for scoring features, 2) Wrapper feature selection method - Recursive Feature Elimination (RFE) with Naïve Bayes for search, and 3) Tree-based feature selection – ExtraTreeClassifier. The purpose of using feature selection methods is to reduce the number of features considered in building the model. More details of feature selection methods are described in the background section (2.1.1).

In this dissertation, we experimented with three classifiers: Decision tree (DT), K-Nearest-Neighbours (KNN), and Support Vector Machine (SVM). More details of classifiers are described in the background section (2.1.2). Each classifier was examined three times using different set of features selected by the different selection methods. In addition, we used the original dataset which contains all features to compare results before and after feature selection.

As an evaluation approach, the standard evaluations measure 'Accuracy' was used to measures the performance of the prediction model. We used 10-fold Cross validation technique which is used to evaluate the prediction model. The technique is based on partitioning the original dataset into ten subsets, nine are used to train the data model and one is used for testing the model. This procedure is repeated ten times by shifting the testing subset, and calculate the accuracy of each cycle. As a result, the average accuracy is calculated.

We measured the average accuracy and execution time of feature selection methods on each dataset to understand the behaviour of each method on the different datasets. Then weI used the three types of classifiers with each feature selection method and created a scheme of (feature selection method \ classifier). Based on classification average accuracy along with the computation timing, the best scheme was selected for prediction. Confusion matrix of the best schemes selected is displayed to describe the performance of the classification model for each dataset. More explanation of evaluation measures and confusion matrix is described in the background section (2.1.3).

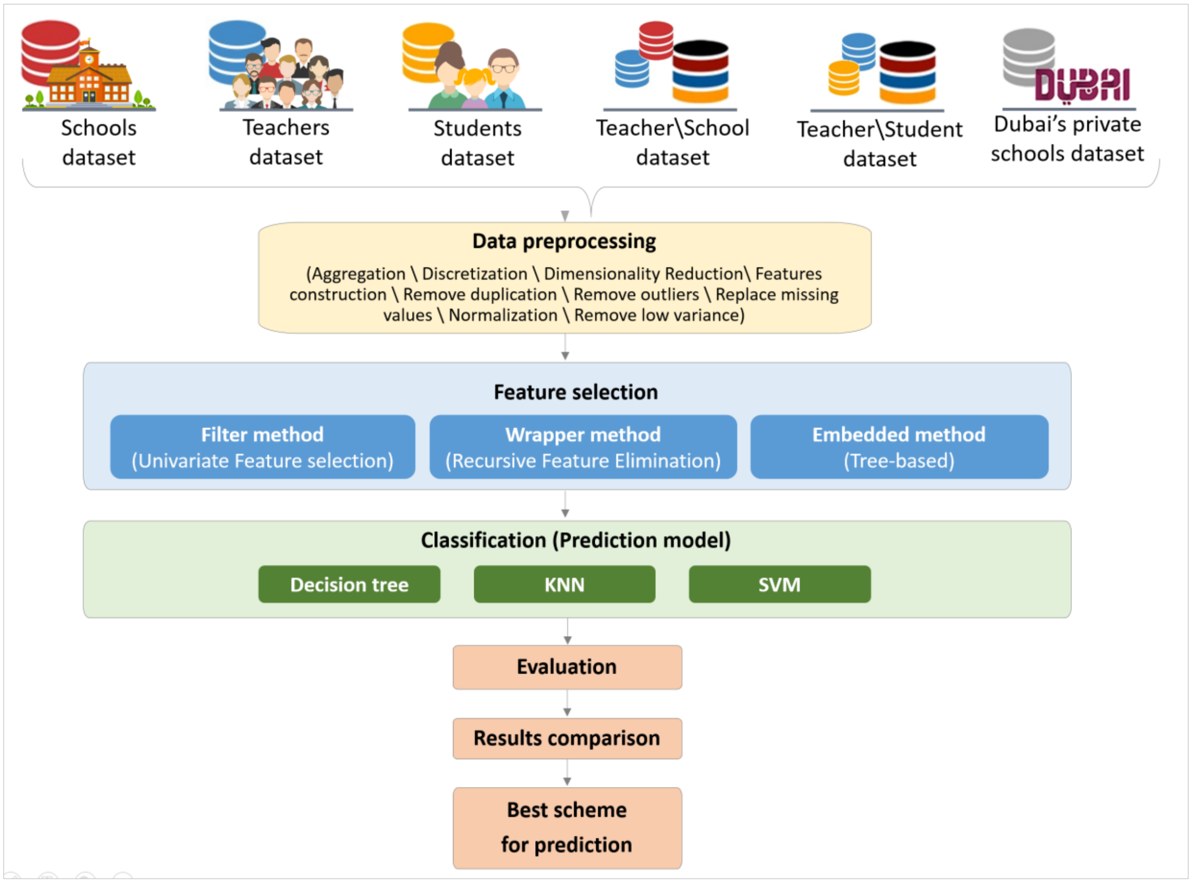


Figure 3: Experimental setup overview structure

1.8. Tools

There are many data mining tools and applications used to apply data mining tasks and machine learning algorithms. Applications vary in the facilities provided for modelling techniques and results' visualization (Leventhal 2010). Selecting an appropriate tool plays a major role in the research process. In this dissertation, we used Jupyter which is a web-based application that allows running IPython code (Nielsen 2017). Python is an object-oriented programming language that has large number of modules and handles manipulating big data in a short time compared to other applications (Layton 2017). Another application used in this dissertation is IEA-IDB analyser which is an application developed by the IEA for merging and analysing TIMSS database files. We used IDB analyser to build the different datasets used in this study. In addition, SPSS and excel were used for basic data analysis and charts building.

1.9. Dissertation Structure

The rest of this dissertation is structured as follows: Chapter Two which gives a background on previous related work in the field. Chapter Three describes the process of data collection and the different datasets created in this study. Chapter Four provides broad and deep descriptive analysis of students' outcomes from different aspects. Chapter Five explains the predictive analysis process with the findings. Finally, chapter six which concludes this dissertation by providing answers to all research questions and exploring possible future work.

2. Chapter Two: Literature Review

2.1. Background

This chapter provides background about feature selection, data mining, and evaluation methods. In addition to describing related work done in feature selection and EDM on international assessments data.

2.1.1. Feature selection

In high dimensional data, there might be features which do not add valuable information to the prediction model. There are several feature selection methods aiming to reduce the number of features used in machine learning tasks. Feature selection is considered to be a pre-processing task for dimensionality reduction. In this section we describe feature selection methods used in this study to reduce the number of features considered in building the prediction model. As a result of each technique, a subset of features is selected to pass to the process of building the prediction model described in section (5.4). Referring to the work of Saeys, Inza & Larranaga (2007) and the work of Chandrashekar & Sahin (2014), Table 2 explains the three main methods, and Table 3 compares filter and wrapper methods.

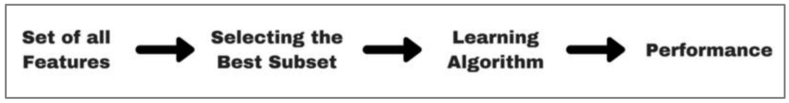
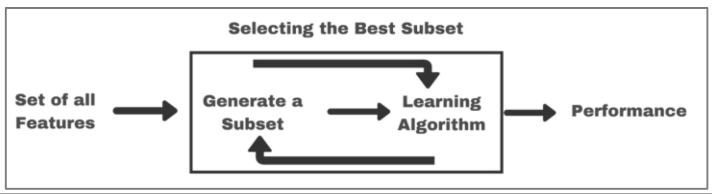
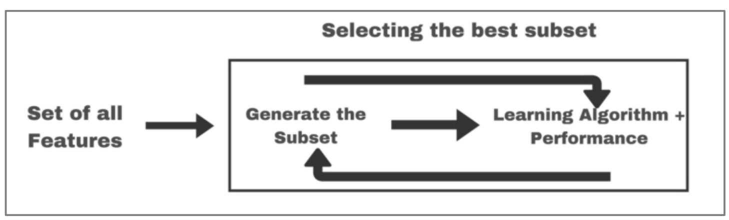
Feature selection method	Description	Example
Filter Method	Filter methods look at all features and score each feature based on its' relevance with the targeted feature. The low scoring features are excluded. As a result, the best subset with the top features is returned. 	Univariate Feature Selection (UFS)
Wrapper Methods	Wrapper methods search for possible feature subsets. The subset is used to train and test a classification model. Based on that, features are added or removed from the subset. To search for all subsets in the data, search algorithm is 'wrapped' around the classification model used. 	Recursive Feature Elimination (RFE)
Embedded Methods	Embedded methods combine both filter and wrapper methods. The search for the best subset of features is embedded with the structure of the classifier. 	Tree-based feature selection

Table 2: Feature selection methods

	Filter method	Wrapper method
Measurement	Measures the correlation between feature and	Measures the effectiveness of features subset
Speed	Faster	Slower
Dependency	Independent of machine learning algorithms	Dependent of machine learning algorithms
Success	Likely fails to find the optimal subset	Always provides the optimal subset
Overfitting	Unlikely overfitting occurrence	Prone to overfitting

Table 3: Comparison between filter and wrapper feature selection methods

2.1.2. Classification (prediction model)

Classification is a supervised data mining technique used to build a prediction model aiming to accurately predict categorical class labels in a dataset. Classification uses two types of data sets, training dataset and testing dataset. Training dataset is used to build the prediction model using given features in addition to the known targeted value. Testing dataset (unseen by the classifier) and contains same data features except the targeted attribute to be predicted. The classifier analyses the input data and predicts the targeted value. The prediction accuracy is calculated based on results which indicates the goodness of the classification model.

There are different classifiers are commonly used in prediction. Each has its own technique and the performance of each depends on the nature of dataset.

Decision tree (DT)

Decision tree is defined as classification algorithm (classifier) that recursively partitions dataset into sub-divisions based on several defined tests on each branch or (node). Decision tree is considered to be a visualization technique for classification model. The resulted tree is composed of root node that corresponds to the first feature used to start deciding the dataset. Internal nodes (splits) include the recursion process for splitting the dataset into sub-sets and terminal nodes (leafs) indicate to labelled classes for the input feature. ID3 is the core algorithm that is used for building the decision tree, it employs top-down and greedy search to split the tree branches and backtracking. ID3 uses Entropy and Information Gain to construct a decision tree. Entropy is partitioning the dataset based on the homogenous values into sub sets and build the top-down tree. Information Gain based on decreasing the Entropy for the split dataset it’s all about creating the tree based on the highest information gain. In the experiment, we created a function that creates the prediction model using both Entropy and Information Gain, then returns the classification model with the heist accuracy among them.

K-Nearest-Neighbours (KNN)

K-Nearest-Neighbours is a type of instance-based learning (lazy learning) that uses similarity measuring to classify new case. The similarity is measured using distance functions between the training objects and the testing objects.

Support Vector Machine (SVM)

Support Vector Machine is a classification algorithm that finds the hyperplane which maximize the margin between two classes. Intuitively SVM finds the optimal hyperplane by maximizing the margin width and operates the vectors into two non-overlaying classes. SVM can perform the Nonlinear classification by finding the hyperplane that maximize the margin and minimize the misclassifications.

2.1.3. Evaluation

As a common evaluation measure used in data mining tasks, ‘Accuracy’ measures how close the predicted value to the standard known value. The higher the accuracy value of the prediction model, the better is the model to predict.

The equation below represents accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

TP: True Positive

TN: True Negative

FP: False Positive

FN: False Negative

Cross validation technique is used to evaluate the prediction model. The technique is based on partitioning the original dataset into ten subsets, nine are used to train the data model and one is used for testing the model. This procedure is repeated ten times by shifting the testing subset, and calculate the accuracy of each cycle. As a result, the average accuracy is calculated.

Confusion matrix of the best schemes selected is displayed to describe the performance of the classification model for each dataset. It shows the number of correct and incorrect numbers predicted by the classification model compared to the real numbers. Figure 4 shows the structure of confusion matrix.

		Actual class	
Predicted class		True	False
	Positive	TP	FN
	Negative	FP	TN

Figure 4: Confusion matrix

Precession is another measure used in evaluating prediction models. Precession represents the proportion of corrected predicted answers among all predicted answers. It is calculated based on the below formula:

$$Precision = \frac{TP}{TP + FN}$$

Figure 5 visualise the difference between accuracy and recall ("Accuracy and precision" 2018).

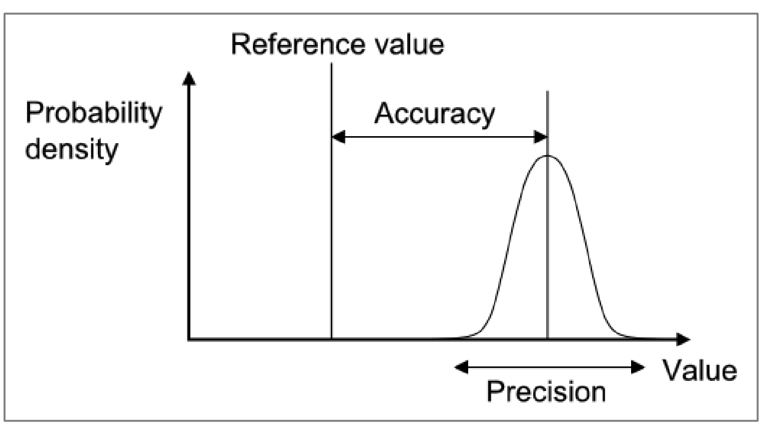


Figure 5: Accuracy and Precission ("Accuracy and precision" 2018)

2.2. Related work

This section provides a brief of related work done previously applying EDM on international assessments.

Paper-1	
Paper name	Identifying the Factors Affecting Science and Mathematics Achievement Using Data Mining Methods
Year of publication	2015
Number of citations	17
Country of focus	Turkeya (focus on Turkish students' scores)
International assessment	TIMSS 1999, PISA 2003 and PISA 2006
Data size	8 features (attributes) 2 predictable variables 7,163 records
Criteria of feature selection	Manual, focused on students' scores mathematics and science in addition to the subdomains of each. Questionnaires background data was not used.
DM task	classification technique (Decision tree), clustering
Purpose of the study	Kiray, Gok & Bozkir (2015) worked on a study aiming to specify the order of importance of attributes affecting middle school students' outcomes in mathematics and science.
Key results	The study resulted that: - reading and problem-solving skills affected both mathematics and science; - mathematics achievement affected science achievement; - science achievement affected mathematics achievement; - Science and mathematics are affected by the same variables.

Paper-2	
Paper name	Using Data Mining to Predict K–12 Students' Performance on Large-Scale Assessment Items Related to Energy
Year of publication	2008
Number of citations	31
Country	US
Data source	TIMSS 1995, TIMSS-Repeat 1999, 2003 TIMSS National Assessment of Educational Progress (NAEP)
Data size	76 features
Criteria of feature selection	Selection of questions related to the field of energy, then building data cubes and exclude variables with low variance

DM task	Prediction using Linear function and Decision Tree (C4.5 and M5 algorithms)
Purpose of the study	Liu & Ruiz (2008) worked on a study using data mining to predict the performance of students from Kindergarten to Grade-12 in the field of energy
Key results	- It is possible to predict the performance of students (from Kindergarten to Grade-12) on assessments question related to energy. - The key predictors are cognitive demand, energy content, context, and grade level.
	Paper-3
Paper name	Identifying the Classification Performances of Educational Data Mining Methods: A Case Study for TIMSS
Year of publication	2017
Number of citations	-
Country	Turkeya (focus on Turkish students' scores)
Data source	TIMSS 2011 (Grade-8 students) – Mathematics data
Data size	12 feature 6,928 students' records
Criteria of feature selection	manual selected features
DM task	Classification (Decision tree, a Bayesian network, a logistic regression, and neural networks)
Purpose of the study	Kılıç Depren, Aşkın & Öz (2017) worked on a study to predict Grade-8 students performance in Mathematics in TIMSS assessments
Key results	- Logistic regression shows the best accuracy in predicting students' performance. - Students' level of confidence found to have the most impact on Grade-8 students performance in mathematics.
	Paper-4
Paper name	A Review on Predicting Student's Performance Using Data Mining Techniques
Year of publication	2015
Number of citations	83
Country	Malaysia
Data source	IEEE Xplore, Spinger Link, Science Direct, ACM digital Library
Data size	12 attributes
Criteria of feature selection	manual selected features
DM task	Classification (Decision tree, Neural network, KNN, SVM)

Purpose of the study	Shahiri, Husain & Rashid (2015) worked on a study to identify the important attributes affecting students' performance in Malaysia. In addition to finding the best prediction method.
Key results	Neural Network and Decision Tree achieve the heist accuracy in predicting students' performance.
Paper-5	
Paper name	Comparing the Predictive and Classification Performances of Logistic Regression and Neural Networks: A Case Study on Timss 2011
Year of publication	2013
Number of citations	6
Country	Turkeya
Data source	TIMSS 2011 – Grade-8 - Mathematics
Data size	6 features 6,928 records
Criteria of feature selection	manual selected features
DM task	Predictive (Logical regression), Classification (Neural network)
Purpose of the study	Askin & Gokalp (2013) worked on a study to compare the performance of logical regression and neural network in predicting Grade-4 students' performance and indicate the level of variables impact on Grade-8 mathematics scores in TIMSS assessments
Key results	- Similar performance of logical regression and neural network in predicting Grade-4 students' performance. - Students' level of confidence found to have the most impact on Grade-8 students performance in mathematics.

Table 4: Comparison of previous work in EDM and international assessments

Feature selection is a process used for data dimensionality reduction and select subset of data variables to be considered in building prediction models. Feature selection is considered to be a pre-processing task aiming to improve the accuracy of prediction model. As per to the research done on EDM using international assessments data, we could not find previous studies used feature selection methods. Therefore, it was useful looking at data mining studies implementing feature selection techniques in other fields. Table 3 presents feature selection related work with data mining in other fields (mechanical engineering, and spam detection).

Paper-1	
Paper name	Bearing fault diagnosis using multiclass support vector machine with efficient feature selection methods
Year of publication	2015
Number of citations	3
Field of study	Mechanical & Mechatronics Engineering
Data size	40 features and 128,000 records
Criteria of feature selection	Feature selection methods: - VM–Recursive Feature Elimination (SVM-RFE) - Wrapper subset method - ReliefF meth-od and Principle component analysis (PCA)
DM task	Classification, Multiclass Support Vector Machines (MSVM) and C4.5 (J48) with different feature selection methods
Purpose of the study	Rajeswari et al. (2015) worked on a study to predict bearing faults using data mining techniques and feature selection methods aiming to help in bearing fault diagnoses and avoid to reduce economic loss.
Key results	- Feature selection improves the accuracy of classifier. - MSVM along with Wrapper subset feature selection showed the best accuracy in predicting bearing faults.

Paper-2	
Paper name	Decision Tree with Optimal Feature Selection for Bearing Fault Detection
Year of publication	2008
Number of citations	8
Field of study	Mechanical & Mechatronics Engineering
Data size	18 features and 1,464 records
Criteria of feature selection	Feature selection methods: - Genetic Algorithm (GA) – introduced in the paper - Principle component analysis (PCA)
DM task	Classification (Decision Tree)
Purpose of the study	Nguyen & Lee (2008) worked on a study that aims to detect bearing faults by using decision tree. They introduced a way for feature selection called Genetic Algorithm (GA), and compared it with PAC feature selection to check which perform better in.
Key results	- GA feature selection performed better in terms of the simplicity of the resulted decision tree - PCA performed better in terms of the accuracy of the resulted decision tree - Using feature selection reduces the complexity and increase the efficiency of the resulted decision tree

Paper-3	
Paper name	Binary PSO with mutation operator for feature selection using decision tree applied to spam detection
Year of publication	2014
Number of citations	162
Field of study	Spam emails
Data size	57 features 6,000 records (emails)
Criteria of feature selection	Wrapper feature selection methods (different algorithms): - Iterated Local Search (ILS) - Genetic Algorithms (GA) - Restarted SA (RSA) - Ant Colony Optimization (ACO) - Article Swarm Optimization (PSO) - Binary PSO (BPSO) - Modified discrete Binary PSO (MBPSO)
DM task	Classification: decision tree (C4.5 algorithm)
Purpose of the study	Zhang et al. (2014) worked on a study to reduce the error cases of labelling non-spam email as spam by using data mining and feature selection methods to create spam detection model.
Key results	- Using Wrapper feature selection increases the classification accuracy - MBPSO wrapper feature selection algorithm gives better classification results than other wrapper algorithms.

Table 5: Comparison of previous work in feature selection

3. Chapter Three: Building Datasets

This chapter describes the process of data collection used in this study and the original features in addition to the datasets created for conducting analysis and extracting experiment results.

3.1. Data collection

TIMSS international assessments database is available online ("TIMSS 2015 International Database" 2018) as an open data for research purpose. TIMSS 2015 database contains data for all the 47 countries who participated in TIMSS 2015. Figure 6 shows the different data available in two formats, SPSS and SAT. The full description of TIMSS database is available in TIMSS 2015 User Guide (TIMSS 2015 User Guide for the International Database 2017).

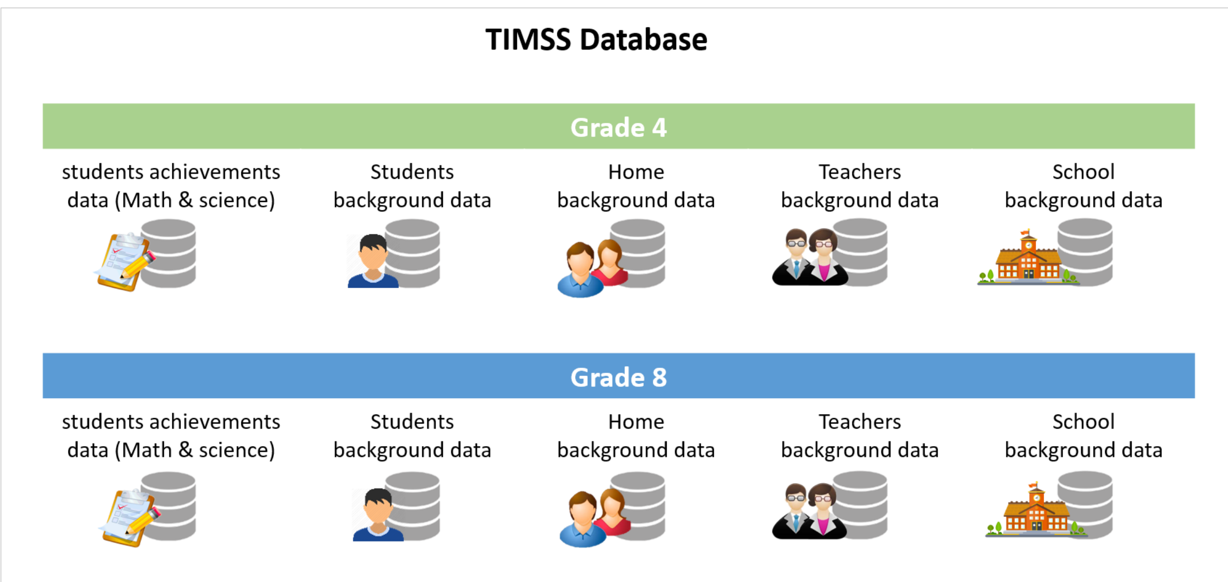


Figure 6: Data merging files

In this dissertation, we focused on Grade-4 students' data in the UAE. In addition to TIMSS data, we collected data about schools inspections conducted in Dubai to link it with Dubai's TIMSS data.

3.2. Datasets construction

TIMSS database is available in separate files. For example, students' scores of each question in the exam is separate than students data collected from the background questionnaire. We used IEA-IDB Analyser to merge TIMSS SPSS data files and create the desired datasets. We created five different datasets for UAE, and two datasets for Dubai's private schools data. Table 6 shows TIMSS data files considered in the construction of the different datasets.

	Dataset	Merged data files
1	UAE students dataset	Achievements data + Students data + Home data
2	UAE teachers dataset	Achievements data + Teachers data
3	UAE schools dataset	Achievements data + Schools data
4	UAE Teachers\Students dataset	Achievements data + Students data + Home data + Teachers data
5	UAE Teachers\Schools dataset	Achievements data + Teachers data + Schools data
6	Dubai classrooms dataset	Achievements data + Students classroom related + Teachers data (classroom)
7	Dubai TIMSS\Inspection dataset	Achievements data + Students data + Home data + teachers data + Schools data + Inspection data

Table 6: Datasets constructions

3.3. Datasets description

TIMSS data files contains different type of features (attributes). This section describes the structure of the different datasets constructed in this study. Table 7 shows the details of datasets dimensions as resulted after merging TIMSS Grade-4 datasets.

	Dataset	Number of rows\ cases	Number features	Number of feature by category
1	UAE students dataset	21,177	766	IDs = 93 Exam questions score = 419 Background features = 254
2	UAE teachers dataset	37,647	742	IDs = 39 Exam questions score = 419 Background features = 284
3	UAE schools dataset	21,177	562	IDs = 40 Exam questions score = 419 Background features = 103
4	UAE Teachers\Students dataset	37,647	1,010	IDs = 53 Exam questions score = 419 Background features = 538
5	UAE Teachers\Schools dataset	37,647	856	IDs = 50 Exam questions score = 419 Background features = 387
6	Dubai classrooms dataset	6,466	48	IDs = 19 Exam questions score = 4 Background features = 61
7	Dubai TIMSS\Inspection dataset	6,466	495	IDs = 26 Exam questions score = 4 Background features = 420 Inspection features = 45

Table 7: Datasets dimensions

As an initial task of data cleaning was done to exclude variables which do not relate to the scope of this study were excluded. The list below shows the excluded variable groups:

- Examination question score (which shows the score of each question in the exam).
- Mathematics exam questions.

Table 8 shows datasets dimensions after the initial data cleaning task.

	Dataset	Number of rows\ cases	Number features	Number of feature by category
1	UAE students dataset	21,177	262	IDs = 3 Exam questions score = 5 Background features = 254
2	UAE teachers dataset	37,647	293	IDs = 4 Exam questions score = 5 Background features = 284
3	UAE schools dataset	21,177	110	IDs = 2 Exam questions score = 5 Background features = 103
4	UAE Teachers\Students dataset	37,647	548	IDs = 5 Exam questions score = 5 Background features = 538
5	UAE Teachers\Schools dataset	37,647	396	IDs = 4 Exam questions score = 5 Background features = 387
6	Dubai classrooms dataset	6,466	71	IDs = 5 Exam questions score = 5 Background features = 61
7	Dubai TIMSS\Inspection dataset	6,466	474	IDs = 4 Exam questions score = 5 Background features = 420 Inspection features = 45

Table 8: Datasets dimensions after initial data cleaning

4. Chapter Four: Descriptive analysis

In this section, we visualise the overview analysis and then show segmented analysis, which is crucial for better understanding of the data. Simply, this section shows ‘What happened’ and builds the historical perspective on the TIMSS2015 data in Grade-4 in the UAE. We looked at data from different angles, each angle is described in a subsection as follows: (4.1) Cross-subject analysis, (4.2) Cross-Emirate analysis, (4.3) Cross-school-type analysis, (4.4) Cross-curriculum analysis, (4.5) Cross-gender analysis, (4.6) Cross-age analysis, (4.7) Cross-educational-qualification analysis, and finally (4.8) Specific analysis on Dubai’s private schools.

4.1. Cross-subject analysis

This section shows analysis conducted to compare students’ achievements and scores in TIMSS2015 in both subjects (science and mathematics). Figure 7 shows students who got scores in the lower performing levels (Low and Below low) are more than students with higher performing levels (High and Advanced). This Applies to both subjects, science and mathematics.

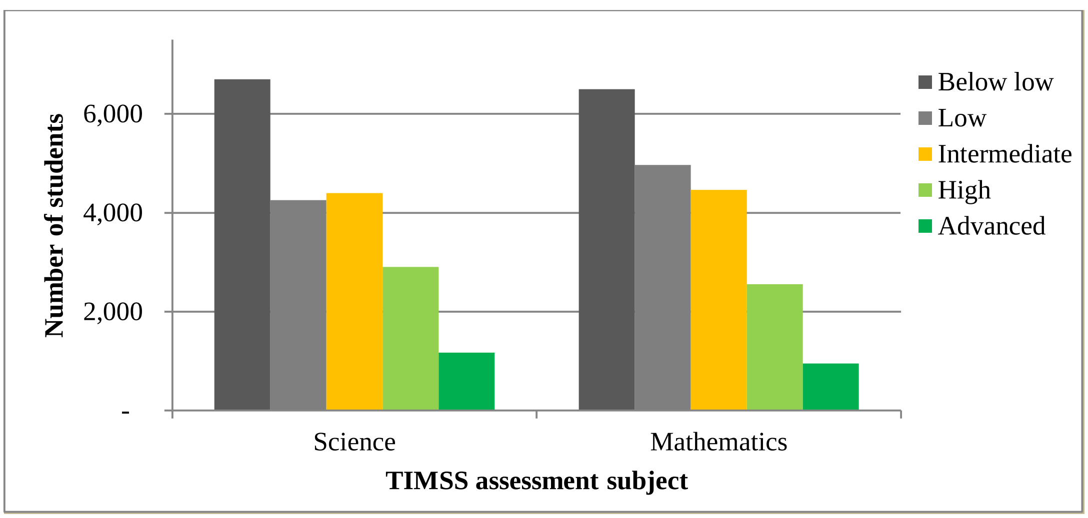


Figure 7: Students distribution in TIMSS benchmarks levels in science and mathematics

Figure 8 shows histograms of students' scores in science and mathematics. The x-axis represent students frequencies, and the y-axis represent number of student in each score obtained. Overall, students' scores in sience and mathematics are similar. Figure 9 shows boxplots with the distribution of students' scores based on the minimum, first quartile, median, third quartile, and maximum. Although the variety of students' scores in science, overall view of students' scores are similar in both subjects.

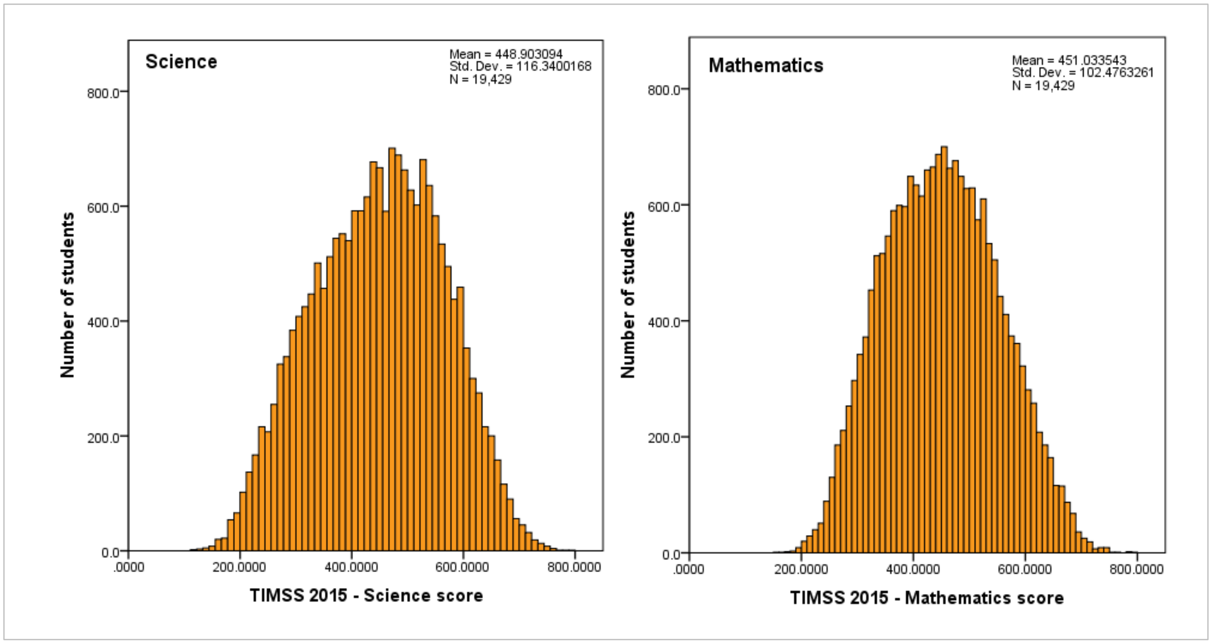


Figure 8: Students frequencies based on their TIMSS scores in science and mathematics

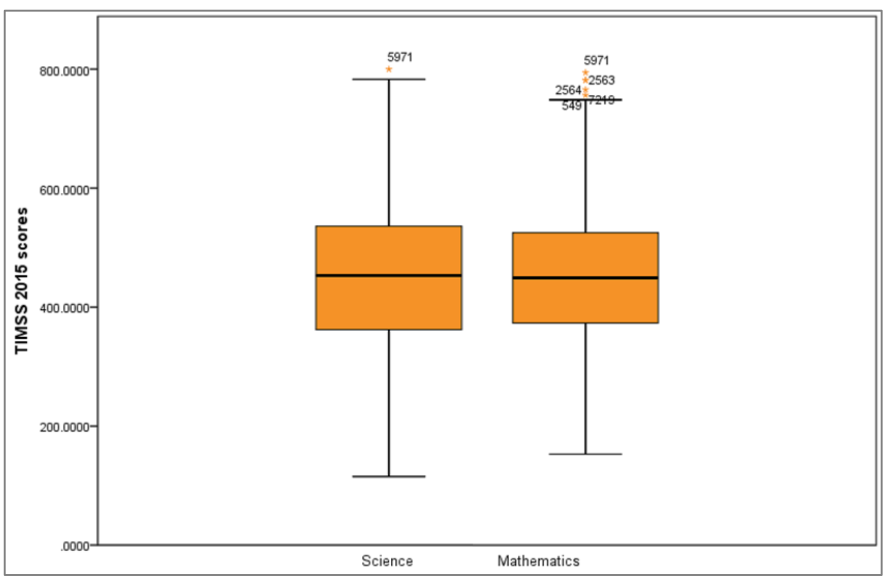


Figure 9: Boxplot of Grade-4 students' scores in science and mathematics

Looking at students' scores in both sujects in another view, Figure 10 and Firure 11 show a high correlation between students' performance in science and mathematics. This means that a student with low score in science got low score in mathematics.

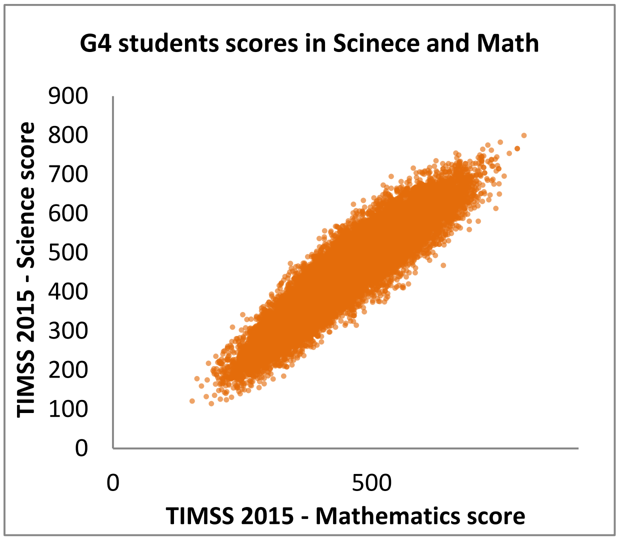


Figure 10: Grade-4 students scores in TIMSS2015

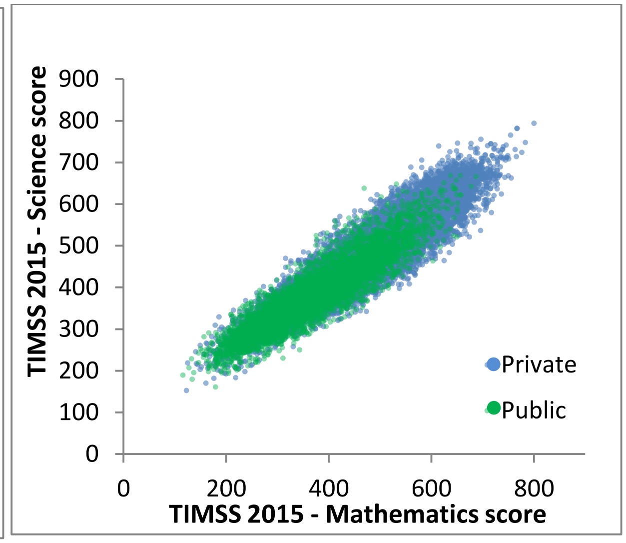


Figure 11: Grade-4 students' scores by school type

Looking deeper into students' scores by school types (public and private), Figure 12 shows that although students in private schools got higher scores than students in public schools, students' performance is similar in science and mathematics within each school school type.

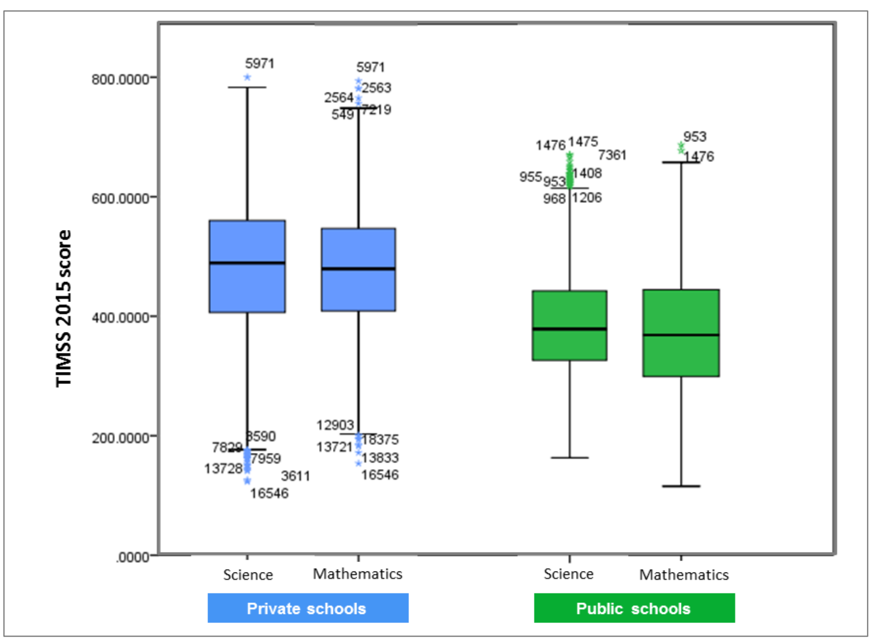


Figure 12: Boxplot of Grade-4 students' scores in science and mathematics by school type

As a result of this section, Grade-4 students' scores in science and mathematics are similar. This applies when looking at all students within different school types. Therefore, in this dissertation we chose to focus on students' scores in science, as it is the subject of focus in TIMSS 2015. The rest of this dissertation refers to Grade-4 students' science scores in measuring and comparing their performance.

4.2. Cross-Emirate analysis

Another angle to describe the data is looking at students outcomes in science in each Emirate in the UAE. Figure 13 shows the number of Grade-4 students who participated in TIMSS 2015 science assessments across different Emirates in the UAE. The highest number of students participated are from Dubai schools.

Figure 14 presents combined graphs showing students' performance in each Emirate. A trend toward lower levels in students' scores is clear in all Emirates other than Dubai, where the story is different. In Dubai, students in 'Intermediate', 'High' and 'Advanced' levels are more than other levels. This shows students in Dubai perform better than their peers in the other Emirates.

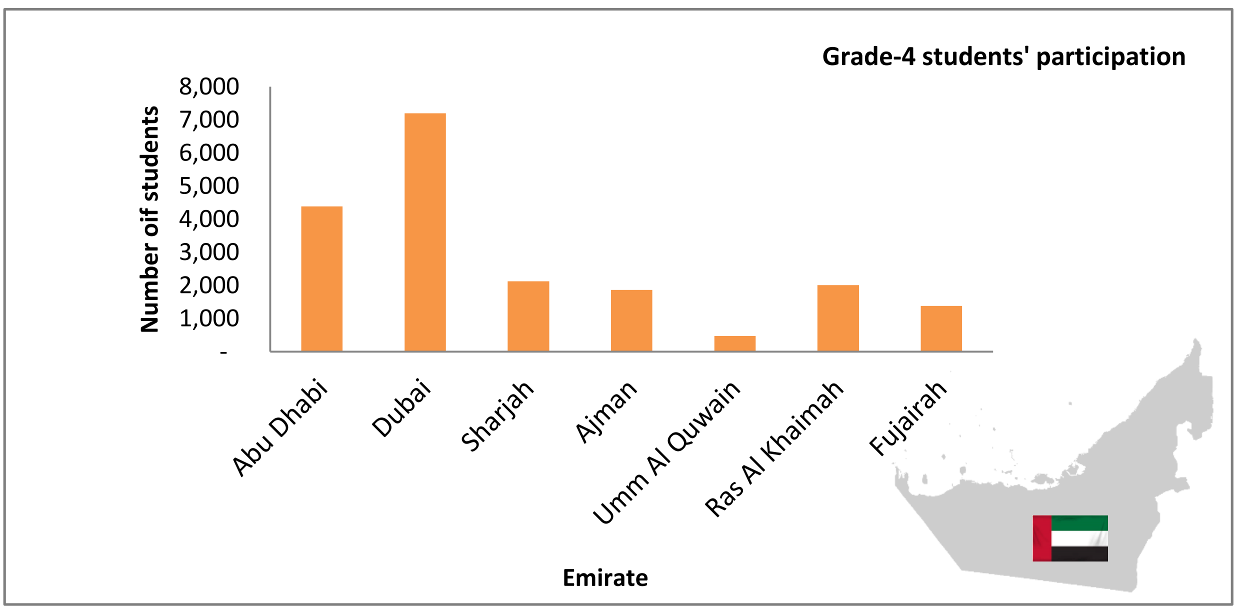


Figure 13: Grade-4 students's participations in TIMSS 2015 across Emirates in the UAE

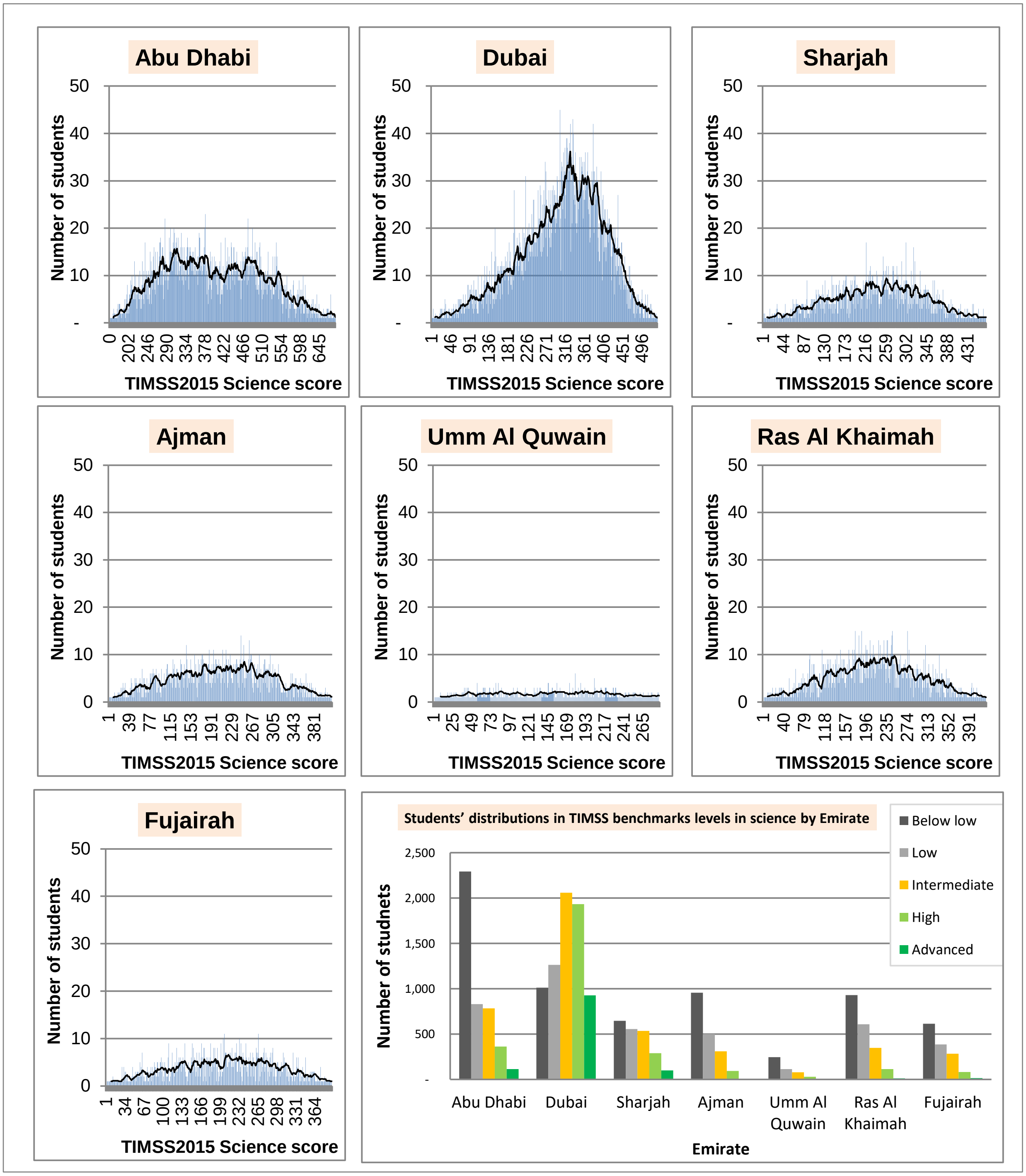


Figure 14: Students scores in science by Emirate

4.3. Cross-school-type analysis

In this section, we look at students' performance in TIMSS 2015 science assessments in the two schooling types followed in the UAE (public and private). Figure 15 shows private schools' students who participated in TIMSS 2015 science assessments are greater than public schools. This can be related to the fact that more students attend private schools than public schools in the UAE (Enrolled Students by Education Type and Nationality - Emirate of Dubai 2018).

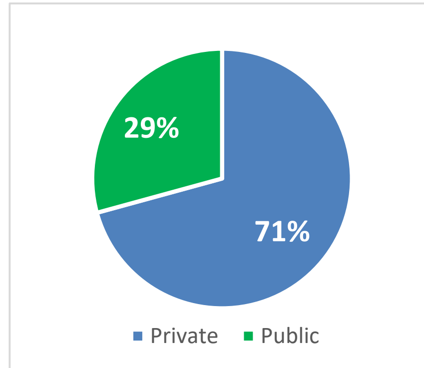


Figure 15: Students participations by scool type

When comparing students' performance in TIMSS science assessments, Figure 16 and Figure 17 show students in private schools scored higher than their peers in public schools.

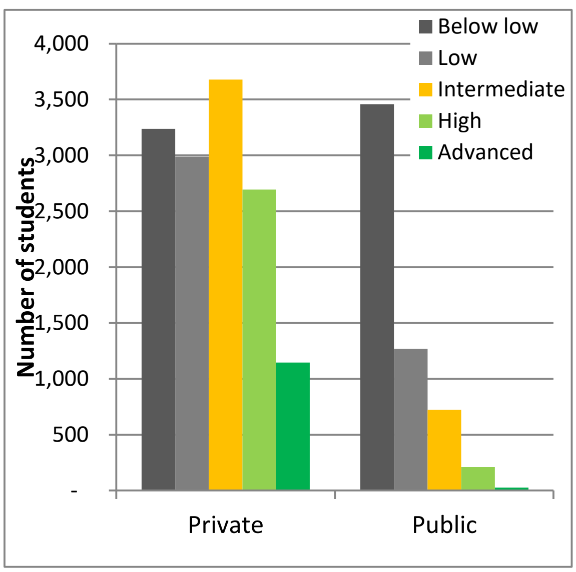


Figure 16: Students distribution in TIMSS benchmarks levels in science by school type

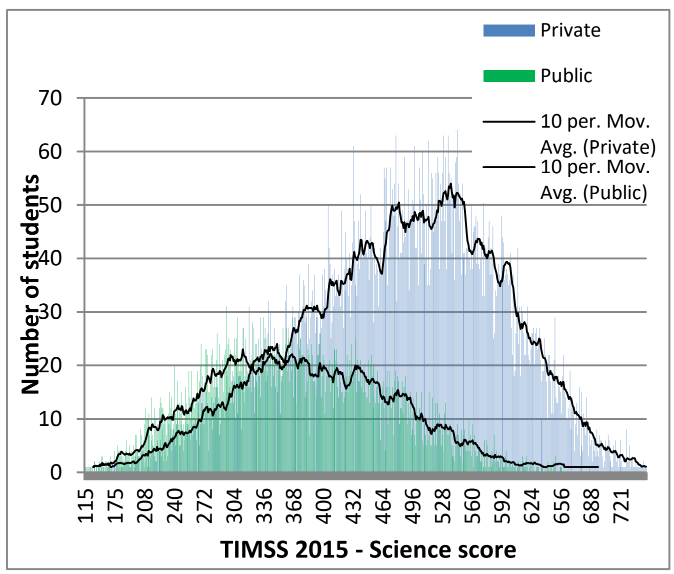
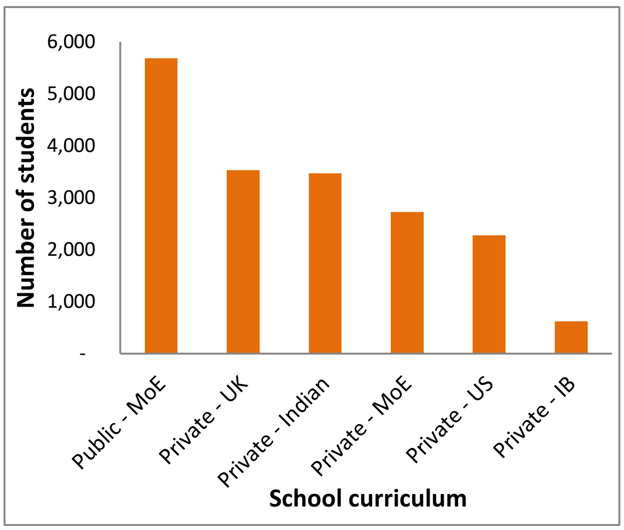


Figure 17: Students scores in science by school type

4.4. Cross-curriculum analysis

In the UAE, schools implement different international curricula. Figure 20 shows Grade-4 students' participation in TIMSS 2015 categorized by the main school curriculum. Figures 19 and 18 clearly shows students attending private schools following the British, Indian, or International Baccalaureate (IB) curriculum performed better than other students. Moreover, it is clear that the majority of students in Public MOE schools fall in the lower performing levels (Low and Below low).



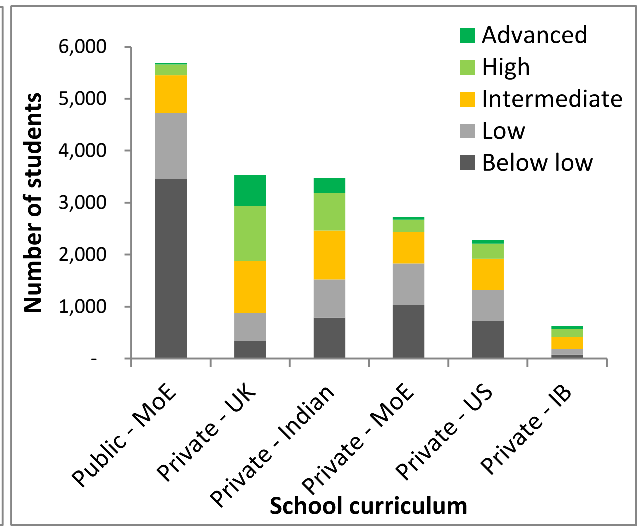


Figure 18: Students distribution school curriculum

Figure 19: Students distribution in TIMSS benchmark levels by school curriculum

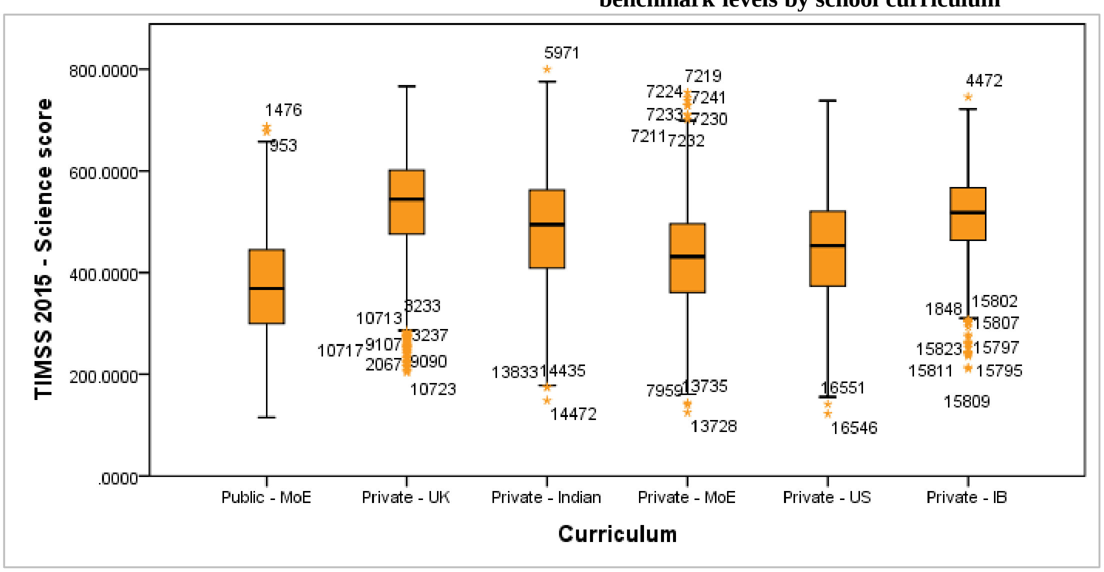


Figure 20: Boxplot of Grade-4 students' scores in science by school curriculum

4.5. Cross-gender analysis

This section shows analysis considering gender of students and teachers.

Students' gender

Figure 21 shows that students' participation was almost equal between boys and girls. Analysis in Figures 22 and 23 show that girls performed better than boys in TIMSS 2015 science assessments. This is clear across all Emirates and all school types.

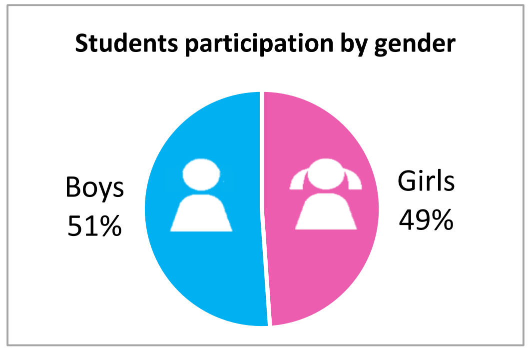


Figure 21: Students participation by gender

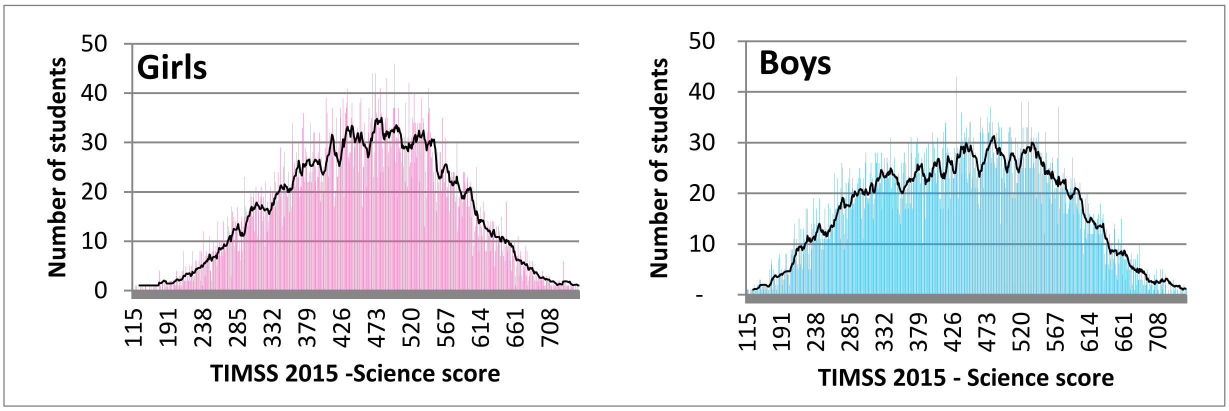


Figure 22: Students scores in science by gender

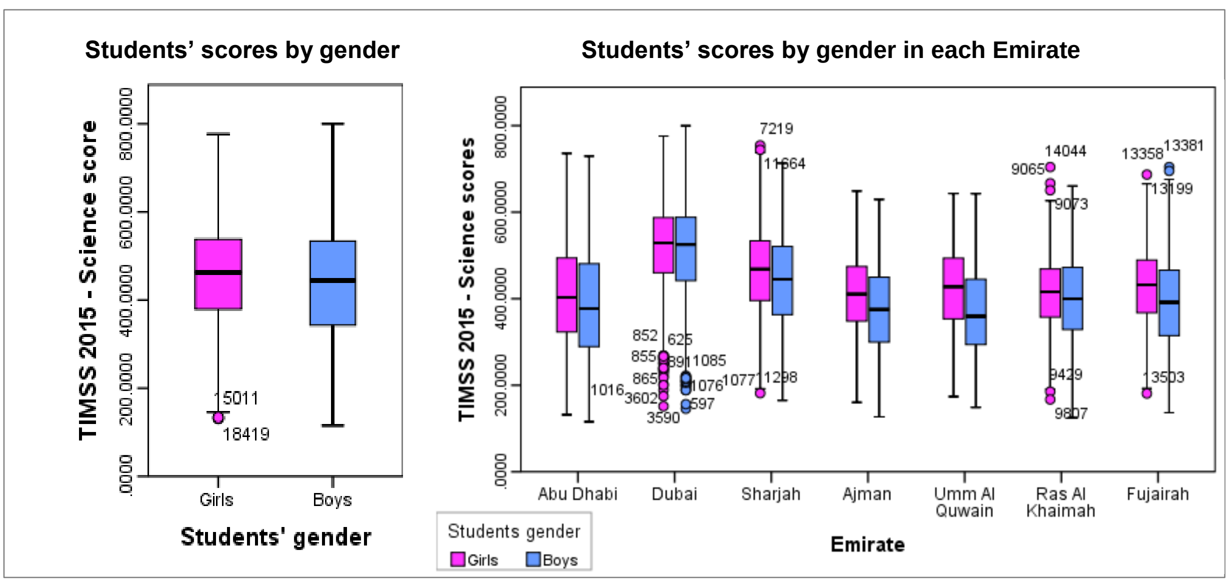


Figure 23: Boxplots of Grade-4 students' scores in all schools and by school Emirate

__Teachers' gender__

We were interested in examining the data in terms of teachers' gender. Figure 24 shows that 89% of Grade-4 teachers are female. When we looked deeper into students' scores based on their teachers' gender, we found that students whose teachers are female scored slightly higher than students with male teachers (shown in Figure 25). However, an interesting piece of analysis was found and presented in Figures 26 and 27. Students who were taught by teachers with different gender than their gender got higher scores than students whose gender is similar to their teachers' gender. Figure 27 shows that boys who were taught by female teachers got better scores than boys who were taught by male teachers. This does not mean that this factor impacts students' scores. It might be related to school types where there are no mixed genders in public schools, which is not the case in private schools.

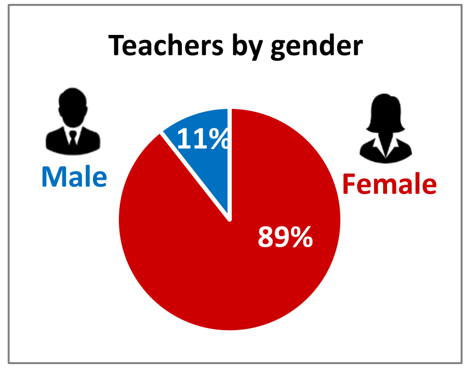


Figure 24: Proportion of teachers by gender

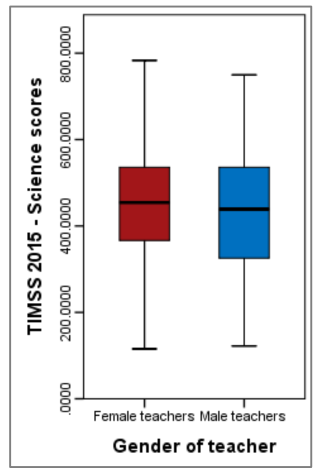


Figure 25: Students scores by teachers gender

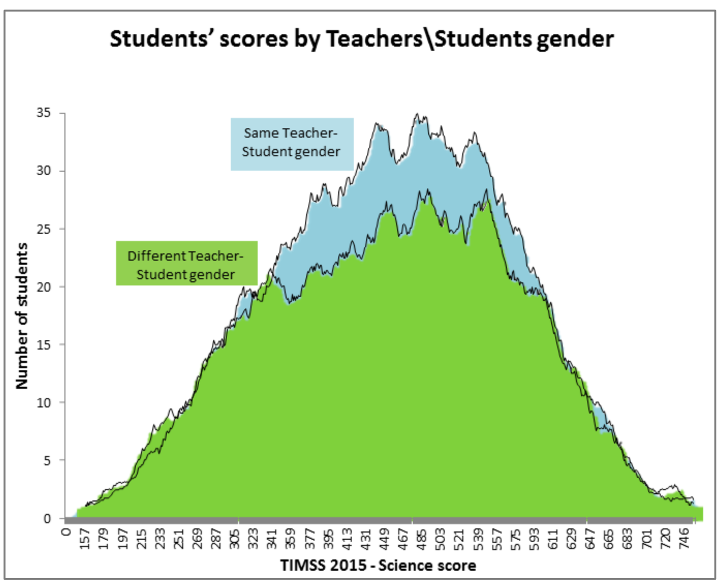


Figure 26: Distribution of students scoree by semilarity of gender with their teaches

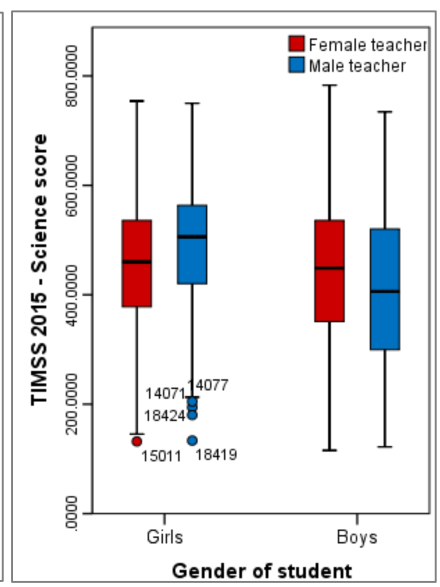


Figure 27: Students scores by teachers' and students' gender

4.6. Cross-age analysis

In this section we look into teachers' age and present students' scores in terms of their teachers' age group. In TIMSS teachers' background questionnaire, teachers were asked about their age group. The data was grouped into six groups as shown in Figure 28. Around 90% of teachers are between 25 and 49 years old. Figures 29 and 30 show that not much difference in students' scores among the different teachers' age groups.

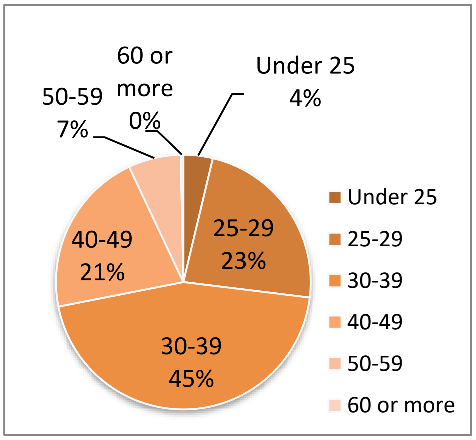


Figure 28: Proportion of teachers by age

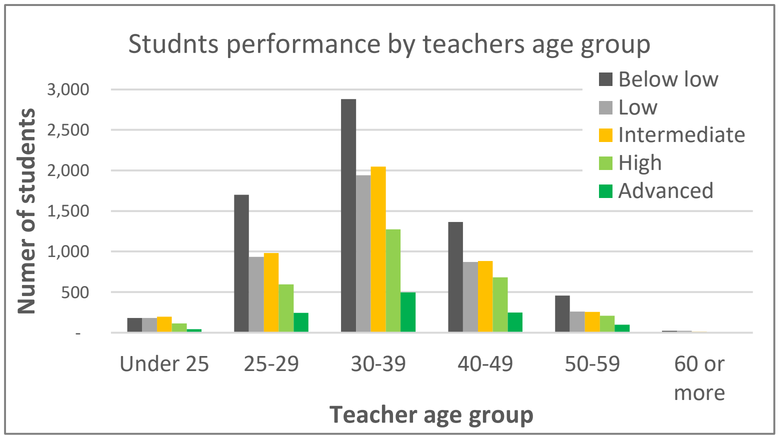


Figure 29: Proportion of teachers by age

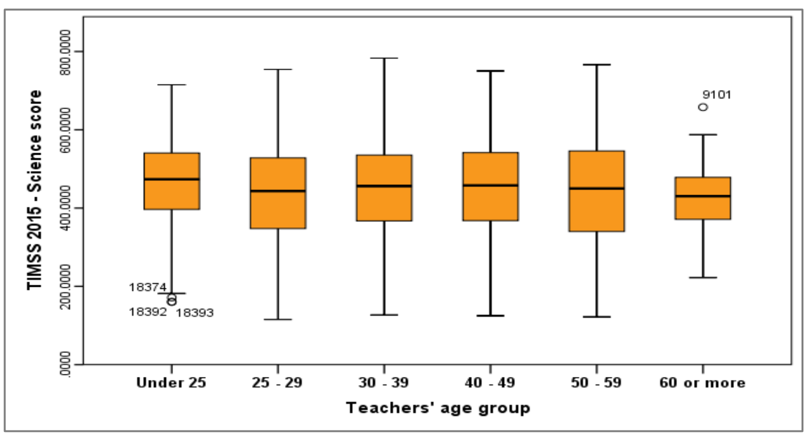


Figure 30: Boxplot of Grade-4 students' scores in science by teachers' age

4.7. Cross-educational-qualification analysis

TIMSS questionnaires contain questions about the highest formal level of education received by fathers, teachers, and principals. In this section, we look into this attribute and show the level of students' achievements corresponding to it.

Fathers' level of education

Looking into Fathers' level of education, analysis show that 50% of Grade-4 students' fathers who participated in TIMSS 2015 in the UAE have Bachelor's or Postgraduate degrees, shown in Figure 31. Figure 32 shows that students whose father's level of education is high got higher scores in TIMSS assessment compared to their peers whose father's level of education is lower.

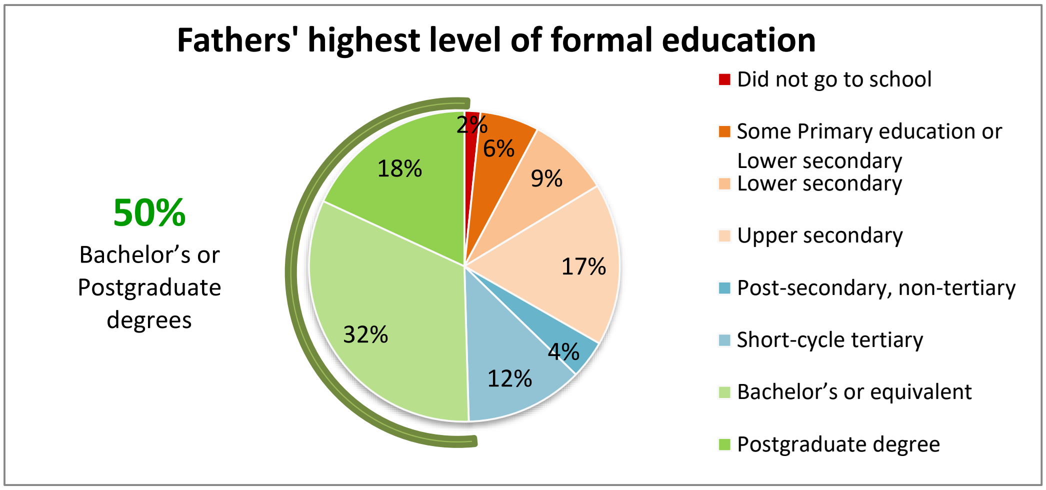


Figure 31: Proportion of father's highest level of formal education

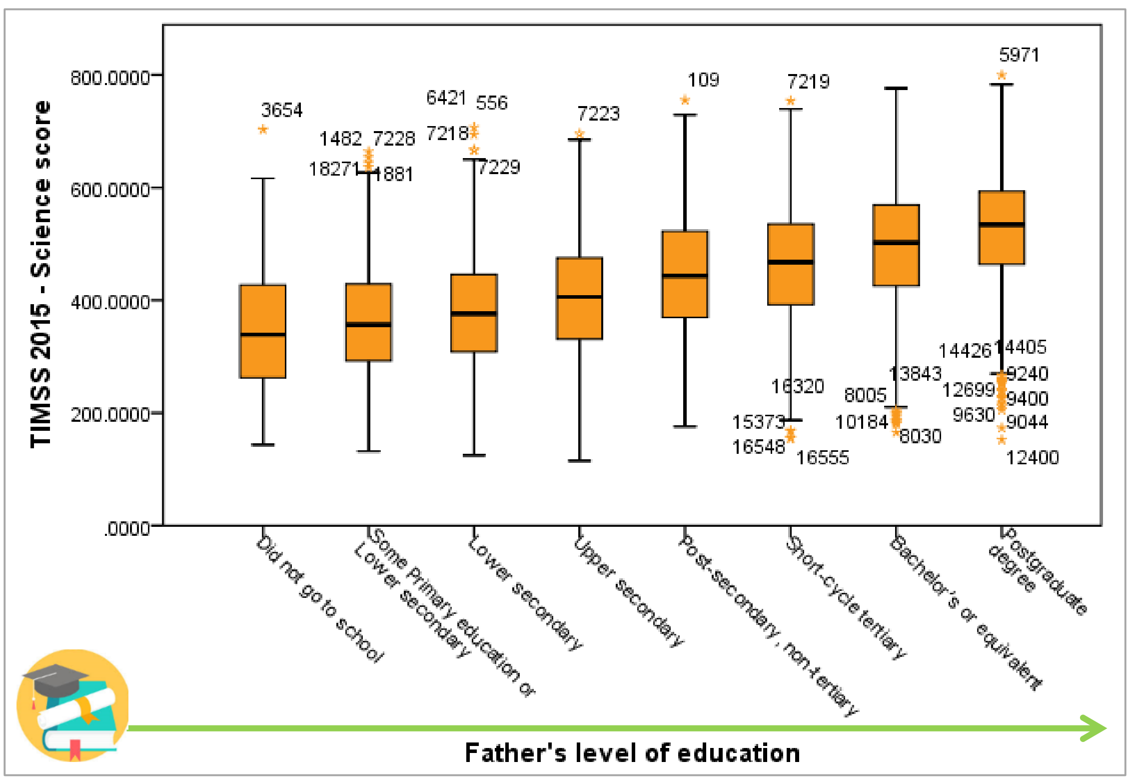


Figure 32: Students scores in TIMSS 2015 science by their fathers' educational level

Teachers' level of education

Looking at teachers' level of education, 95% of teachers have a Bachelor's or Master's degree as presented in figure 33. Figures 34 and 35 focus on those two groups and shows that students whose teachers are with a Masters' degree got slightly higher scores than those whose teachers are Bachelor degree qualified.

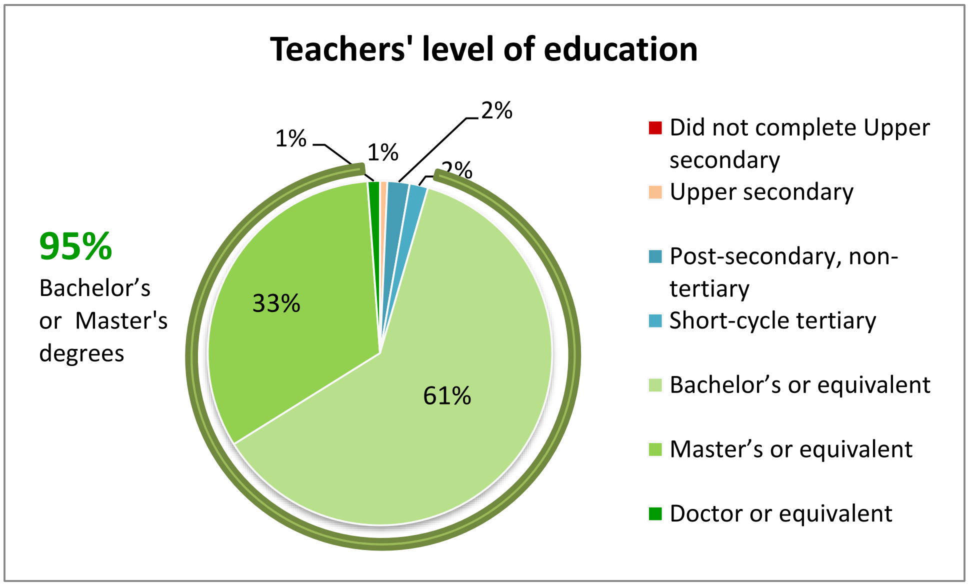


Figure 33: Proportion of science teachers' highest level of formal education

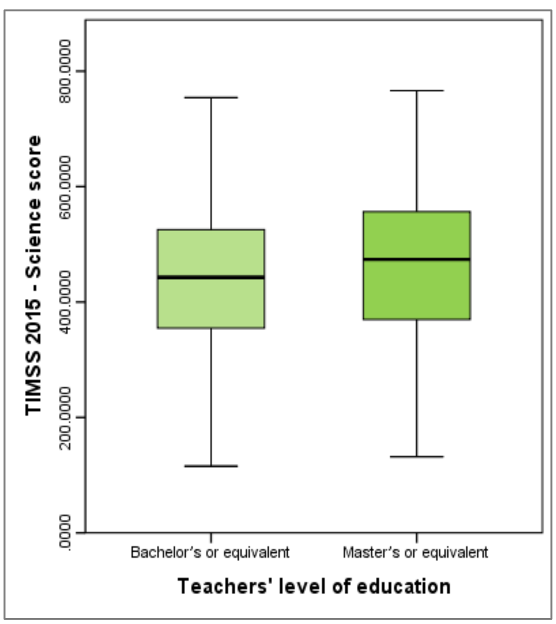


Figure 34: Students scores by theachers' level of education

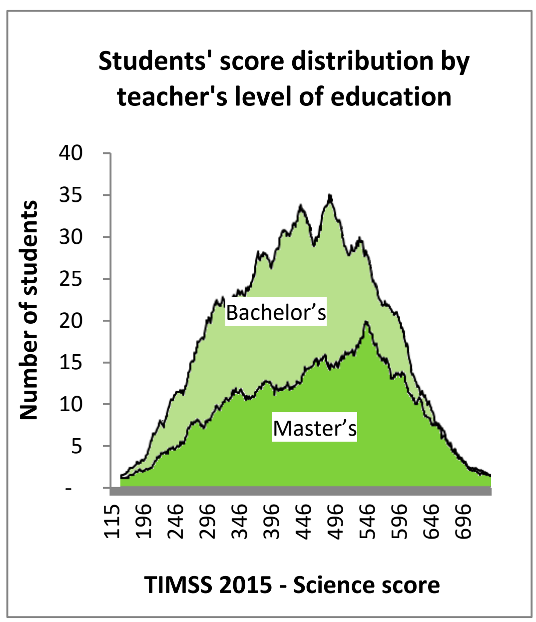


Figure 35: Distribution of students scoree by their teachers' level of education

Principals' level of education

Figure 36 shows that more than half of principals have Bachelor's degree, where 40% have Master's and 7% having a PhD. In figures 37 and 38 we look into student scores in those three groups. Students in schools led by principals with Master's degree scored higher than students in schools led by principals who are holding Bachelors or Doctoral Degrees.

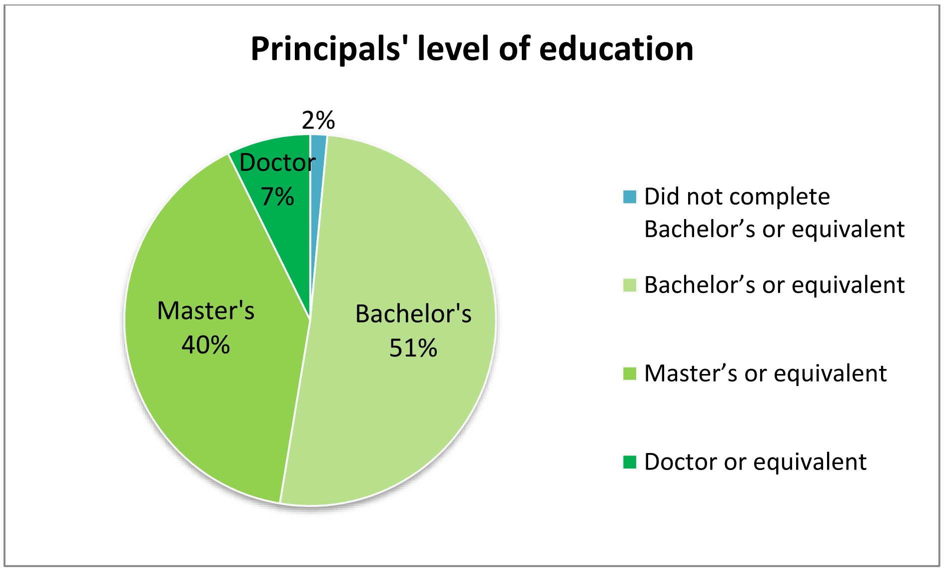


Figure 36: Proportion of school principals' highest level of formal education

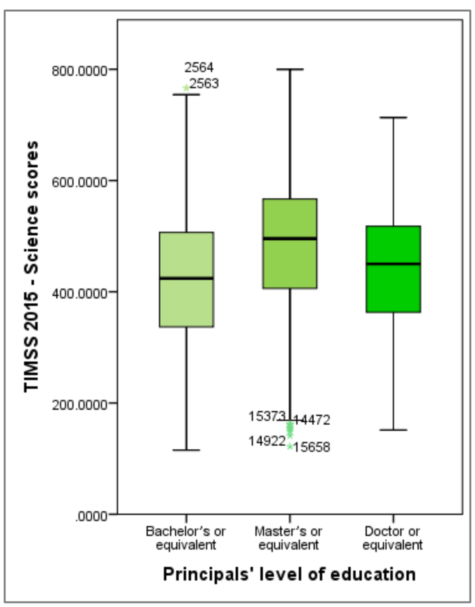


Figure 37: Students scores by principals' level of education

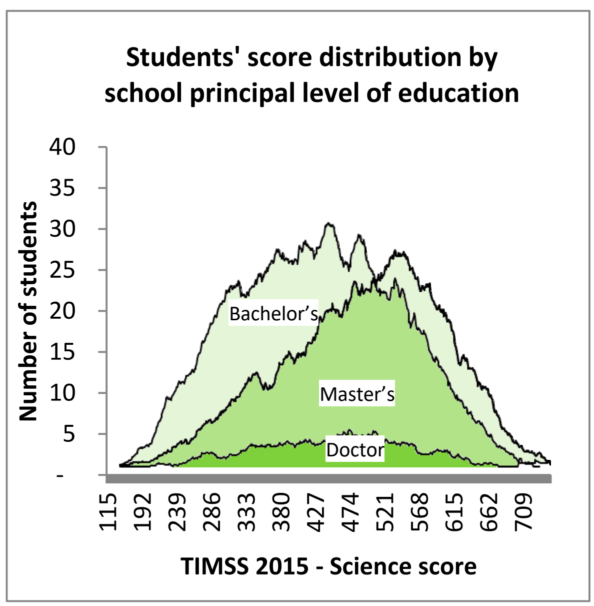


Figure 38: Distribution of students scoree by school principals' level of education

4.8. Specific analysis on Dubai's private schools

The private educational sector in Dubai is growing year by year. More than 89 % of all students in Dubai were attending private schools in the year 2014 ("Publications Article | Knowledge and Human Development Authority" 2014). Section 4.2 in this dissertation shows students in Dubai schools achieved higher scores compared to students in the other Emirates. In addition, the Knowledge and Human Development Authority (KHDA) started conducting inspections on private schools on an annual bases. This was a motivation to investigate and analyse the private education in Dubai considering TIMSS scores, TIMSS surveys data, and inspection results. This section presents a specific figures about private schools in Dubai.

Figures 39 shows that 90% of Grade-4 students in Dubai who participated in TIMSS 2015 science assessments attend private schools, which represent 140 schools. Figure 40 shows students' distribution in each benchmark level in Dubai schools, private and public. Students in private schools in Dubai are doing better than students in public schools in Dubai.

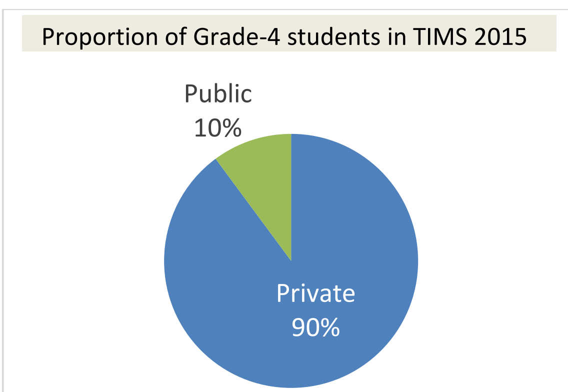


Figure 39: Proportion of students participation by school type in Dubai's private schools

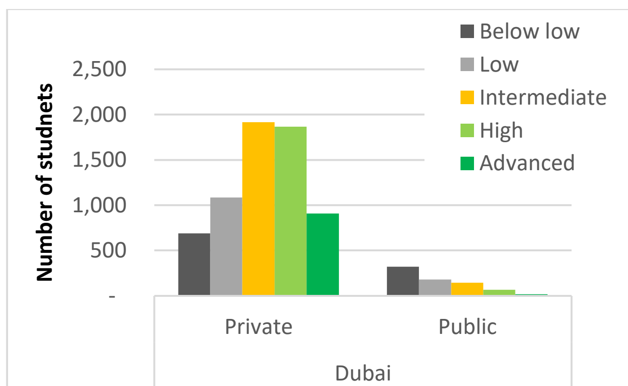


Figure 40: Grade-4 students levels in Dubai's schools by school type

Looking at students' scores in Dubai's private schools by curriculum, figure 41 shows that students' scores in UK, Indian and IB curriculum schools are higher than students' scores in schools following the MOE or US curriculum.

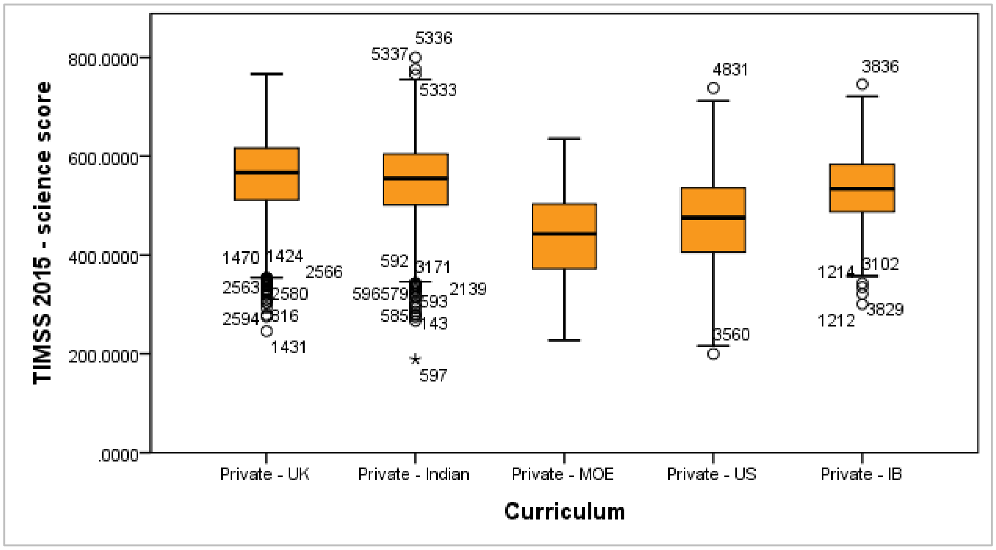


Figure 41: Students scores by curriculum in Dubai's private schools

It is interesting looking at students' scores in TIMSS assessments based on schools' inspection rating as reported by KHDA inspections in the year 2015. Figure 42 shows that the better the inspection rating of the school, the higher students' scores in science.

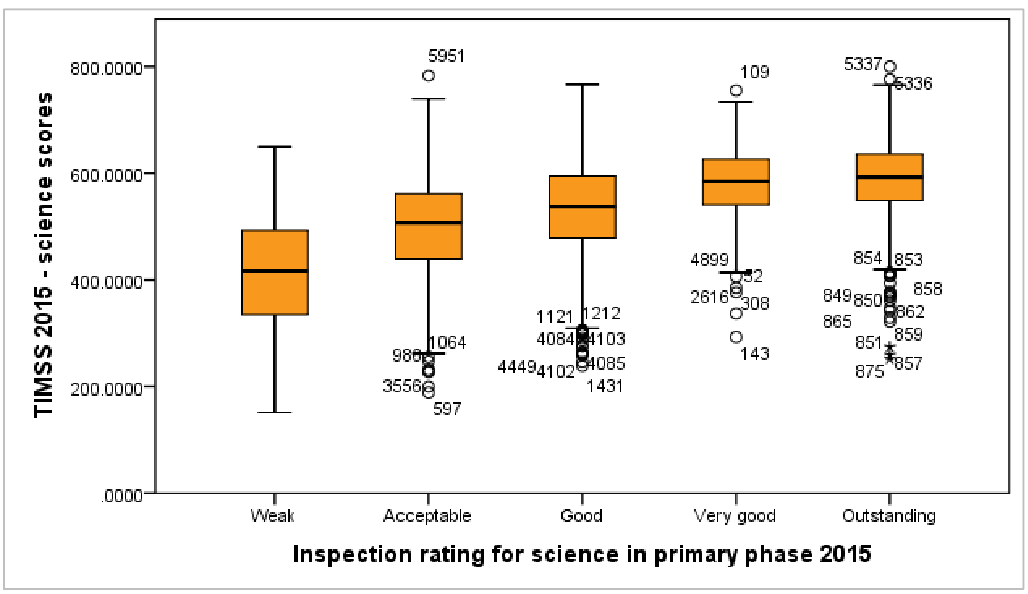


Figure 42: Students scores by 2015 inspection ratings in Dubai's private schools

Dubai private schools between 2011 and 2015

Looking at the trend of students' scores in TIMMS science, we compared Dubai private schools average scores in TIMSS 2011 and TIMSS 2015. This comparison was completed at school level, not student level because of the following reasons. Firstly, TIMSS data do not track students over the years. For example, X student (IDSTD = 001) who was in Grade-4 in the year 2011 and participated in TIMSS-2011 is not the same student with (IDSTD = 001) in Grade-8 data of TIMSS 2015. Secondly, not all student participate in TIMSS assessments. A representative sample is being selected to do the assessment every cycle.

Figure 43 shows number of schools participated in TIMSS 2011 and TIMSS 2015. Between 2011 and 2015, forty schools providing education to Grade-4 students opened and participated in TIMSS international assessments for the first time in the year 2015. We looked at the results of the 100 schools participating in TIMSS 2011 and 2015 cycles. Figure 44 shows that more than three-quarter of schools improved between 2011 and 2015.

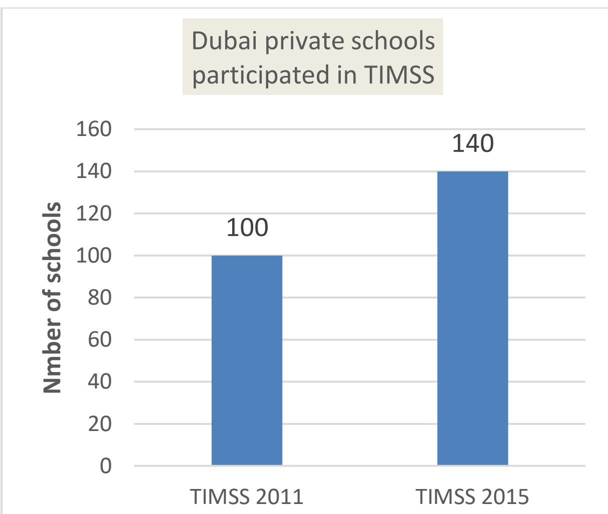


Figure 42: Schools participation in TIMSS 2011 and 2015

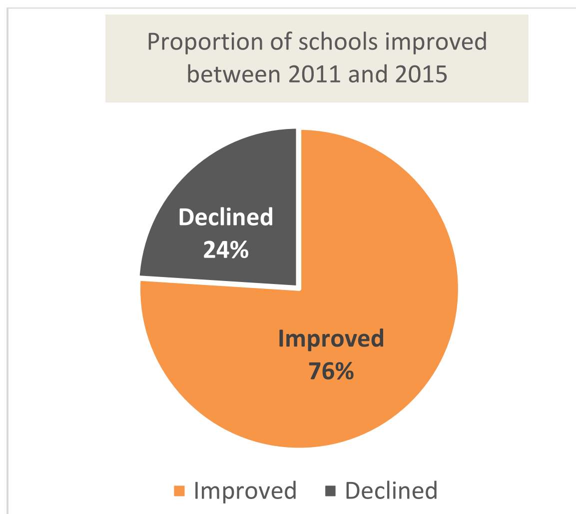


Figure 44: Proportion of schools improved and declined

Figures 45 and 46 show that improvements were more in low performing schools levels than schools with high students' scores in TIMSS 2011.

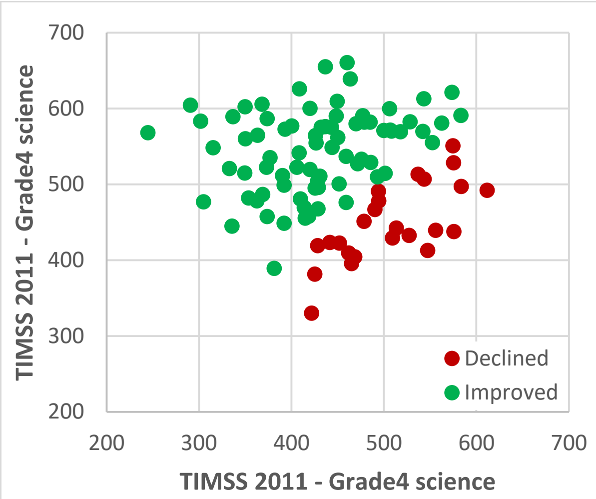


Figure 45: Schools participation in TIMSS 2011 and 2015

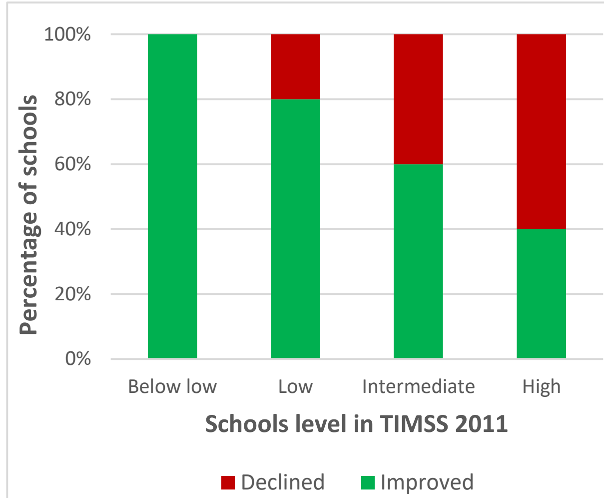


Figure 46: Proportion of schools improved and declined

The rest of this section provides a closer look into teachers in relation to their schools' leadership, their well-being, and what happens in their classrooms.

Teachers\ schools' leadership analysis

Looking at the relation between teachers with schools' leadership, figure 47 shows that the schools that improved in TIMSS between 2011 and 2015 had more collaboration between teachers and leadership to plan instructions. From another view, figure 48 shows that students in schools where the collaboration is higher between teachers and leadership did better than students where the collaboration is lower.

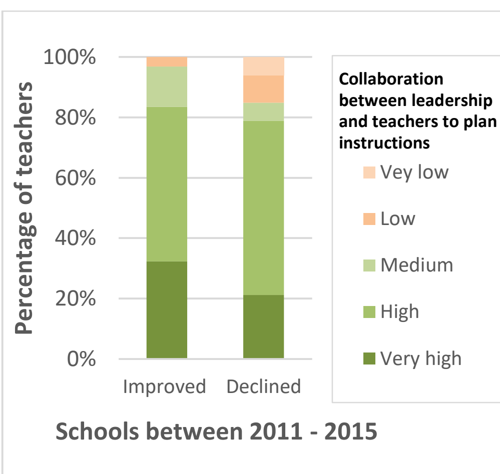


Figure 47: Collaboration between leadership and teachers in improved and declined schools

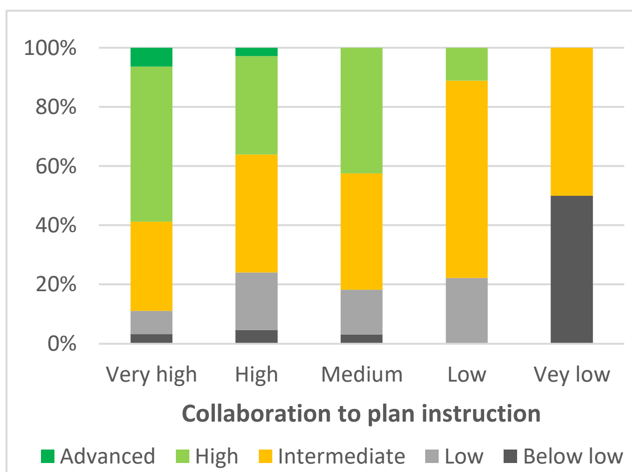


Figure 48: Proportion of students inTIMSS benchmark levels by the level of collaboration between leadership and teachers

We looked at the level of support provided to teachers by their school's leadership. The analysis below reflects teachers' responses in TIMSS questionnaire about the support provided to them by schools' leadership. As presented in figure 49, leadership support is higher in schools that improved. Figure 50 shows that students' scores are higher in schools where leadership provide more support to teachers.

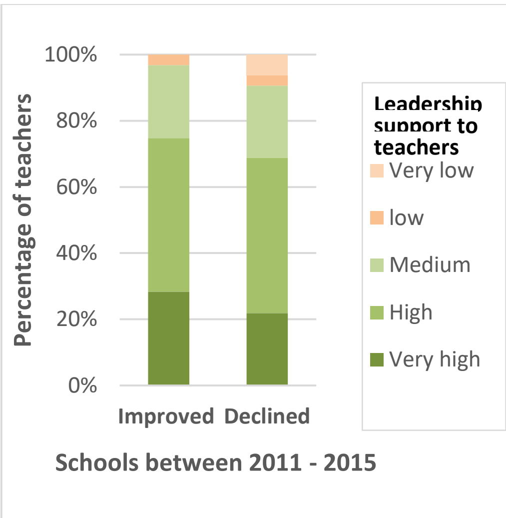


Figure 49: Leadership support to teachers in improved and declined schools

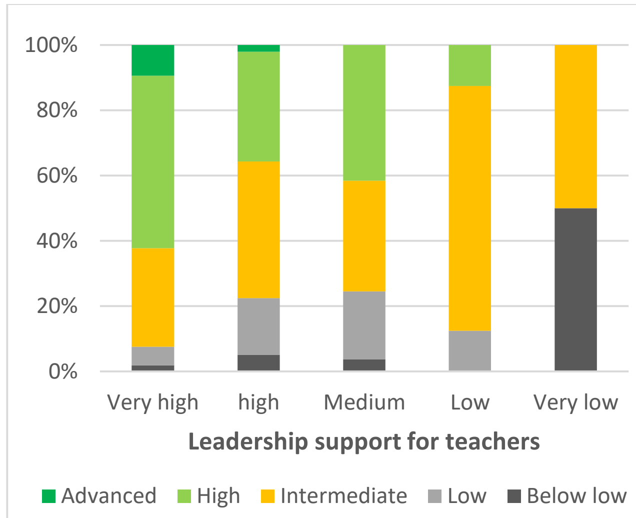


Figure 50: Proportion of students inTIMSS benchmark levels by the level of leadership support to teachers

Teachers' wellbeing

Looking at teachers questionnaires, there are questions related to teachers' well-being. We looked at teachers' satisfaction being a teacher at the school they are in. Figure 51 shows the distribution of students' scores by the level of their teachers' satisfaction. Students whose teachers are more satisfied got higher scores than students whose teachers are not satisfied being a teacher at their school.

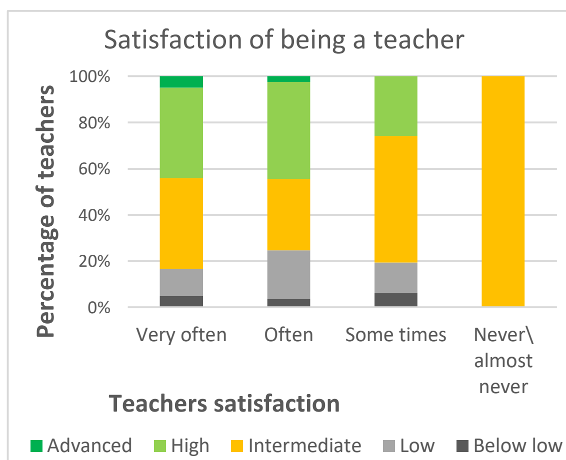


Figure 51: Proportion of students inTIMSS benchmark levels by their teachers' satisfaction

Figures 48 shows teachers' answers about their satisfaction being teachers at schools that improved in TIMSS between 2011 and 2015.comparing figure 52 and figure 53, more teachers are satisfied in schools that improved than those who dropped.

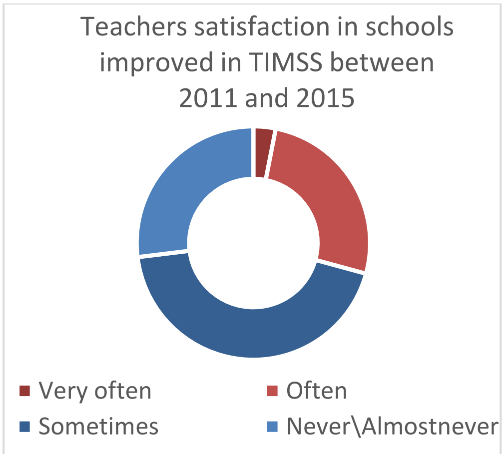


Figure 52: Teachers satisfaction in schools improved in TIMSS between 2011 and 2015

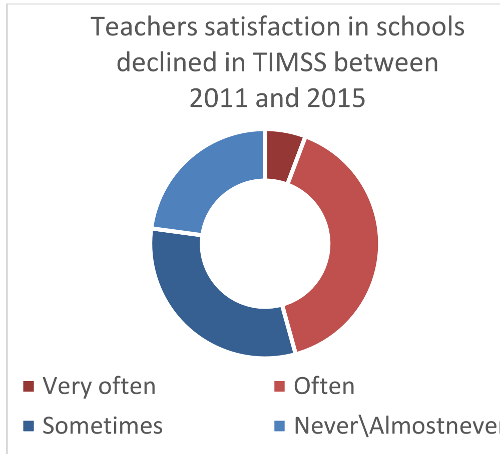


Figure 53: Teachers satisfaction in schools improved in TIMSS between 2011 and 2015

Teachers' in classrooms

This section shows students' scores in relation to what happens in their science classrooms. Figure 54 shows students based on how often their science teacher asks them distributes them in mixed ability groups. Figure 55 shows that Grade-4 students in schools declined in work more in mixed ability groups during science classes compared to students in schools that improved.

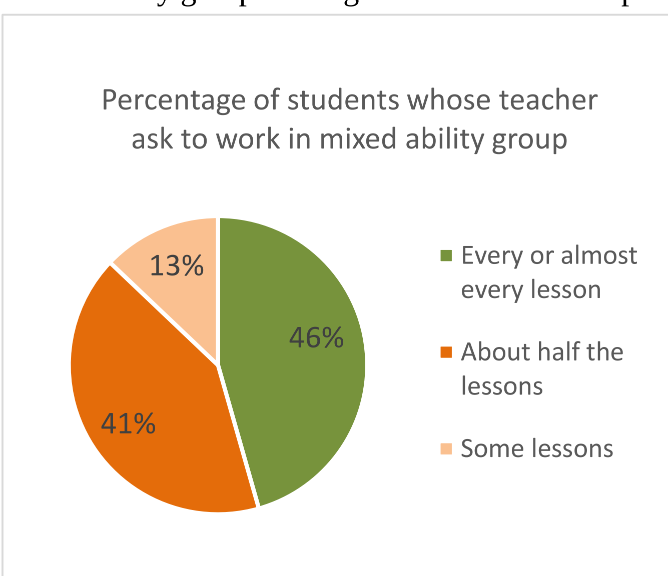


Figure 54: Proportion of students working in mixed ability groups

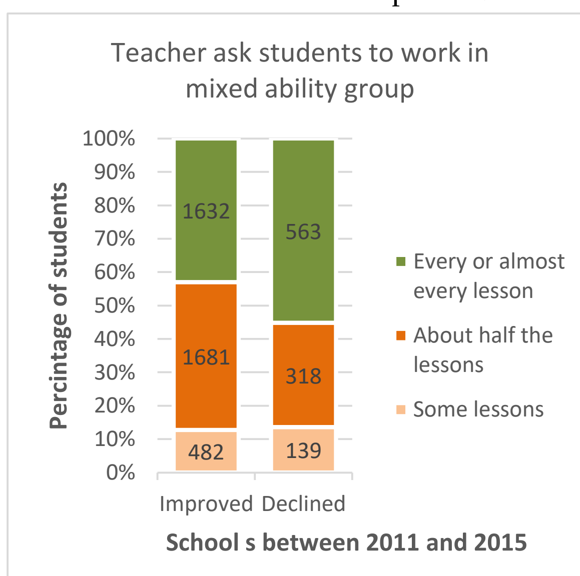


Figure 55: Students working in mixed groups in schools that improved\ declined

5. Chapter Five: Predictive analytics

Predictive analytics is a type of analysis that uses historical data and applies machine learning algorithms to create prediction models which can be used for future forecast and behaviour projection. This chapter presents results of applying the proposed methodology described in section (1.7) on the different datasets.

5.1. Data Pre-processing

Data pre-processing is a set of techniques to prepare the data for data mining tasks. This step is essential in transforming the real world data in to a form that helps machine learning algorithms understand the data and get accurate results. This section explains the pre-processing tasks done in this dissertation. Table 6 at the` end of this section shows the number of attributes in each dataset after each pre-processing task.

Aggregation

As described in chapter three, TIMSS data shows five plausible values for each student, in this phase of data-pre-processing, we aggregated the five plausible values into on value showing the average. In this case each student will have one average score for science and one for mathematics. Then in the teachers and schools data sets, we calculated the average science and mathematics score for each teachers and each school. Then, excluded students' records and duplicated teachers\ schools records which were mapped to students.

Discretization

Here we categorized students' score in science and mathematics based on the IEA international Benchmark levels. We created two new variables in all data sets to contain the categorical value of student score. Referring to IEA international Benchmark levels, there are five groups as follows:

- Advanced (625 or above),
- High (550-624),
- Intermediate (475-549),
- Low (400-474),
- Below Low (400 or below).

Dimensionality Reduction

Dimensionality reduction is excluding features that don't add information to the data. In the case of TIMSS data, there were many features which might confuse the machine learning process. Thus, this step was very important to clean up the data. The following group of attributes were excluded for the data considered in this study:

- IDs
- Students answers to each question
- Mathematic related variables

Features construction

Feature construction is the process of producing new features from the existing descriptive features. This step is important to build more effective features for the machine learning task. For example, the feature (Student_gender) had two values (1=Girl) and (2=Boy). These numbers were placed to describe students' gender. For the machine learning algorithm, these numbers will be interpreted to give Boys more value than girls. Thus, was replaced this feature by two new features (Boys) and (Girls). Each will have a value of 1 in case of true and 0 if false. The below features were constructed in TIMSS data:

- Teachers' gender
- Students' gender
- School type
- School curriculum
- School region

Remove duplication

Some of the variables are repeated, this was expected as we merged many datasets, and some have the same attribute in each.

Remove outliers

We applied the below formulas to decide the outlier boundary.

$(OT < LB)$ OR $(UB < OT)$

Where:
OT: Outlier value
LB: Lower bound of accepted values
UB: Upper bound of accepted values

$LB = Q1 - ((\mathrm{Q3} - \mathrm{Q1}) \times 1.5)$
$UB = ((\mathrm{Q3} - \mathrm{Q1}) \times 1.5) + Q3$

Where:
$Q3$: value of the 75$^{\text{th}}$ percentile
$Q1$: value of the 75$^{\text{th}}$ percentile

Replace missing values

We replace all missing values of the value "Null" by the average of the attribute. It was found that the proportion of missing values in the different data sets are 3% in Students data, 12% in home data and 4% in teachers data.

Normalization – scaling

TIMSS surveys data had different scales. Some question had option answers scaled from 1-6, while others are scaled from 1-4. This will cause inaccurate results in classification processes. We used normalization to map all features to have values from 0 to 1. We used min-max normalization: to [0, 1]. The below formula:

$$min = 0 \text{ , } max = 1$$

$$v' = \frac{v - min_A}{max_A - min_A}(new\_max_A - new\_min_A) + new\_min_A$$

Remove low variance

Variables that have the same value in all records have a variance of zero. These variables have less predictive power and affect the prediction process. We removed all variables that have variance near to zero (variance <0.05).

Table 9 shows size of the different datasets after each pre-processing task.

	Pre-processing task	Students dataset	Teachers dataset	Schools dataset	Techer\ School dataset	Teacher\ Student dataset	Dubai classrooms dataset	Dubai TIMSS\ Inspection dataset
1	Aggregation	19,429 × 195	735 × 201	527 × 110	735 × 288	21,177 × 392	5,466 × 71	5,466 × 470
2	Discretization	19, 429 × 196	735 × 202	527 × 111	735 × 289	21,177 × 393	5,466 × 72	5,466 × 471
3	Dimensionality Reduction	19, 429 × 158	735 × 186	527 × 93	735 × 211	21,177 × 344	5,466 × 50	5,466 × 461
4	Features construction	19, 429 × 175	735 × 203	527 × 110	735 × 228	21,177 × 361	5,466 × 67	5,466 × 444
5	Remove duplication	19, 429 × 173	735 × 203	527 × 110	735 × 288	21,177 × 356	5,466 × 61	5,466 × 440
6	Remove outlier	18,808 × 173	735 × 203	527 × 110	735 × 288	19,403 × 356	5,149 × 61	5,149 × 440
7	Replace missing values	18,808 × 173	735 × 203	527 × 110	735 × 288	19,403 × 356	5,149 × 61	5,149 × 440
8	Normalization	18,808 × 173	735 × 203	527 × 110	735 × 288	19,403 × 356	5,149 × 61	5,149 × 440
9	Remove low variance	18,808 × 169	735 × 198	527 × 104	735 × 281	19,403 × 351	5,149 × 61	5,149 × 440

Table 9: Datasets size track after each pre-processing task (cases × features)

5.2. Feature selection results

We applied the three feature selection techniques described in section (1.1.2) on the five datasets. Table 10 shows a comparison of execution time needed by feature selection methods to return the selected set of features. Figures 56 and 57 show that Wrapper feature selection method (RFE-Naïve Bayes) need much more time for execution compared to UFS and Tree-based methods. Among the three feature selection methods, embedded feature selection method (Tree-based) shows to be the fastest across all datasets.

	Feature Selection method	Students dataset	Teachers dataset	Schools dataset	Techer\School dataset	Teacher\ Student dataset
1	No Feature selection	0.0	0.0	0.0	0.0	0.0
2	Univariate Feature Selection (UFS)	0.27	0.27	0.17	0.39	0.30
3	Recursive Feature Elimination (RFE)	7.78	1.28	1.69	2.61	32.10
4	Tree-based	0.17	0.09	0.09	0.09	0.24

Table 10: Execution time of feature selection methods (seconds) on different datasets

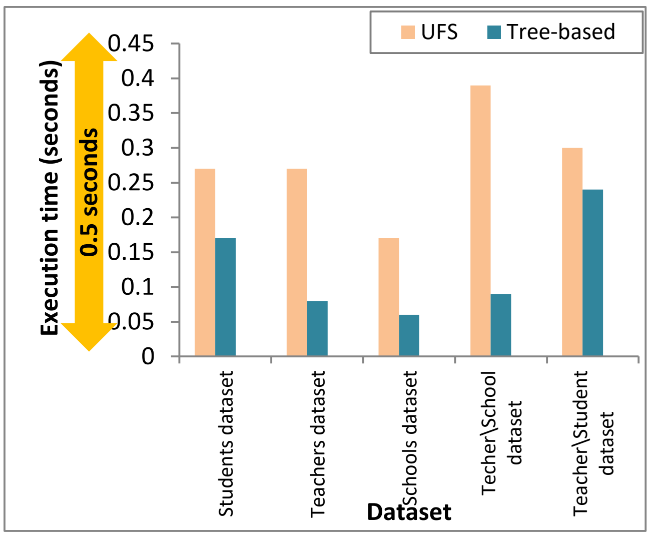


Figure 56: Execution time of UFS and Tree-based

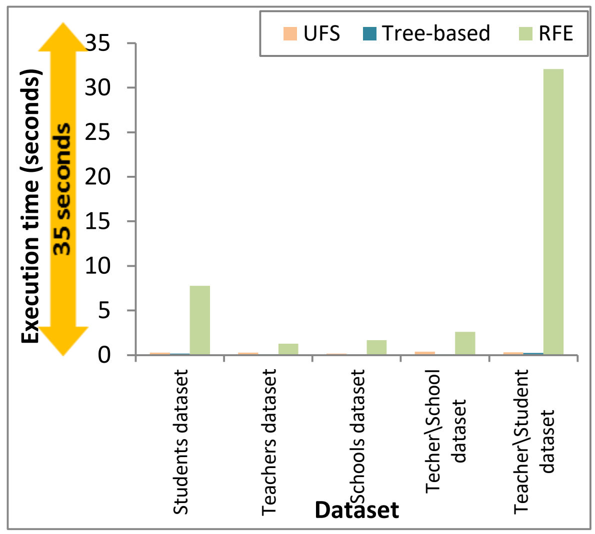


Figure 57: Execution time of all feature selection methods

Table 11 shows the number of features selected by feature selection methods on the different datasets. Those selected features will be fed into the different classifiers for building the prediction model. Tree-based feature selection method returns fewer number of features compared to the other feature selection methods.

	Feature Selection method	Students Dataset	Teachers dataset	Schools dataset	Techer\School dataset	Teacher\ Student Dataset
1	No Feature selection (All features)	All (169)	All (200)	All (104)	All (281)	All (351)
2	Univariate Feature Selection (UFS)	10	10	10	10	10
3	Recursive Feature Elimination (RFE)	10	10	10	10	10
4	Tree-based	8	7	7	10	8

Table 11: Number of features selected by different methods from different datasets

Looking at the performance of feature selection methods, Table 12 displays the average accuracy of 10-fold cross validation technique. Figure 58 plots the average accuracy of the three feature selection. Tree-based feature selection showed the lowest accuracy on all datasets. However, this was not considered to be the measure for the best feature selection method. In the next section (section 5.6.2), we will measure the accuracy of the prediction scheme which will be evaluated to check the best performing scheme.

	Feature Selection method	Students Dataset	Teachers dataset	Schools dataset	Techer\School dataset	Teacher\ Student Dataset
1	No Feature selection	-	-	-	-	-
2	Univariate Feature Selection (UFS)	59.20 %	67.60 %	64.20 %	68.50 %	58.50 %
3	Recursive Feature Elimination (RFE)	59.09 %	66.00 %	64.85 %	66.94 %	57.85 %
4	Tree-based	56.73 %	62.59 %	64.14 %	62.59 %	56.30 %

Table 12: Accuracy of feature selection methods on different datasets

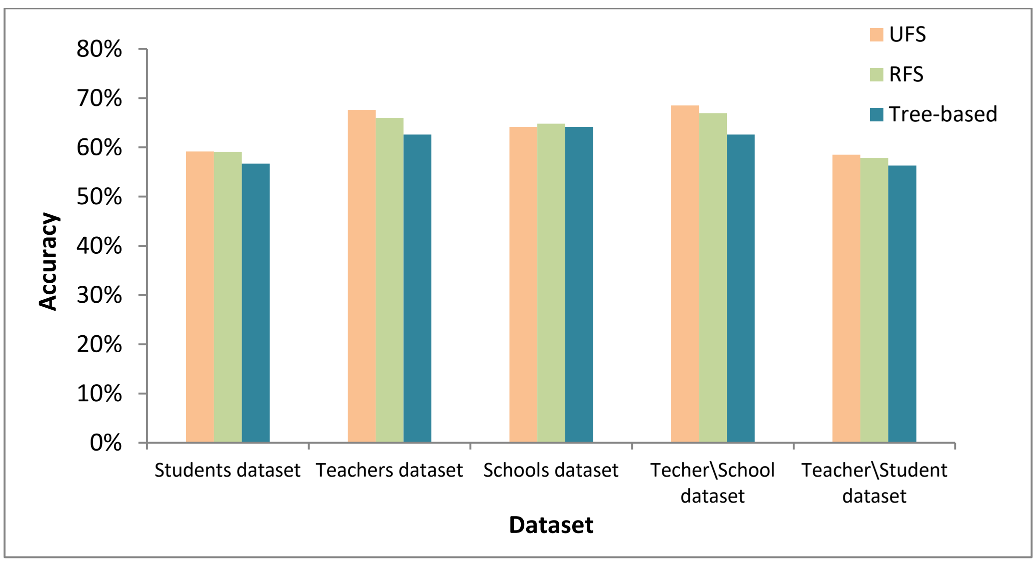


Figure 58: Accuracy of feature selection methods on different datasets

5.3. Classification (prediction model) and results comparison

In this section we shows results of all prediction models. We built a set of schemes by mapping the three classification algorithms (Decision tree, KNN and SVM) with the three feature selection methods. In addition, we included the results of classifiers without feature selection (Including all features in datasets). This resulted of having twelve prediction schemes for being prediction models for each dataset. In addition to presenting the resulted figures and findings of implementing the proposed experimental methodology on the different datasets.

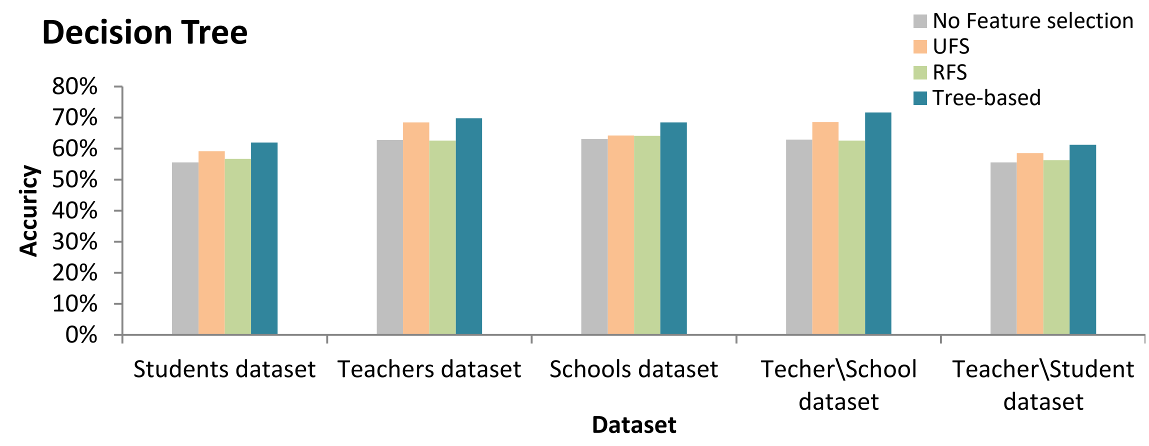
Tables 13 and 14 show the average accuracy of the twelve prediction schemes examined in this study. Decision tree classifier (DT) along with Tree-based feature selection method (DT\Tree-based) shows the highest performance on the first four datasets. SVM \ Tree-based scheme shows the highest performance on Teaches\Schools dataset. Considering execution time of feature selection method, Tree-based feature selection method shows to be the fastest. As a result, those schemes are selected to be the best to build an accurate prediction model for the different datasets. Details of the best prediction schemed selected are described in section (5.6).

Classifier	Feature Selection Scheme	Students dataset	Teachers dataset	Schools dataset	Techer\School dataset	Teacher\Student dataset
Decision tree (DT)	No Feature selection	55.6 %	62.80 %	63.10 %	62.90 %	55.60 %
	Univariate Feature Selection (UFS)	59.20 %	68.40 %	64.20 %	68.50 %	58.50 %
	Recursive Feature Elimination (RFE)	56.70 %	62.60 %	64.10 %	62.60 %	56.30 %
	Tree-based	**61.90 %**	**69.80 %**	**68.40 %**	**71.60 %**	61.19 %
KNN	No Feature selection	59.24 %	63.26 %	64.70 %	65.44 %	41.47 %
	Univariate Feature Selection (UFS)	55.36 %	61.79 %	67.38 %	66.67 %	54.57 %
	Recursive Feature Elimination (RFE)	36.84 %	50.85 %	52.30 %	52.08 %	41.60 %
	Tree-based	55.09 %	65.45 %	65.48 %	68.98 %	59.89 %
SVM	No Feature selection	59.73 %	63.00 %	65.27 %	63.81 %	57.13 %
	Univariate Feature Selection (UFS)	56.11 %	66.14 %	61.90 %	65.32 %	55.46 %
	Recursive Feature Elimination (RFE)	57.19 %	62.59 %	64.14 %	62.59 %	56.80 %
	Tree-based	55.98 %	67.37 %	61.90 %	66.14 %	**62.81 %**

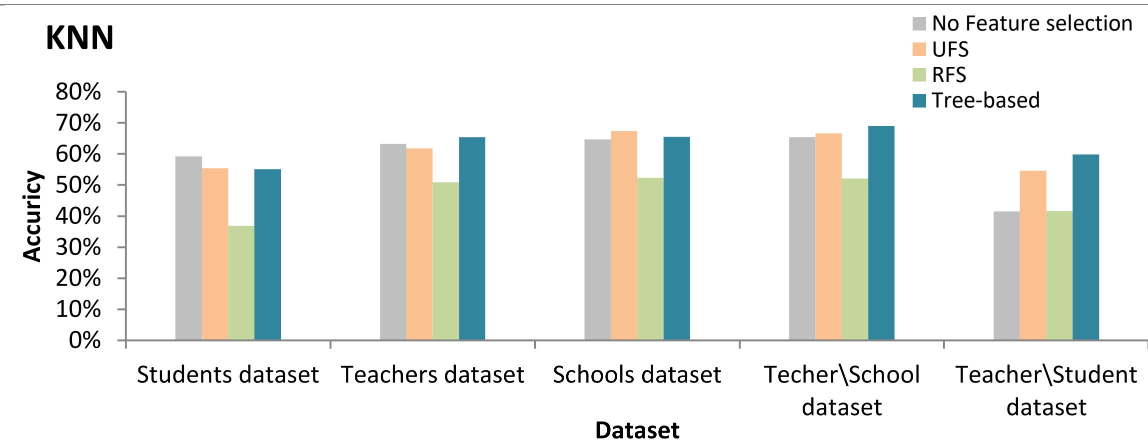
Table 13: Accuracy of classifier with features selection methods on different datasets

Classifier

Decision tree



KNN



SVM

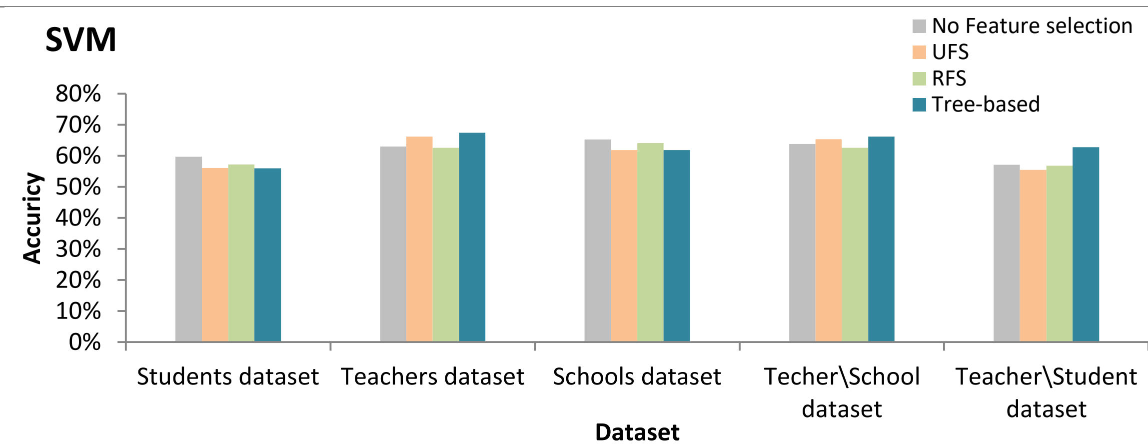


Table 14: Classification accuracy

5.4. Best scheme for prediction and results discussion

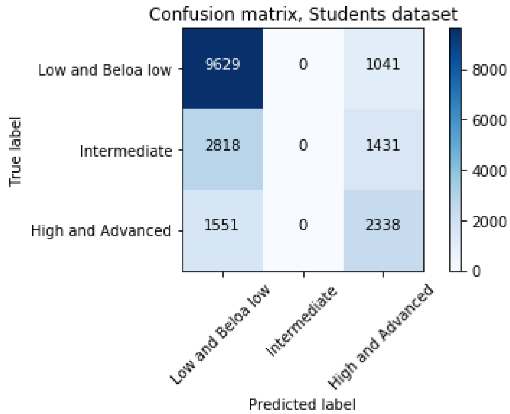
In this section, we discuss the best prediction scheme for each dataset showing the selected features and prediction model performance. Table 15 shows accuracy and precision of the best prediction scheme selected for each dataset based on tenfold cross validation as a test method. Decision tree with tree-based feature selection (DT \ Tree-based) showed the highest accuracy of 62% on students' dataset, 70% on teachers' dataset, 68% on schools' dataset and 72% on teachers\schools dataset. SVM with tree-based feature selection (SVM \ Tree-based) showed the highest accuracy of 63% on teachers\students dataset. The class (Low and Below low) was selected to show the precision of the best prediction scheme.

	Dataset	Best scheme	Average accuracy	Precision of Low and Below low class
1	Students dataset	DT \ Tree-based FS	**61.90 %**	68.79%
2	Teachers dataset	DT \ Tree-based FS	**69.80 %**	81.38%
3	Schools dataset	DT \ Tree-based FS	**68.40 %**	78.92%
4	Techer\School dataset	DT \ Tree-based FS	**71.60 %**	81.94%
5	Teacher\student dataset	SVM \ Tree-based FS	**62.81 %**	65.06%

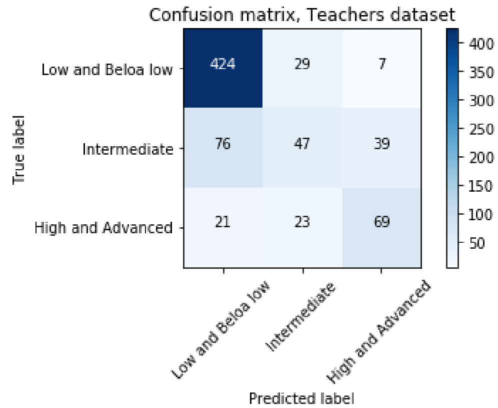
Table 15: Details of the best prediction scheme on the different datasets

Looking at confusion matrixes of the best prediction models displayed in Table 16 in addition to the precision values in Table 15, it is noticed that the prediction performance is lower in datasets that include students' related attributes (Students dataset) and (Teacher\student dataset). This can refer to the 12% of missing data 'Home questionnaire data' which is part of students' dataset, as mentioned in section 5.1 about the pro-processing task of replacing missing values by the mean of the attribute. Farhangfar, Kurgan & Dy (2008) found that using the mean as an imputation method for missing data between 5% and 50% does not improve classification results. This can explain the low performance of datasets that contain students' data.

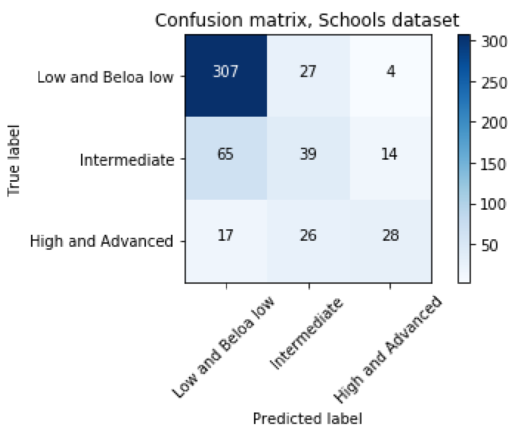
Students dataset



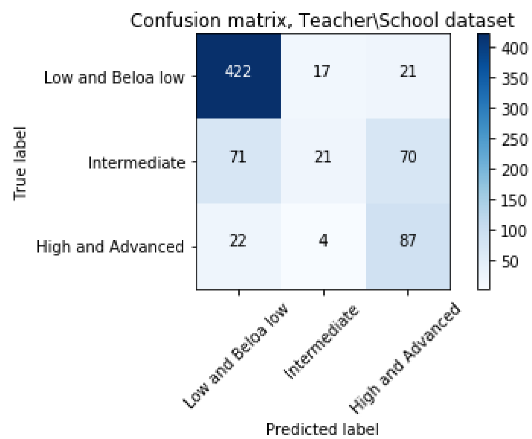
Teachers dataset



Schools dataset



Techer\School dataset



Teacher\student dataset

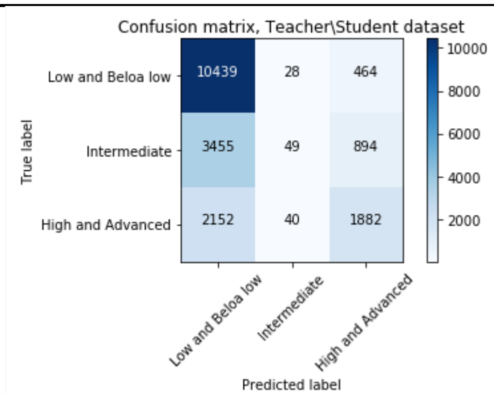


Table 16: Confusion matrixes of the best prediction models for the different datasets

As a result of applying the experimental methodology on the different datasets, Table 17 shows a list of features selected from each dataset. This gives an indication of the important factors influencing Grade-4 students' performance in TIMSS international assessments.

Students dataset

```
1)  'ASBH02A'      → Before your child began primary/elementary school, parents used
                      to reads books with their child
2)  'ASBH02B'      → Before your child began primary/elementary school, parent used
                      to tell the child stories
3)  'ASBH03A'      → Student was born in the country
4)  'ASBH10BA'     → Student attend extra lessons or tutoring for Mathematics
5)  'ASBH16A'      → Parents' believe that most occupations need
                     skills in math, science, or technology
6)  'ASDHENA'      → Chile had Numeracy activities before joining primary school
7)  'ASDHEDUP_Parents_Highest_Education_Level'→ Parents' highest level of education
8)  'ASDHOCCP'     → Parents' Highest Occupation Level
9)  'Private - UK' → School is following UK curriculum
10) 'Dubai'        → School is in Dubai
```

Teachers dataset

```
1)  'ATBG06C'      → Teachers' expectations for student achievement within the school
2)  'ATBG06G'      → Teachers' characterization of parental commitment to ensure
                      that students are ready to learn within the school
3)  'ATBG07D'      → The students behave in an orderly manner
4)  'ATBG08A'      → school building needs significant repair
5)  'ATDGLSN'      → Teachers view that teaching is limited by student needs
6)  'Public - MoE' → School is following MOE curriculum
7)  'Dubai'        → School is in Dubai
```

Schools dataset

```
1)  'ACBG14CC'     → Schools capacity to provide library resources relevant
                      to science instruction
2)  'ACBG15K'      → Students' desire to do well in school
3)  'ACDGSRS'      → School's capacity to provide instructional science
4)  'ACDGEAS'      → School Emphasis on Academic Success
5)  'Private'      → School is private
6)  'Public - MoE' → School is following MOE curriculum
7)  'Dubai'        → School is in Dubai
```

Techer\School dataset

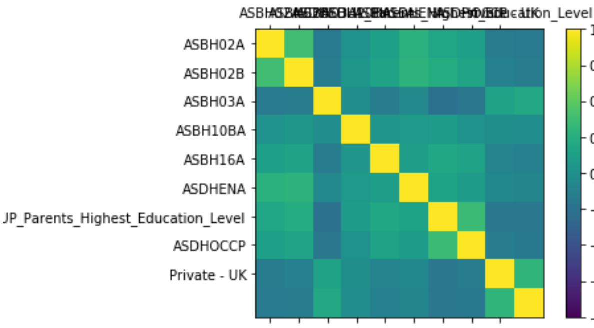
```
1)  'ATDGLSN'      → Teaching is limited by student needs
2)  'ACBG03A'      → Students come from economically disadvantaged homes
3)  'ACBG14AA'     → Schools' capacity to provide Instructional materials
4)  'ACBG14AE'     → Schools' capacity to provide Instructional space
5)  'ACBG15H'      → Schools' characterization of parental expectations for
                      student achievement
6)  'ACBG15L'      → Schools' characterization of students' ability to reach
                      school's academic goals
7)  'ACBG16E'      → Profanity is a problem among the school
8)  'Public - MoE' → School is following MOE curriculum
9)  'Dubai'        → School is in Dubai
```

Teacher\student dataset
1) 'ATBS04A' → students in class have computers (including tablets) available to use during their science lessons 2) 'ASBG08' → Students' absence 3) 'ASBS06C' → Students believe that "I am just not good in science" 4) 'ASDGHRL' → Student has resources for learning at home 5) 'ASBH02A' → Before your child began primary/elementary school, parents used to reads books with their child 6) 'ASBH02B' → Before your child began primary/elementary school, parent used to tell the child stories 7) 'ASDHELN' → Chile had literacy and numeracy activities before joining primary school 8) 'ASDHEDUP_Parents_Highest_Education_Level'→ Parents' highest level of education 9) 'Private - UK' → School is following UK curriculum

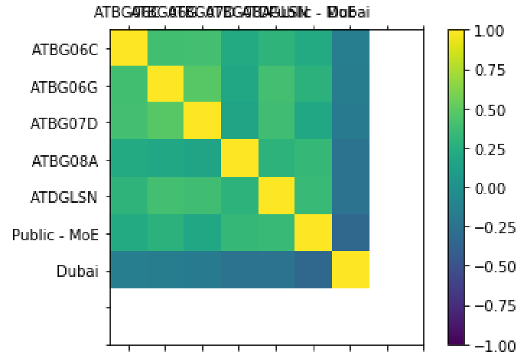
Table 17: Key factors\features used to build prediction models with the highest accuracy

Table 18 shows correlation matrixes of selected features. Confusion matrixes provides a visual representation of the correlation between features. The x- axis and y-axis list the selected features which makes the confusion matrix looks symmetrical. The colour code of each square represents the level of correlation between two variables based on the colour scale presented beside the matrix. This explains the yellow colour in the diagonal line which reflects a perfect positive correlation of a feature with itself.

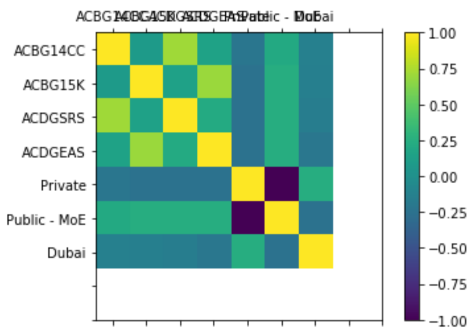
Students dataset
(Tree-based feature selection)



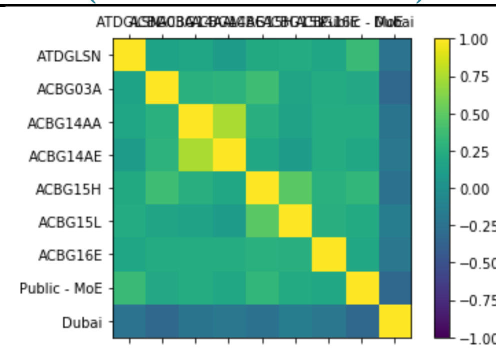
Teachers dataset
(Tree-based feature selection)



Schools dataset
(Tree-based feature selection)



Techer\School dataset
(Tree-based feature selection)



Teacher\student dataset
(Tree-based feature selection)

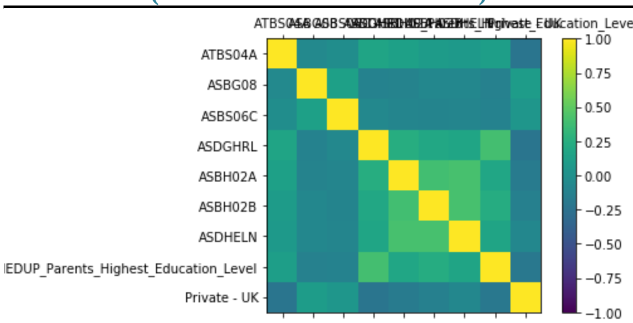
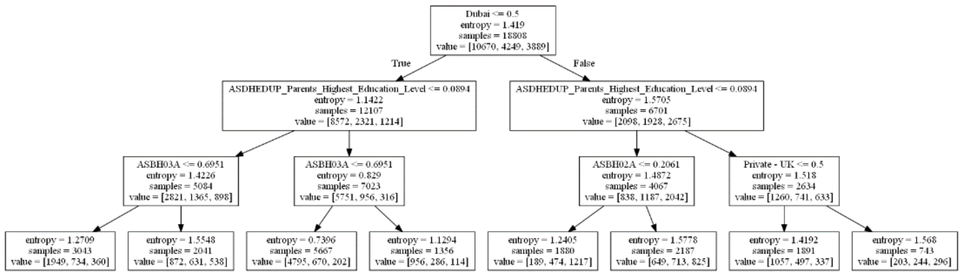


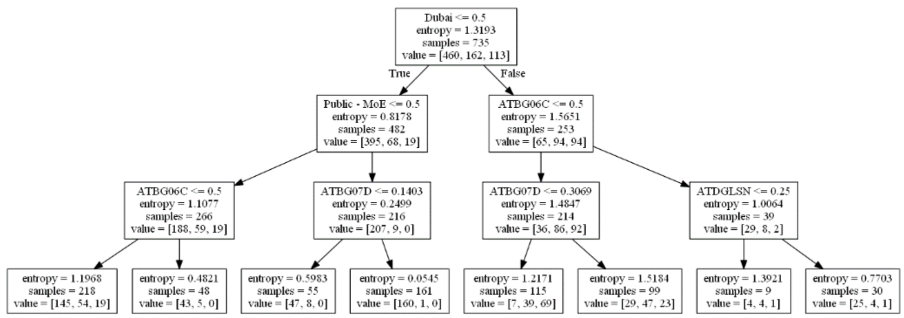
Table 18: Correlation matrixes of the selected features in the best prediction models

Table 19 presents the resulted decision tree of the best prediction models.

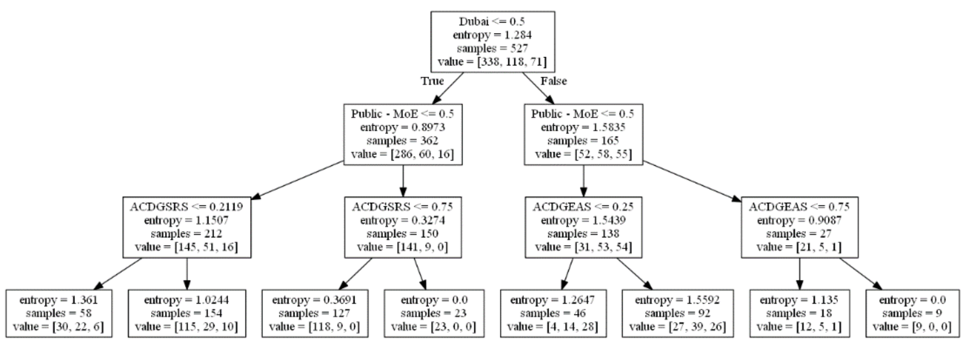
Students dataset



Teachers dataset



Schools dataset



Techer\School dataset

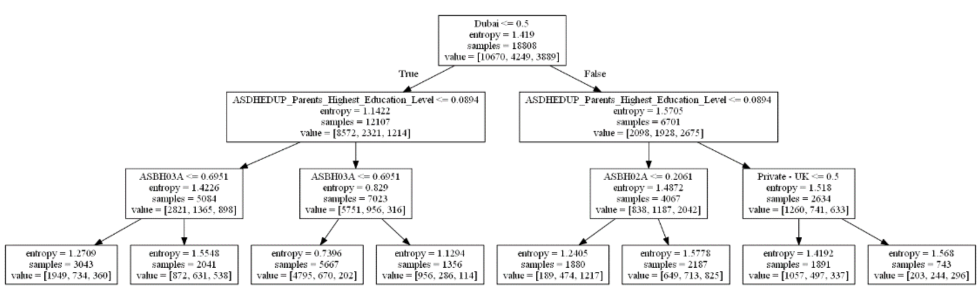


Table 19: Decision trees of the best prediction models

5.5. Predictive analysis on Dubai private schools

In this section we apply feature selection and classification techniques on Dubai's private schools data. Our aim is to predict if a school improves in TIMSS assessments considering all data features available.

We created two datasets, 'Dubai classrooms dataset' and 'Dubai TIMSS\Inspection dataset'. 'Dubai classrooms dataset' has selected attributes about classrooms from teachers and students surveys to look at what in classes affect students' performance. 'Dubai TIMSS\Inspection dataset' contains all TIMSS data in addition to inspection ratings in the year 2015 for Dubai's private schools.

Tables 20, 21 and 22 show the execution time, number of features selected, and the accuracy of the feature selection methods on Dubai's private schools datasets. Figures 59 and 60 plots the figures represented in Tables 20 and 22. Tree-based feature selection takes the lowest execution time to select features and it has the highest accuracy when applying it to 'Dubai TIMSS\Inspection dataset'. RFS feature selection method showed higher accuracy compared to other feature selection methods when applying them on 'Dubai classrooms dataset'. However, we consider the accuracy of the scheme of (feature selection method \ classifier) to get the best classification model.

	Feature Selection method	Dubai classrooms dataset	Dubai TIMSS\Inspection dataset
1	No Feature selection	0.0	0.0
2	Univariate Feature Selection (UFS)	0.06	1.75
3	Recursive Feature Elimination (RFE)	1.59	20.28
4	Tree-based	0.09	0.11

Table 20: Execution time of feature selection methods (seconds) on Dubai datasets

	Feature Selection method	Dubai classrooms dataset	Dubai TIMSS\Inspection dataset
1	No Feature selection	All (61)	All (440)
2	Univariate Feature Selection (UFS)	10	10
3	Recursive Feature Elimination (RFE)	10	10
4	Tree-based	9	9

Table 21: Number of features selected by different methods from Dubai datasets

	Feature Selection method	Dubai classrooms dataset	Dubai TIMSS\Inspection dataset
1	No Feature selection	-	-
2	Univariate Feature Selection (UFS)	78.00%	75.4%
3	Recursive Feature Elimination (RFE)	81.45%	71.99%
4	Tree-based	76.62%	76.62%

Table 22: Accuracy of feature selection methods on Dubai datasets

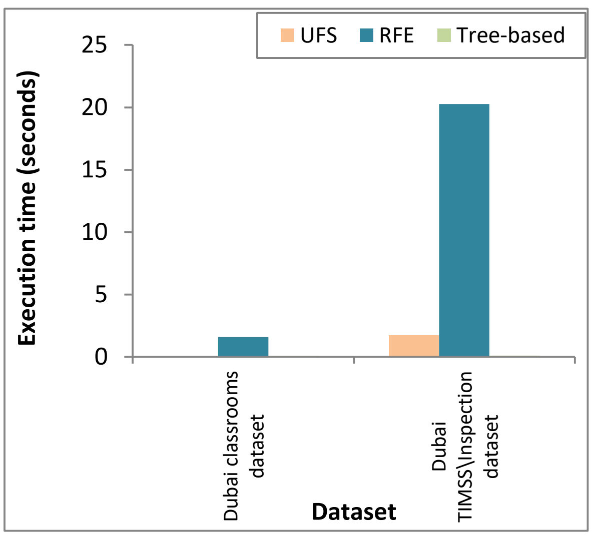


Figure 59: Execution time of different feature selection methods on Dubai's datasets

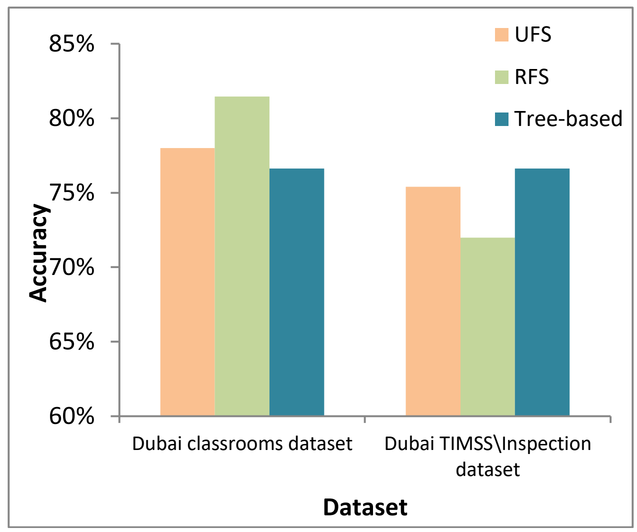


Figure 60: Accuracy of different feature selection methods on Dubai's datasets

Similar to what was presented in section 5.5.2, Table 23 shows the average accuracy of the (classifier \ feature selection method) schemes on Dubai's data. Tree-base feature selection method showed the highest accuracy of 98.5% with DT classifier for classroom dataset, and around 79% with KNN for TIMSS\Inspection dataset. Tables 24 and 25 show the performance of the best scheme selected and the resulted the confusion matrix. It is clear that the prediction performance on Dubai private schools' data is better than data of the UAE.

Classifier	Feature Selection Scheme	Dubai classrooms dataset	Dubai TIMSS\Inspection datase
Decision tree (DT)	No Feature selection	79.70%	63.00%
	Univariate Feature Selection (UFS)	78.80%	74.2%
	Recursive Feature Elimination (RFE)	81.00%	77.4%
	Tree-based	**98.50%**	77.3%
KNN	No Feature selection	94.95%	76.40%
	Univariate Feature Selection (UFS)	93.16%	76.91%
	Recursive Feature Elimination (RFE)	52.88%	42.03%
	Tree-based	96.91%	**78.81%**
SVM	No Feature selection	90.33%	72.83%
	Univariate Feature Selection (UFS)	82.64%	77.30%
	Recursive Feature Elimination (RFE)	78.64%	76.23%
	Tree-based	82.73%	74.87%

Table 23: Accuracy of classifier with features selection methods on Dubai datasets

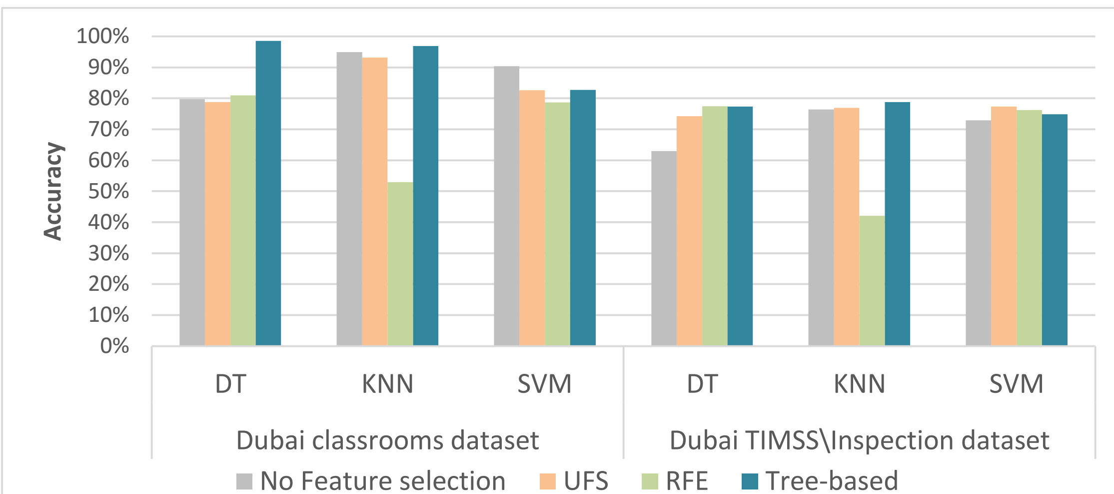


Figure 61: Classification accuracy

	Dataset	Best scheme	Average accuracy	Precision of improved class	Execution time (Sec)
1	Dubai classrooms dataset	DT \ Tree-based FS	**98.5 %**	99.42%	32.67
2	Dubai TIMSS\Inspection dataset	KNN \ Tree-based FS	**78.81 %**	99.07%	0.02

Table 24: Details of the best prediction scheme on the different datasets

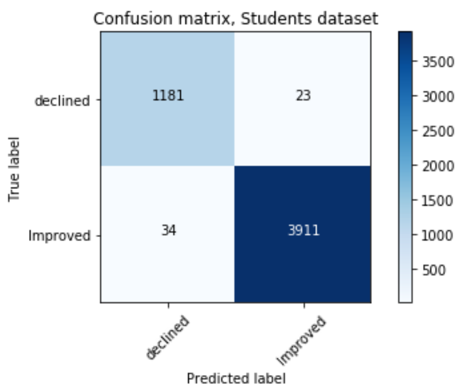
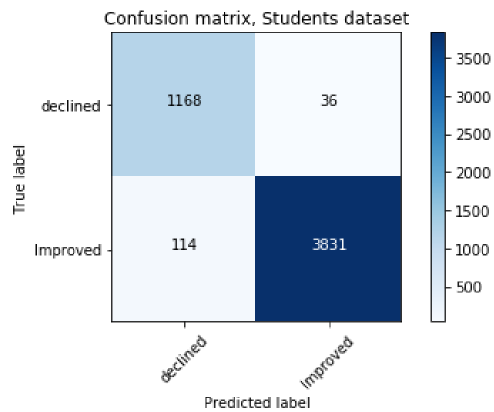
Dubai classrooms dataset	Dubai TIMSS\Inspection dataset
	

Table 25: Details of the best prediction scheme on the different datasets

Table 26 show the list of features selected from Dubai datasets based on the best prediction scheme.

Dubai classrooms dataset

1) 'ATBG14E' → Encourage classroom discussions among students
2) 'ATBG15E' → In your view, to what extent do the following limit how you teach this class? Uninterested students
3) 'ATBG15F' → In your view, to what extent do the following limit how you teach this class? Students with physical disabilities
4) 'ATBS03C' → Watch me demonstrate an experiment or investigation
5) 'ATBS03I' → Teacher ask students to read their textbooks or other resource materials
6) 'ATBS03J' → Have students memorize facts and principles
7) 'ATBS03L' → Take a written test or quiz
8) 'ATBS03M' → Work in mixed ability groups
9) 'ATBS06B' → When you assign science homework to the students in this class, about how many minutes do you usually assign?

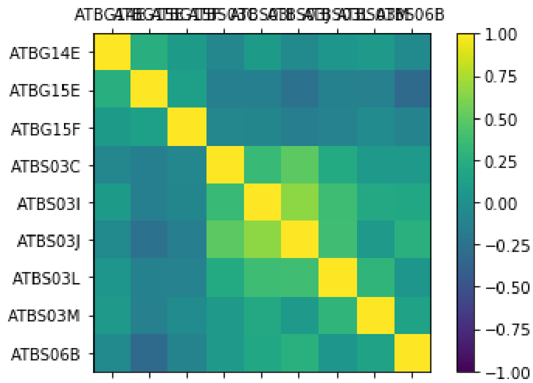
Dubai TIMSS\Inspection dataset

1) 'ASBH04B' → Language that child speak before he/she began school
2) 'ACBG16F' → Vandalism is a problem among Grade-4 students in school
3) 'ACBG18C' → About how many of the students in your school can do the following when they begin the <first grade> of primary/ elementary school? Read sentences
4) 'ACBG18J' → About how many of the students in your school can do the following when they begin the <first grade> of primary /elementary school? Do simple addition
5) 'ACBG22B' → Do you hold the following degrees in educational leadership? <Doctor or equivalent level—ISCED Level 8>
6) 'Insp_2015_Math_Att' → Inspection rating about attainment in Mathematics
7) 'Insp_2015_SEF' → Inspection rating about schools' self-evaluation
8) 'Insp_2015_Leadership'→ Inspection rating about the quality of school's leadership

Table 26: Key factors\features used to build prediction models with the highest accuracy

Looking at Table 27, the correlation matrix shows how strong are features selected. In the correlation matrix of Dubai TIMSS\Inspection dataset, it is clear that inspection ratings of mathematics, school self-evaluations and leadership are correlated and impact have impact on students' performance in science.

Dubai classrooms dataset



Dubai TIMSS\Inspection dataset

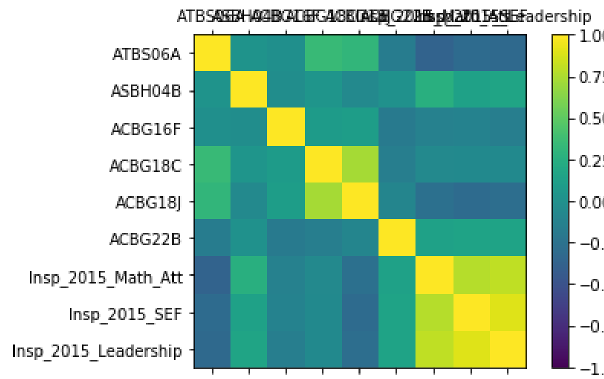


Table 27: Correlation matrixes of the selected features in the best prediction models

As a result of the predictive analysis, figure 62 shows the decision tree of the model predicting schools' improvements on Dubai classrooms dataset.

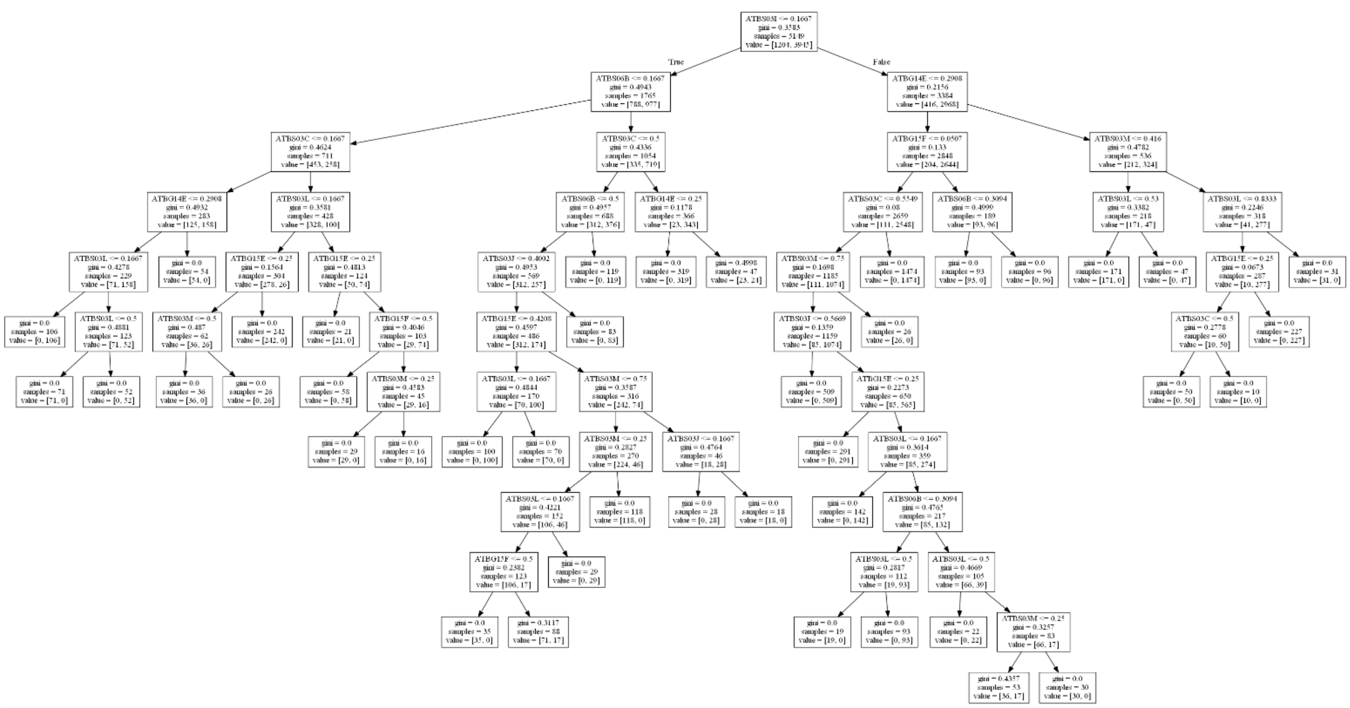


Figure 62: Decision tree of the best model predicting schools' improvements on Dubai classrooms dataset

6. Chapter Seven: Conclusion and Future Prospects

6.1. Conclusion

This dissertation provides clear breadth and depth analysis about Grade-4 students' performance in the UAE, and specifically in Dubai private schools. It presents a cause driven understanding about students' performance. Data mining was used to build a prediction model that forecasts future students' performance and scores improvements. In addition, we examined different feature selection methods associated with different prediction algorithms to build prediction models. It was proven that feature selection improves the accuracy and efficiency of the prediction model. Tree-base feature selection method was found to be the best feature selection method for TIMSS data.

6.2. Research Questions Answers

Based on the results presented in this dissertation, here are the answers of the research questions:

- **Does the use of feature selection methods improves the accuracy of predicting Grade-4 students' performance in science?**

Educational data is unique in the level of complexity and variety causing a big number of attributes in the data. Results in this study show that using feature selection methods increase the accuracy of predicting students' performance levels in science. In addition to reducing the execution time of the prediction model. As a result, embedded methods of feature selection (Tree-based feature selection method) showed the best performance when associated with classification algorithm.

- **Which classification method has the best performance in predicting Grade-4 students' achievement in the UAE?**

  Results discussed is section 5.4 showed that Decision Tree and SVM classifiers associated with tree-based feature selection method had the best performance in predicting Grade-4 students' performance in science on UAE datasets. Decision Tree and KNN showed the best accuracy of prediction on Dubai's private schools data reaching an accuracy of 98.5% on Dubai classroom dataset and 78.8% on the data of TIMSS and inspections.

- **What are the key factors impacting on Grade-4 students' science achievement in the UAE?**

Referring to the selected features considered in building the most accurate prediction model, the points below indicates the most important factors linked to Grade-4 students' achievements in science in the UAE:

- The development of children's literacy and numeracy skills in early years
- Extra tuition\ support in mathematics
- Parents' levels of education and qualification
- Parents believe that most employment needs skills in math, science, or technology
- Students' desire and needs and level of confidence
- School resources to provide library resources related to science, instructional materials and space
- School being a private school in Dubai
- School following UK or MOE curriculum

- **What are the key factors affecting the improvement of Grade-4 students in science at Dubai private schools?**

Referring to the selected features considered in building the most accurate prediction model, the points below indicates the most important factors which impact on Grade-4 students' achievements in science in the Dubai private schools:

<u>Factors to be considered in science classes:</u>

- Teacher encourage classroom discussions
- Uninterested students or students with physical disabilities in the classroom
- Teachers demonstrations of experiment and investigation
- Students reading of their science textbooks and other resources
- Students memorization of facts and principles
- The amount of written science tests and quizzes
- Students work in mixed ability groups
- Time given for science homework

<u>Factors to be considered in schools:</u>

- Science homework assigned to students
- The development of children's literacy and numeracy skills in early years
- Principals' degree in educational leadership
- Vandalism among Grade-4 students
- Inspection ratings of Mathematics, quality of leadership and schools' self-evaluation.

6.3. Future prospects

As a future work to be done,

- Conduct descriptive and predictive analysis on grade-8 students data in the UAE
- Consider students' scores in the sub categories of TIMSS assessment. For example, looking at students' scores in cognitive domains under TIMSS science in addition to the overall score in science
- Conduct wider analysis by including other International assessments results. For example: PIRLS and PISA
- Conduct analysis comparing UAE students' performance with other Arab countries
- Conduct a survey study on the current work done using EDM in international assessments

7. References

Mullis, I. (2012). *TIMSS 2011 international results in mathematics*. Boston:TIMSS & PIRLS International Study Center.

Martin, M., Mullis, I., Gonzales, E. & Chrostowski, S. (2004). *TIMSS 2003 international science report*. Chestnut Hill, Mass.:TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Hutchison, D. & Schagen, I. (2018). Comparisons between PISA and TIMSS: Are we the man with two watches. Lessons learned: What international assessments tell us about math achievement, (2007), pp. 227--261.

International surveys: PISA, TIMSS, PIRLS. (2015). Cambridge:Cambridge International Examinations,. Viewed 6 April 2018.

PISA 2015 Results in Focus. (2016). Paris:OECD Publishing.

International Findings from PIRLS. (2016). International Association for the Evaluation of Educational Achievement. Viewed 6 April 2018. http://pirls2016.org/download-center

Olson, D. & Delen, D. (2008). Advanced Data Mining Techniques. Springer Science \& Business Media.

Romero, C. & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. Expert Systems with Applications, vol. 33 (1), pp. 135-146.

Padhy, N. (2012). The Survey of Data Mining Applications and Feature Scope. International Journal of Computer Science, Engineering and Information Technology, vol. 2 (3), pp. 43-58.

"First-Rate Education System". (2018). [Accessed 6 April 2018]. Available at: https://www.vision2021.ae/national-agenda-2021/list/first-rate-circle

Liu, X. & Ruiz, M. (2008). Using data mining to predict K–12 students' performance on large-scale assessment items related to energy. Journal of Research in Science Teaching, vol. 45 (5), pp. 554-573.

Kılıç Depren, S., Aşkın, Ö. & Öz, E. (2017). Identifying the Classification Performances of Educational Data Mining Methods: A Case Study for TIMSS. educational sciences: theory & practice, vol. 17 (5).

Shahiri, A., Husain, W. & Rashid, N. (2015). A Review on Predicting Student's Performance Using Data Mining Techniques. Procedia Computer Science, vol. 72, pp. 414-422.

Askin, O. & Gokalp, F. (2013). Comparing the Predictive and Classification Performances of Logistic Regression and Neural Networks: A Case Study on Timss 2011. Procedia - Social and Behavioral Sciences, vol. 106, pp. 667-676.

Rajeswari, c., Sathiyabhama, B., Devendiran, S. & Manivannan, K. (2015). Bearing fault diagnosis using multiclass support vector machine with efficient feature selection methods. International Journal of Mechanical \&Mechatronics Engineering, vol. 15 (1), pp. 1-12. [Accessed 6 April 2018].

Leventhal, B. (2010). An introduction to data mining and other techniques for advanced analytics. Journal of Direct, Data and Digital Marketing Practice, vol. 12 (2), pp. 137-153.

Nielsen, F. (2017). Jupyter (formerly IPython notebook. [Accessed 6 April 2018].

Layton, R. (2017). Learning data mining with Python. Packt Publishing Ltd.

"TIMSS 2015 International Database". (2018). [Accessed 6 April 2018]. Available at: https://timssandpirls.bc.edu/timss2015/international-database/

TIMSS 2015 User Guide for the International Database. (2017). International Association for the Evaluation of Educational Achievement. Viewed 1 April 2018. https://timssandpirls.bc.edu/timss2015/international-database/downloads/T15_UserGuide.pdf

"Accuracy and precision". (2018). [Accessed 2 April 2018]. Available at: https://en.wikipedia.org/wiki/Accuracy_and_precision

"Publications Article | Knowledge and Human Development Authority". (2014). [Accessed 8 April 2018]. Available at: https://www.khda.gov.ae/en/publications/article?id=10170

Saeys, Y., Inza, I. & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. Bioinformatics, vol. 23 (19), pp. 2507-2517.

Chandrashekar, G. & Sahin, F. (2014). A survey on feature selection methods. Computers & Electrical Engineering, vol. 40 (1), pp. 16-28.

Enrolled Students by Education Type and Nationality - Emirate of Dubai. (2018). Dubai:Dubai Statistical Center. Viewed 1 March 2018. https://www.dsc.gov.ae/Report/Copy%20of%20DSC_SYB_2016_04%20_%2001.pdf

"Home | Knowledge and Human Development Authority". (2018). [Accessed 1 April 2018]. Available at: https://www.khda.gov.ae/en/Website

Nguyen, N. & Lee, H. (2008). Decision tree with optimal feature selection for bearing fault detection. Journal of Power Electronics, vol. 8 (1), pp. 101--107. [Accessed 14 April 2018].

Zhang, Y., Wang, S., Phillips, P. & Ji, G. (2014). Binary PSO with mutation operator for feature selection using decision tree applied to spam detection. Knowledge-Based Systems, vol. 64, pp. 22-31.

Farhangfar, A., Kurgan, L. & Dy, J. (2008). Impact of imputation of missing values on classification error for discrete data. Pattern Recognition, vol. 41 (12), pp. 3692-3705.