# Customs Trade Facilitation and Compliance for Ecommerce using Blockchain and Data Mining

# تسهيل التجارة والامتثال في الإدارات الجمركية للتجارة الإلكترونية باستخدام تقنية البلوكتشين والتنقيب عن البيانات

**by**

**OMAR ALQARYOUTI**

**Thesis submitted in fulfilment**

**of the requirements for the degree of**

**DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE**

**at**

**The British University in Dubai**

**July 2021**

# Customs Trade Facilitation and Compliance for Ecommerce using Blockchain and Data Mining

# تسهيل التجارة والامتثال في الإدارات الجمركية للتجارة الإلكترونية باستخدام والتنقيب عن البيانات البلوكتشين تقنية

**by**

**OMAR ALQARYOUTI**

**Thesis Supervisor**
**Professor Khaled Shaalan**

**Approved for award:**

| | |
|---|---|
| _____ | _____ |
| Name | Name |
| Designation | Designation |
| _____ | _____ |
| Name | Name |
| Designation | Designation |

Date: _____

## DECLARATION

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that the content of this research is the direct result of my own work and that any use made in it of published or unpublished copyright material falls within the limits permitted by international copyright conventions.

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

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

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

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

_____

Signature of the student

## COPYRIGHT AND INFORMATION TO USERS

**Dedication**

I dedicate this work to the people who made a difference in my life

*For My Father*

*(May Allah bless his soul)*

*For encouraging me to always seek knowledge and aim for the best*

*For My Mother*

*For her continuous prayers that made me stronger and full of confidence*

*For My Wife*

*For her love and the continuous motivation and support all the way*

*For My Children*

*For the energetic and positive power that they always give me*

*For My Brothers and Sisters*

*For being by myside throughout my life*

**Acknowledgement**

At the beginning, I would like to thank Allah for the countless blessings and for providing me with the strength and patience to complete this thesis. I hope that I have provided something useful that could contribute toward the advancement of science and society.

I would like to express my sincere gratitude to my supervisor Prof. Khaled Shaalan for the continuous support throughout my PhD journey and related research, for his patience, motivation, and immense knowledge. I could not have imagined having a better supervisor and mentor for my PhD study.

I would like to thank Dubai Customs for their support in facilitating obtaining the required dataset. I would also like to extend my thanks to Dr. Khaled Almiani for his support to overcome various obstacles that I came upon. Furthermore, great gratitude goes to my dear friend and colleague Oussama Jousou for giving me his precious time to discuss various business challenges and for his help in providing valuable inputs.

Special thanks go to my parents, my brothers and my sisters for their continuous encouragement and spiritual support throughout my life.

Big thanks to my family for all the support they have shown me through this research. For my kids, who continually provided the required breaks and the motivation to finish my degree with expediency! Finally, for my wife, I thank her for the time and unconditional support throughout the entire journey, without her tremendous understanding and encouragement, it would be impossible for me to complete my achievement.

## Abstract

Electronic commerce (ecommerce) has penetrated every society, organization, business and household and changed consumers' habits. It enabled businesses in some nations to trade beyond local borders and reach global proportions. This led to the explosive growth in demands for ecommerce platforms over the last few years and the increased popularity in cross-border trade interactions. The popularity became more evident in times of crisis such as COVID-19 for critical food and medical supplies and products. However, it was disrupted in other markets due to societies going on lockdown, which were further accentuated by borders being shut down. Ecommerce cross-border trade is impacted by regulations of each country. The challenges facing global trade and Customs administrations in particular cover many dimensions. Customs being tasked with protecting society and the smooth trade flow can no longer rely on traditional practices. A coordinated and consorted effort is required to disrupt illegitimate activities and support the mission of Customs. This work first aims to determine factors that drive the adoption of Blockchain technology. Blockchain is characterized for providing visibility, integrity, provenance and immutability across participants through the shared ledger capabilities. Therefore, blockchain technology is used in this study to build a framework to enhance trade facilitation and increase compliance while eliminating risks. This framework will provide advance access to information from various sources and will enable real-time discovery of risks. Accordingly, two off-chain clustering algorithms are proposed to determine value manipulation in ecommerce transactions and increase the efficiency of Customs Audit process. The Software Development Life Cycle (SDLC) methodology is adopted to build the framework. An integrated web application is developed to mock up the end-to-end process in ecommerce. Additionally, the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology is employed for modelling the two proposed clustering algorithms to identify transactional risks. The usability of the proposed framework is evaluated using the System Usability Scale (SUS) resulting in overall high acceptability levels across all users. Furthermore, accuracy measures are used to evaluate performance of the proposed clustering algorithms, reaching 86% for valuation assessment and 87% for risk identification in customs audit. The proposed framework will revolutionize the way trade supply chain is handled. It will create a shift from reactive limited visibility to proactive full visibility mode and properly manage various scenarios such as the current health hurdles and any future challenges lurching around.

## الملخص

اخترقت التجارة الإلكترونية (التجارة الإلكترونية) كل مؤسسة وقطاع أعمال ومجتمع وغيرت عادات المستهلكين على مستوى العالم. لقد مكّنت التجارة الإلكترونية الشركات في بعض الدول من التجارة خارج الحدود المحلية والوصول إلى أبعاد عالمية. أدى هذا إلى النمو الهائل في الطلب على منصات التجارة الإلكترونية على مدى السنوات القليلة الماضية وزيادة الاستخدام في التعاملات التجارية عبر الحدود. هذا الواقع جعل الحكومات تنضم إلى ركب التطور وبدأت تركز بشدة على تسهيل إجراءات التجارة الإلكترونية. أصبحت أهمية التجارة الإلكترونية أكثر وضوحًا في أوقات الأزمات مثل كوفيد-19 من خلال زيادة الطلب على الأغذية والمنتجات الطبية والإمدادات الحيوية. ومع ذلك، فقد تعطلت في أسواق أخرى بسبب استمرار المجتمعات في الإغلاق مما زاد من حدة إغلاق الحدود. تطلع الهيئات الجمركية المكلفة بحماية مجتمعها واقتصادها المحلي بتغيير الممارسات التقليدية. وهذا يعني الاضطرار إلى العمل بطريقة منسقة مع المعنيين ورفع درجة الاستعداد. في هذا العمل، تم اقتراح إطار عمل قائم على تقنية البلوك تشين. يهدف هذا الإطار إلى تبسيط عمليات تداول التجارة الإلكترونية باستخدام تقنية البلزك تشين مع ضمان احتياجات جميع الأطراف المعنية. سيحدث الإطار المقترح ثورة في الطريقة التي يتم بها التعامل مع سلسلة التوريد التجارية وخلق تحول من العمل في رد الفعل إلى الوضع الاستباقي من خلال تسهيل الوصول إلى المعلومات المسبقة التي يتم مشاركتها. سيتيح ذلك تنبؤ واكتشاف المخاطر في الوقت الفعلي وزيادة مستوى توافق الأنشطة التجارية. يتم استخدام المعلومات المسبقة من خلال الإطار المقترح لتحديد التلاعب بالقيمة الجمركية وتحديد المخاطر التي تشكلها الأطراف التجارية باستخدام تقنيات التنقيب عن البيانات. تم استخدام منهجية التنقيب عن البيانات من خلال العملية المعيارية لبيئات العمل المختلفة للتنقيب عن البيانات (CRISP-DM) لتمثيل وتحليل البيانات لتحقيق الأهداف المرجوة من هذا العمل. يتم تقييم التقييم الجمركي باستخدام النهج العنقودي (Clustering) والتقنيات القائمة على الكثافة. علاوة على ذلك، يهدف النهج القائم على تمثيل البيانات بطريقة كونفيكس (Convex) إلى تحديد ما إذا كانت شحنة شركة معينة تشكل أي خطر. سيعزز هذا الإطار حوكمة الأنشطة التجارية ويقيم مستوى الثقة بين المعنيين. تم تقييم إطار عمل البلوك تشين المقترح باستخدام مقياس قابلية استخدام النظام (SUS) الذي أظهر مستويات قبول عالية بشكل عام من قبل جميع فئات المستخدمين. كما أشار أداء النهج العنقودي المقترح بشأن تقييم التقييم الجمركي والنهج القائم على تمثيل البيانات بطريقة كونفيكس في التدقيق الجمركي إلى نتائج مهمة بدقة 86٪ و 87٪ على التوالي. يضمن الإطار المقترح التعامل السليم مع السيناريوهات المختلفة بما في ذلك العقبات الحالية التي يواجهها العالم وأي تحديات مستقبلية.

# Table of Contents

# List of Tables

# List of Figures

# List of Functions

# List of Abbreviations

| | |
|---|---|
| AJAX | Asynchronous JavaScript and XML |
| API | Application Program Interface |
| B2B | Business-to-Business |
| B2C | Business-to-Consumer |
| B2G | Business-to-Government |
| BoL | Bill of Lading |
| C2C | Consumer-to-Consumer |
| COO | Certificate of Origin |
| CRISP-DM | The Cross Industry Standard Process for Data Mining |
| DHL | Dalsey Hillblom Lynn |
| HS-Code | Harmonized System Code |
| IoE | Internet of Energy |
| IoT | Internet of Things |
| KDD | Knowledge Discovery in Databases |
| K-NN | K-Nearest Neighbour |
| LOF | Local Outlier Factor |
| MVC | Model-View-Controller |
| PBF | Permissioned Blockchain Framework |
| PBFT | Practical Byzantine Fault Tolerance |
| RFID | Radio Frequency Identification |
| SDLC | Software Development Life Cycle |
| SUS | System Usability Scale |
| TAM | Technology Acceptance Model |
| USV | User Selection Value |
| WCO | World Customs Organization |
| WTO | World Trade Organization |

# List of Publications

This thesis is based on three publications in international peer-reviewed journals and papers submitted for publications co-authored with Khaled Shaalan as follows:

1. **Alqaryouti, O.** & Shallan, K. (2020). Trade Facilitation Framework for E-commerce Platforms using Blockchain. *International Journal of Business Information Systems*. DOI: 10.1504/IJBIS.2020.10028490.
2. **Alqaryouti, O.** & Shaalan, K. Outlier Detection for Customs Post Clearance Audit Using Convex Space Representation. *Technology in Society, Elsevier*. (Under Review)
3. **Alqaryouti, O.** & Shaalan, K. Customs Valuation Assessment Using Cluster-Based Approach. *SN Computer Science, Springer*. (Under Review)

# 1. Chapter 1: Introduction

Trade supply chain is characterized by operating silos. Information is shared among parties using paper documents, making them vulnerable to manipulation and forgery. A cross-border transaction goes through many stages where each stage can have its own documentary requirements and data elements. Every bit of information exchanged is stained with suspicion and mistrust either directly or by association. The trade supply chain lacks much required visibility and consensus to function in a cost effective and efficient manner. This chapter provides a background on these aforementioned problems and details a solution for the key issues and challenges using blockchain technology and data mining techniques. This chapter also outlines the statement of the problem, purpose and objectives of this work, the research questions, the research rationale, and the focus and key contributions.

## 1.1. Background

### 1.1.1. Trade Supply Chain Process

The cross-border trade through the sea channel comprises exchange of various documents between the trade supply chain parties. The process of moving goods across-borders starts once an agreement is established between the exporter and the importer. As part of the requirements for this process, critical documents are required such as the Bill of Lading (BoL), the original commercial invoice, and the Certificate of Origin (COO). The BoL represents the document that is issued by the carrier to the shipper to acknowledge the receipt of the cargo for shipment. And the COO is a widely used international document in the global trade supply chain that indicates and attests the provenance of products.

These documents must be shared in original form between the exporter and importer. Both exporter and importer will use these documents to submit the customs declaration either directly or through a customs broker each in their own intended country in both source and destination. For instance, the exporter submits the export declaration at the origin country. The shipment is then shipped through a shipping agent. After that, the importer submits the import customs declaration at which the clearance process for the imported shipment will commence.

The customs clearance processes include risk assessment for the shipment. Part of this assessment is identifying malicious intent of value manipulation. The risk assessment process results in tagging risky shipments for mitigation which may result in physical inspection. The customs clearance will be issued following the payment of customs duties by the importer in order to clear the goods.

Once the shipment is cleared, the customs administration is responsible to target specific shipments for post clearance auditing. This process verifies all information and documents for the targeted shipments. In case of any discrepancies, the customs post clearance committee will contact the importer to clarify these discrepancies. If the importer fails to justify the variations with evidences, the customs administration will issue fines and penalties in addition to any additional customs duties. The process of post clearance audit heavily depends on a human determining the criteria based on which he will assess cleared transactions and gather the data for all transactions matching these criteria along with historical interactions and it is dependent on the skill-level and experience of the auditor. This procedure lacks accuracy, consistency, efficiency and effectiveness. Mistakes

made will result in wrong decision taken which leads to multiple negative effects such as cost of manual activities, wrong actions taken, and reduced customer happiness.

### 1.1.2. Ecommerce and Customs Declaration Process

The customs declaration application must be submitted to the customs administration in order to start the shipment clearance process. In ecommerce domain, the method of submitting this application depends on the trading scenario whether it is Business to Consumer (B2C) or Business to Business (B2B). In B2C, each shipment consists of few items that are shipped for personal consumption. In this scenario, shipping agents such as Dalsey Hillblom Lynn (DHL) perform bulky customs declarations. Therefore, shipments that are expected to arrive at the same time are declared together using the same application. The bulky customs declaration is important to increase the overall process efficiency due to the expected huge volume of trade in B2C (Ding, Huo & Campos 2017).

In B2B, the customs declaration process depends on the volume of trade. For small B2B shipments, the customs declaration process can be performed similar to the B2C scenario. However, for large shipments, each shipment will be declared separately. A customs declaration application may have over 120 fields. This is a time-consuming process to fill in the customs declaration application specifically in ecommerce trading domain due to the high volume of trade. This highlights the demand of automating the customs declaration process in the ecommerce trade.

Being that the process involves multiple parties and documents as well as handles various scenarios, the potential for risk and cost increase. Tight coupling between systems and practitioners is complex and comes with a high cost. What is needed is a universal

system that streamlines the exchange of information and introduces a trust system to establish a smooth flow for cross-border transactions. This system will drastically reduce the cost of doing business and at the same time, it will eliminate common risks associated with the trade community.

## 1.2. Statement of the Problem

Every cross-border trade transaction gets handled by various parties covering multiple industries both in public and private sectors. Some of these industries are governed by age-old formalities. These transactions involve processing and handling through various facilities and channels by different regimes. Each of the components has its own operational procedures, systems, infrastructure, conduct and regulations. The scope of these parties differs based on the width and breadth of its operations, market strength and outreach. Connecting all these parties and getting them to speak a universal language is unfeasible with traditional capabilities and will face a lot of resistance as it involves many competing interests. Regulations today create a challenge for ecommerce because each country has its own restrictions and prohibitions for certain goods and products. Ecommerce traders armed with knowledge of these regulations and rules are better equipped to determine whether goods or products can be shipped to destination country or not. This visibility is crucial to the smooth flow of cross-border ecommerce trade and can save consumers and companies unnecessary time or extra cost. Lack of sufficient knowledge of the rules and regulations governing certain cross-border trade in goods may lead to transactions taking longer to reach the consumer and increase the cost of transaction. A key challenge for ecommerce companies is the losses incurred due to returned goods for noncompliance with prohibitions, restrictions and regulations of the intended destination.

Authorities take actions and make decisions based on multiple factors using available data as supporting evidences. In some cases, these actions and decisions may be detrimental to local economies and societies. In part, this is due to either lack of sufficient data to determine legitimacy of transactions or the interpretation of it. By large, these actions are decided based on human perception and lack sufficient scientific basis that creates uniformity on how risks are determined and mitigated. At the same time, this prevalent practice is further accentuated by the reliance on manual activities and low level of automation, standardization and integration per industry. The protection of borders falls heavily on the preparedness of customs administrations who have to deal with sourcing information from all different industry formalities to complete the trade chain.

Most industries operate based on age-old laws and regulations. Most of their formalities are as ancient as the industry itself. Till date, these industries are paper-based and lack clear visibility to preceding trade supply chain activities.

Nations and companies are struggling to navigate this global maze and incurring cost along the way, which leads to raising the cost of doing business. Re-investing in reforming the principals that shape the industry is a daunting and costly task as well as difficult to establish consensus for. The increased intersections between industries are necessary in today's world. However, these interactions typically use industry-specific business language, which may have different interpretations in the different industries. This may lead to misinterpretation or fault mapping among parties. The impact of this scenario reflects in a negative way on decision-making, processing, handling and correlations. This lack of cross-industry harmonization inflames the mistrust between participating parties. Each industry practitioner adopts what they consider to be mature enough practices without

proper consideration for the end-to-end capabilities of partners in other industries. As a result, there are evidences of risks going unhandled properly and the efficiency and effectiveness of business conduct suffer. Majority of trade supply chain parties acknowledge these challenges and aspire to address them but, at the same time, they recognize that change will cause disruption to normal business operations which can be costly and not feasible within a desired timeframe.

In summary, the trade supply chain has to deal with these major challenges creating complex cross-border conditions and affecting global trade. The breadth and width of change is hindering the reform and modernization efforts as many parties with competing interests are involved in the decision-making process. Despite advances in technology and data mining techniques and their potential for improving the trade supply chain, there is no unified approach among trade parties including the Customs formalities.

## 1.3. Research Rationale

The challenges facing the cross-border trade supply chain are of global scale. Their impact spans economic, communication, technology, social, health and political realms. A much-needed change can only be achieved through an orchestrated active participation and contributions of concerned parties while recognising the complexity of the work involved. A Survey conducted by KPMG (a global network of firms providing advisory services) in 2019 indicates global technology industry leaders consider that ecommerce platforms represent the most disruptive business model over the next three years (Scally 2019). They expect ecommerce spending to increase from $3.5 trillion in 2019 to $6.5 trillion in 2023 (Lipsman 2019; Scally 2019).

As trade activities depend on paper-based documents and involve a number of manual steps, the rationale is to reduce the complexity into small manageable chunks. There are many areas to cover in the trade supply chain. However, this study aims to establish a unified digital approach to address critical issues in electronic commerce (ecommerce) from Customs perspective. The factors this study is taking into account relate to integrity, provenance, traceability and risk. Blockchain deals with storing and processing the data between the participating entities in the network. The need of an agreement between the participating entities guarantees the data integrity by establishing consensus. Moreover, the collaborative blockchain network set up between trade supply chain participants identify provenance and establish traceability of complete information about shipments from origin to market. Furthermore, the data mining techniques are used to analyse data collected through a unified approach to enable real-time discovery of risks and increasing the conformance level of trade activities in a proactive manner.

The beneficiaries of this work will see multiple values based on where they sit in the trade supply chain. The benefits will be most visible with the final recipient and the gate keeper embodied in customs authorities. This will guarantee that the end-to-end process is covered from a holistic viewpoint starting with initiation of trade transactions and ending with customs clearance for local market.

Today, advances in technology come with promising capabilities that revolutionize the inherited old-school practices. Trade supply chain practitioners see a huge potential in the adoption of blockchain technology to rectify the aforementioned challenges. It is perceived to transform the global trade supply chain and ecommerce domain which is evident in the growing number of blockchain initiatives disrupting traditional business models in various

sectors (Chang, Iakovou & Shi 2020). Therefore, this work illustrates how blockchain addresses the issues under consideration in trade supply chain and ecommerce domains from Customs perspective including integrity, provenance and traceability. Furthermore, this work is complimented with scientific data mining clustering techniques for risk and fraud detection to provide the necessary risk detection measures. The proposed solution will drastically reduce or eliminate the paper dependency and provide a digitized trade platform that guarantees faster to market delivery at reduced cost while ensuring compliance with rules and regulations.

## 1.4. Purpose and Objectives

The appetite for online shopping is growing exponentially causing certain governments to include ecommerce as part of their national strategy. Thus, the need for nations and organizations to be prepared to enhance every aspect of the end-to-end online shopping experience is crucial for global competitiveness. The ability to sell goods online by any person or small business introduces higher for trade supply chain parties including customs to ensure they meet their objectives with the most efficiency level. At the same time, ecommerce provides buyers with a wide variety of options to fulfil their needs.

The number of global ecommerce transactions has increased drastically over the last ten years and is expected to continue its upward trend (Lipsman 2019; Scally 2019). Compliance with shipping destination rules and regulations is an important factor for ecommerce platforms and trade supply chain parties to meet the ever-changing consumer demands. Therefore, ensuring sufficient data is available about the selling party, the product components, provenance, selling price, packaging information, shipping

information and many more which will enable faster delivery to intended consumer can be achieved seamlessly and risk free.

This ever-growing trading domain requires thorough and well-studied solutions from all parties, including customs administrations, to control the unprecedented ecommerce trading volume. These solutions should be able to address the lack of standardized global policies, and to tackle potential trade risks (WCO 2018a). Furthermore, these solutions should provide an automated environment that requires minimal physical interventions and ensures near real-time clearance for the ecommerce related shipments.

Researchers have investigated issues related to simplifying the ecommerce trade (Chen et al. 2017; Xie et al. 2018; Ying, Jia & Du 2018). However, this study aims to contribute to this growing domain by building a framework that can be integrated to any ecommerce platform. Also, this framework handles trade-related security issues by providing mechanisms to identify risks associated with the end-to-end trade conduct.

A necessary feature of the proposed solution is the ability to prevent any form of data manipulation. Through research, the only technology today with this capability is blockchain as immutability is one of its top characteristics (Zheng et al. 2017; Ahmad et al. 2021). Therefore, in this work, a blockchain-based framework is proposed. This research aims to streamline ecommerce trading processes using blockchain technology while ensuring the needs of all concerned parties. The proposed framework will revolutionize the way trade supply chain is handled and create a shift from working in reactive limited visibility to proactive full visibility mode through facilitating access to

advance information sharing. This ensures that challenges related to simplicity, traceability, scalability, efficiency and trade risks are addressed.

This study seeks to address the following objectives:

1. Determine factors that influence the adoption of blockchain technology in ecommerce platforms.
2. Build customs enabled blockchain-based framework to enhance trade facilitation and compliance in ecommerce domain.
3. Develop a clustering technique that utilizes the proposed blockchain-based framework to identify risks in shipments for Customs post clearance audit by utilizing the proposed blockchain-based framework
4. Develop a clustering technique to determine risks related to value manipulation in the ecommerce transactions by utilizing the proposed blockchain-based framework

## 1.5. Research Questions

Although there are many challenges to address in the trade supply chain, the target in this study is to enhance the cross-border ecommerce activities from customs perspective. This study attempts to establish synergy through the use of advance technology and analytical capabilities to reduce the economic effects of illegitimate trade activities. With the research objectives in mind, this work aims to answer the following questions:

1. What factors will influence the adoption of blockchain technology in ecommerce platforms?
2. How can blockchain enhance trade facilitation and compliance in ecommerce trade domain?

3. Can clustering techniques be used to identify risks in shipments for Customs post clearance audit by utilizing the proposed blockchain-based framework?
4. Can clustering techniques be used to determine risks related to value manipulation in the ecommerce transactions by utilizing the proposed blockchain-based framework?

## 1.6. Focus and Key Contributions

Ecommerce is an attractive platform for trading in goods to fulfil various needs through multiple channels. One of the key steps of ecommerce cross-border trade is the role of customs. Customs are typically focused on dutiable goods and less interested in small retail related transactions for personal use. today, it costs customs around the world large sums of money and time spent handling ecommerce transactions without real value in return. This is not economical for customs administrations and clear distinction between retail and non-retail ecommerce transactions has to be made to reduce the financial burden on both customs and consumers.

This work investigates the development of a mechanism for sharing data among the various interested trade parties that allows customs to process these activities with the minimal possible validation and verification steps to reduce unnecessary costly interventions. Additionally, the use of scientific data mining clustering methods that enhance risk and audit capabilities is leveraged by recognizing trade patterns and detecting fraudulent activities.

### 1.6.1. Build Customs Enabled Blockchain-Based Framework to Enhance Trade Facilitation and Compliance in Ecommerce Domain

Consumer sentiments are negatively impacted by unexpected disruptions in their ecommerce trade journey. The complex regulations and procedures hamper efforts to facilitate trade activities and leads to less than desired outcomes.

To enhance the consumer experience when trading over ecommerce channels, a customs-enabled blockchain-based framework is proposed. This framework aims to streamline and simplify the handling and clearance of ecommerce trade while ensuring proper compliance is in place. This should help customs administrations focus on transactions with economical value and fast-track the process for goods with zero duty.

The capabilities of blockchain cover many of the desired techniques needed as part of transforming the trade supply chain; shared ledger capability helps synchronize data among all participating parties, immutability guarantees integrity of data, provenance helps determine origin of goods and products to prevent fraud and establish traceability and smart contracts capability ensures the consensus to establish cross-party workflow. These capabilities give blockchain technology a leading edge over other offerings that do not have these needed capabilities out of the box and will require larger investment.

Ecommerce transactions consist of features that, if shared properly by the original source with the rest of the trade supply chain parties, will result in faster deliver-to-market process. Consequently, this will translate into cheaper trade transactions and more effective and efficient procedures. This data will provide visibility and traceability with ensured integrity supported by the blockchain consensus and smart contract capabilities to facilitate

trade and ensure compliance to regulations. Additionally, it will result in increased levels of customer satisfaction and productivity leading to reduced operational cost.

### 1.6.2. Develop a Clustering Technique That Utilizes the Proposed Blockchain-based Framework to Simplify the Post Clearance Audit in Customs

Audit plays an important customs role in which the past transactions of a company are investigated for non-conformance. Traditionally, audit is conducted based on a selection of a sample group of transactions for investigation, which at times results in wrongful audit. Many resources are wasted in trying to prove the legitimacy of company activities. This process is also based on human perception and analytical capabilities without sufficient evidence of the reasons a company was selected for audit.

Using the analytical capabilities of the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology, Customs will be able to assess company behaviours and detect outliers in the data received as part of Customs declarations and employing clustering algorithm that uses convex representation to establish patterns in data and rely on certain data-driven facts to determine abnormal behaviour. This clustering technique analyses the behaviour of a trader with its peers based on historical data to determine deviations from the norm. Thus, any company related data can be analysed to determine abnormal behaviours with companies in the same line of business. This technique will provide much needed capabilities to protect the revenue and secure the trade supply chain from illegitimate activities. This approach will significantly reduce or eliminate false audits and provide better resource management, thus reducing operating cost for both customs and the entity being audited. CRISP-DM has been used in many different domains and

have been proven to be successful. However, it was never applied in ecommerce trades from customs post clearance audit perspective.

### 1.6.3. Develop a Clustering Technique to Determine Value Manipulation in the Ecommerce Transactions Through the Proposed Blockchain-based Framework

Customs duty obligations are primary revenue sources for nations. Any under or over value can result in economic instability. The role of customs is safeguarding society and local businesses from criminally guided value manipulations.

To calculate a custom value properly, we propose a technique in which the shipment related information is represented into geometrical space to determine whether a wrongdoing is developed. This representation simplifies the process by identifying relationships among the various shipment features. Consequently, correlations are established to detect anomalies in declared values. The proposed framework is based on blockchain technology which will enable the use of real-time information to identify value manipulation. The CRISP-DM methodology is employed to represent data in clusters using density-based techniques that allows customs to analyse data on the fly and determine if there is any manipulation in the declared goods values. CRISP-DM was never applied in ecommerce trades from customs value manipulation perspective.

In the proposed approach, the similarity factor for a given commodity with that of similar past transactions is determined. This technique provides much needed capabilities to protect the revenue and securing the trade supply chain from illegitimate activities.

## 1.7. Thesis Outline

This thesis includes seven chapters that illustrate the collective proposed solution. These chapters use materials from three publications or papers that are under review in international per-reviewed journals by the author and co-authored with Khaled Shaalan (Alqaryouti & Shaalan n.d. b, n.d. a; Alqaryouti & Shallan 2020). Chapter 4 is based on (Alqaryouti & Shallan 2020), chapter 5 is based on (Alqaryouti & Shaalan n.d. b) and chapter 6 is based on (Alqaryouti & Shaalan n.d. a). These chapters have their own substructure that also include the methodology and discussion of results.

This chapter provided an introduction to the topic of this thesis and illustrated the key problems and challenges that will be addressed in this work. The remainder of the thesis is organized as follows:

- **Chapter 2** provides a comprehensive literature review on the work done in the area of cross-border trade using blockchain technology with focus on ecommerce. In addition, it provides details on the approaches used in the literature to detect outliers and abnormal behaviours.
- **Chapter 3** illustrates the research methodology followed in designing the customs enabled ecommerce framework using blockchain technology and data mining techniques.
- **Chapter 4** provides the details in which the blockchain-based framework is built and how it streamlines the ecommerce trading process from customs perspective through the use of blockchain technology.

- **Chapter 5** presents the clustering technique that utilizes the proposed blockchain-based framework in chapter 4 to simplify the post clearance audit in customs and identify company shipments that pose suspicious behaviour.
- **Chapter 6** presents the clustering technique that is used to determine customs value manipulation in the ecommerce transactions through the proposed blockchain-based framework in chapter 4.
- **Chapter 7** concludes the study and explains how trade-related traceability and security issues with the end-to-end trade conduct are handled. Furthermore, this chapter summarizes the thesis and the key findings, and it discusses the theoretical and practical implications, limitations of this study and the recommendations for further study.

## 1.8. Conclusion

Historically, the trade supply chain faced many challenges resulting from age-old formalities and limitations in adopted technologies. Today, organizations are experimenting with technological capabilities that come with the promise of revolutionizing industries and practices that are ancient. The trade supply chain sees potential in automating much of its activities, which became possible with the introduction of advanced technologies such as blockchain. This has paved the way for more open and connected global trade. This chapter provided a background on the research topic and tackled various challenges that face the trade supply chain from customs perspective. Furthermore, this chapter highlighted the statement of problem, purpose and objectives, the research questions, the research rationale, and the focus and key contributions of this study. The next chapter will investigate the factors that influence the adoption of

blockchain technology in ecommerce platforms and how it can improve the global trade supply chain. Additionally, it discusses the various data mining clustering techniques that can be adopted to determine abnormalities and outliers.

# 2. Chapter 2: Background and Literature Review

In the contemporary era, there are a number of factors that created opportunities for the advancement of different fields such as trade and commerce, both locally and globally. Progress has been witnessed in the global business environment wherein it underwent significant shifts such as leveraging on innovation and technological advancements. Blockchain technology is one of the most promising technology introduced in the modern time being adopted by organizations in lieu of enhancing capabilities particularly in supply chain, ecommerce and international business. The application of blockchain transforms traditional business activities and inspires new ways of doing business (Hooper & Holtbrügge 2020). Thus, blockchain applications provide various improvements such as automating trade and offering online payment alternatives for ecommerce and international transfers (Hooper & Holtbrügge 2020). This suggests that blockchain can be viewed as being useful in global trade supply chain and ecommerce.

Blockchain technology provides four key features that can contribute to enhancing the coordination and integration of key players of trade supply chain including transparency, validation, automation and tokenization (Blossey, Eisenhardt & Hahn 2019). In addition, blockchain offers promising opportunities for enhancing capabilities of ecommerce innovations such as payment alternatives like Bitcoin, higher security and enhancing trust for platforms and gateways (Lim et al. 2019). In relation to these opportunities offered by blockchain, this study focuses on the application of blockchain towards improving global trade supply chain and ecommerce.

Successful global trade supply chains centers on comprehensive and synchronized management of the flow of product, processes, information and financial resource and plays a pivotal role in the development of ecommerce businesses. Blockchain applications are argued to transform global trade supply chains and ecommerce practices and activities. Therefore, this review aims to understand the role of blockchain in improving global trade supply chain and ecommerce.

The majority of the companies worldwide are attempting to innovate new forms of businesses towards achieving enhancements of capabilities in cost-efficient ways. The motivation of this research stems from the desire to understand the opportunities that emerging technologies, such as blockchain, provide in improving business processes and practices towards achieving increased competitiveness and future success. According to McDaniel and Norberg (2019), Blockchain presents opportunities for reducing the time and expenses required in the facilitation of trade, provision of real-time information on flow of goods, and improvement on management of supply chains, which are important for fostering successful businesses. Thus, the interest in blockchain technology as well as its applications in businesses has grown over the last few years.

Many studies on blockchain focus on the application of bitcoin and cryptocurrency. However, the application of blockchain in supply chain management and ecommerce has been given little attention in literature (Blossey, Eisenhardt & Hahn 2019). As such, the importance of this research is that it attempts to fill the knowledge and awareness gaps on the use of blockchain in improving global trade supply chain and ecommerce. This chapter aims at investigating the factors that influence the adoption of blockchain technology in ecommerce platforms and how it can improve the global trade supply chain. Furthermore,

it aims to exploring the various approaches to mine the available data for the detection of abnormalities and outliers.

## 2.1. Blockchain Overview

Blockchain was introduced to address security and traceability issues in untrusted environments (Zheng et al. 2016). Blockchain deals with storing and processing the data between the participating entities in the network, where the processing part is performed through the use of smart contracts. Smart contracts are digital protocols in which rules, regulations and business rules are defined. They provide digital facilitation, verification and enforcement for business logic between the participating entities (the network nodes). For instance, in the ecommerce domain, the network nodes can represent the customs administration, the seller, the buyer and the trading platform (Zheng et al. 2017).

In blockchain technology, information is stored in a ledger that represents the transaction data. A ledger consists of several blocks that are connected to each other through a hash value. A block hash value is obtained using a hash function, where each block contains the hash value of the previous block. The hash value will change in case of any revisions in the block transaction values, denoting the blockchain nature of irreversibility and tractability. Thus, any update in the block's details changes the hash value, resulting in breaking the chain. Each participating entity (network node) in the business network have exactly the same replicated copy of the ledger, where the ledger represents the personal copy of the transaction data. This ensures the immutability and authenticity of the overall transactions' history. This blockchain structure is considered as a decentralized technology (Zheng et al. 2017).

In order to add a new block to the blockchain, 51% of the participating entities should agree on this block details (Efanov & Roschin 2018). This is achieved using the consensus mechanism in which entities are required to solve a mathematical puzzle. The entities involved in solving this puzzle are known as miners. Based on the applications domain, the miners will be rewarded with cryptocurrency for their efforts. To solve a mathematical puzzle, each miner has to obtain a nonce (number of digits with particular size) such that entering the nonce and the block to the hash function results in a pre-determined range of value (David et al. 2018). This requires a high computational power to resolve the mathematical puzzle. The entities who first solve the puzzle broadcast its solution to the other entities to confirm it. The block is announced as part of the blockchain when 51% of the participating entities confirms the correctness of the solution. The complexity of the puzzle and the consensus mechanism strategy depends on the application domain and the overall security requirements. The example presented above describes the proof-of-work consensus mechanism, where lighter mechanism can be employed in less restricted environment (Puthal et al. 2018). The main benefit of this consensus mechanism is to guarantee the data integrity.

Blockchain is categorized into three types, namely private, public and consortium blockchain. These types of networks are mainly distinguished by the user's identity, the employed consensus mechanism, and the role of each node in the network (Lin & Liao 2017).

In public blockchain, the identity of the participants is anonymous (unknown) and anyone can join the network. All nodes in public blockchain are equally treated, and all nodes play the role of maintaining the ledger and endorsing the transactions. The possibility

of joining the network by any user increases the security pressure in this type of network. Therefore, strong consensus mechanisms such as proof-of-work must be deployed to ensure the integrity of the data. In public blockchain, all nodes play the same role of maintaining the information. This reduces the scalability of the overall architecture. For instance, the Ethereum blockchain network can process up to 20 transactions per second. In term of application domain, this type of blockchain is suitable in cases where the identity of the users can be kept anonymous and anyone has the right the join the network. These characteristics highlight this type of blockchain as a suitable environment to build ecommerce related applications. This is the main reason for designing the proposed framework on the Ethereum blockchain network (Lin & Liao 2017; Zheng et al. 2018).

In private blockchain, the identity of the participating entities is known and confirmed. This reduces the importance of the consensus mechanism because participating entities have an established level of trust. Additionally, the role of maintaining the ledger and endorsing the transactions is not necessarily performed by the same nodes, which increases the scalability of the private blockchain architecture. Consortium blockchain is similar to private blockchain, but several trusted entities handle the administration tasks unlike the private blockchain which operates under the authority of a single entity (Lin & Liao 2017; Zheng et al. 2018).

The cost associated with building a blockchain-based application depends on the type of the employed blockchain. In private blockchain, the cost is mostly related to establishing the required infrastructure to host and maintain the ledger. In public blockchain, the ledger is hosted on a public blockchain such as Ethereum. In this case, users are charged based on the amount of computations they perform (Zheng et al. 2018).

## 2.2. Blockchain-Based Applications

Since the emergence of blockchain, there has been an increasing amount of literature on a variety of its applications (Seebacher & Schüritz 2017). Previous research has focused on how blockchain can address the traceability, security, and scalability challenges in various domains such as cryptocurrency (Liu, Li, et al. 2018; Alqaryouti et al. 2019b, 2019a), trade supply chain (Feng Tian 2017), and energy (Li et al. 2017). Traceability is considered as a ground base in ecommerce trading process since improper traceability impacts the overall process performance and reduces the satisfaction level among participants. In terms of security, the anonymity of participating customers highlights the importance of employing proper security mechanisms to ensure the legitimacy of trade. This is due to the normal registration process of ecommerce public platform may not involve any user identification process. Scalability in any ecommerce trading platform is a major challenge caused by the fast-evolving trading environment. Thus, it is essential to maintain the overall performance of any ecommerce trade platform.

Blockchain technologies are emerging in different business processes, systems, models and practices. The introduction of blockchain prompted disruptions to the traditional business processes thereby suggesting the importance of blockchain applications in businesses and industries (Casino, Dasaklis & Patsakis 2019). As such, blockchain technology and its applications are emerging as promising solutions to rebuilding the way businesses operate. In line with this, supply chain management is one of the potential applications of blockchain (Chang & Chen 2020). According to Wang *et al.* (2018), blockchain-based applications such as cryptocurrency, internet of things, risk management and financial services have a number of benefits including decentralization, persistency,

anonymity and auditability. This suggests that different blockchain-based applications present opportunities for improving business operations, processes and practices towards moving to a sustainable future.

This section provides detailed discussion on proposals in the literature according to components that are designed to optimize. Thus, the following subsections are divided into performance and trade improvement. In performance improvement, particular performance related issued are addressed. Whereas in trade improvement, the improvement of different trade domain functionalities is studied.

### 2.2.1. Blockchain for Performance Improvement

Xia et al. (2017) developed the blockchain-based system (MedShare) that facilitates the sharing of the patient's medical data between cloud providers. This sharing is established while ensuring the consensus of the patients and the actual behavior for each entity is monitored to ensure that no unauthorized actions are performed by any party. This system is divided into the following layers: user layer, data query layer, data structuring layer, and database infrastructure layer. User layer defines and classifies all of the users that are allowed to access the data for research tasks. The data query layer defines how the users can access the data. In such system, data is expected to be stored in different databases and the data structure layer handles how and who can access the data and also record all of the initiated requests to the blockchain. The database infrastructure layer handles the passing of the datasets between the peers. This system targets the main two features which are behind the existence of the blockchain technology, namely traceability and security. However, the regulator entity should be owned and handled by trusted authority since it has access to very critical information.

Likewise, Jiang et al. (2018) built a blockchain solution termed BlockHIE. This solution consisted of three components: blockchain network, medical institutes, and patients who are willing to store their data on the system. In this solution, blockchain is used to store the patient users and manage the information sharing between the medical institutes. Medical institutes use the stored information during the diagnosis and treatment of the patients. Beside the traditional way of uploading medical reports to the systems, the system gathers the medical information through the use of IoT devices such as smart watches and smart thermometers. Once the hospital creates a medical record for a patient, three copies of this record are generated. The first copy is sent to the patient and the second copy is kept at the hospital. These two copies are totally identical. The third copy, which is kept in the blockchain, represents a summary of the report information and it acts as a proof-of-existence for the other two copies. This is due to the limited storage capabilities on the blockchain. In addition, uploading relatively large amount of data to the blockchain requires a huge throughput. The provided system ensures the traceability of the initialized report in emergency related situations. However, accessing the actual historical tests in timely manner is more important than locating their existence. Similarly, Turkanović, Hölbl and Košič (2018) proposed Blockchain-based solution to address and manage the process of initiating student's certificates and attestation.

To reduce the amount of computation used by the consensus protocol, Alzahrani and Bulusu (2018) suggested a random validator selection mechanism. In this mechanism, nodes are assigned randomly to validate each other. In addition, once a block is required to be validated, only $\log\frac{n}{3}$ of the node will perform the mining process. These nodes are selected at random with the objective of distributing and producing the energy consumption

related to the consensus activities. This system is designed mainly for public blockchain and this highlights the importance of the author provided consensus mechanism. However, such system architecture can be easily moved to private blockchain and that will reduce the pressure on the consensus mechanism.

Chen et al. (2017) presented a Blockchain-based framework for information sharing and quality control. This framework consists of four layers, namely IoT layer, data layer, contract layer and business layer. All of these layers work together to support and provide the required information to help achieving the business goals. The IoT layer consists of sensor and RFID technology that can be used to monitor and capture the required measure from the business perspective. The data layer is the blockchain layer and all of the capture information in the sensor layer is added to the blockchain layer to support the other two layers decisions. The contract layer represents the logic of the system and it is implemented through the use of the smart contract. This layer supports the business layer by providing the required inputs. The business layer represents the global objective of the network and it uses the provided information by the contract layer to make the final business decision. This architecture seems to be self-contained and logical. However, detailed investigation about the objectives and drawback of adopting such architecture must be performed by the authors.

Lu and Xu (2017) developed origin-chain, a blockchain-based solution to provide tamper-proof data traceability and regulation validation management system. In this system, the product supplier contacts the traceability provider in order to arrange for the traceability contract. This contract represents the type of validation and must be performed by the service provider. The business flow in this scenario consists of three types of entities,

namely suppliers, labs and traceability service provider. The actual process of tracing is performed through the use of the factory contract. This contract issues two types of services: registry contract and service contract. Registry contract contains the legal information for the agreement between the service provider and the supplier. On the other hand, service contract is used to add or remove conditions to the original traceability agreement. The lab periodically uploads the test results based on the type of the agreement at the traced product.

Li and Wang (2018) proposed a blockchain-based system that ensures the privacy of the transactions in financial platforms. The main parties in this system are: the verifiers, the participants and the supervisors. Any participant can play the role of the verifier to validate the transaction on the blockchain. Participants refer to the customer and service provider where the supervisor represents the third-party authority, which ensures the satisfaction of the financial constraints. In this system, the supervisor makes sure that the participating entities in the trade are behaving within the legitimate boundaries. Once two or more participants start conversation action to perform a specific deal, this supervisor monitors the information exchange and interfere whenever necessary. Again, this system targets the security and traceability of information through the use of blockchain technology. However, the scalability factor of this system should be investigated.

### 2.2.2. Blockchain for Trade Improvement

Transactions of international trade are described to be generally high value, which means that one of the most important roles in relation to their execution lies with partner countries' public authorities (Belu 2019). As such, the application of new and innovative solutions is needed to promote simplification of the transaction mechanism in the context

of international trade. According to Belu (2019), the application of blockchain technology in international trade can offer a number of advantages including real time verification of documents, disintermediation, decentralization of contract execution, transparency in proof of ownership and reduction in transaction costs. Thus, the application of blockchain technology can contribute to revolutionizing international trade but it must be noted that it must not be seen as a solution to everything (Ganne 2018). This means that whilst there are some challenges, blockchain can open a number of opportunities towards enhancing process efficiencies in the international trade context.

International trade can generally be described as exchanging of goods in a global scale. This means that international trade is complex in nature because its processes are facilitated by various factors including financial, socio-political and technological (Derindag, Yarygina & Tsarev 2020). As such, the same authors noted that the application of blockchain technology has the potential to revolutionize international trade on a continuous basis and can contribute to reducing diverse trade barriers. In addition, the application of blockchain technology can also influence technological reengineering of financial processes in trade finance (Chang, Luo & Chen 2020). Therefore, the application of blockchain technology in international trade can serve as an innovative platform towards improving traditional trade processes whilst enhancing transparency on trade transactions.

The application of blockchain technology in international trade can contribute to the digitalization of processes and enable automation. Accordingly, blockchain technology can drive automation of processes particularly in relation to cross validation of import declaration without the need for a central ledger (Segers et al. 2019). Therefore, it can be noted that blockchain is an emerging technology that can play an important role towards

automating international trade processes and enhancing cross validation in custom administration.

In this part, related literature is categorized according to the trading domain into energy trade, trade supply chain and ecommerce trade.

#### *2.2.2.1. Energy Trade*

With the presence of smart city, Internet of Energy (IoE) has emerged as a promising approach to increase the efficiency of energy usage and sustainability. In this context, Alcarria et al. (2018) developed an authentication system that uses blockchain technology. This system aimed to facilitate smart cities application process and provides trusted data approach. Nowadays, every household is equipped with smart devices that can help in manage day-to-day needs (smart meters). In this work, the system considered the amount of energy used by the households as the main assets. In the electricity usage use-case, the authors assumed that the smart city is equipped with high-tech sensors and devices. Each user knows his own electricity usage pattern, and therefore he can use this information to plan a head his electricity demand. The authors assume that there will be devices that can be used to sell electricity to other households in similar settings. However, it is essential to consider and respect the user's privacy and this can be achieved by the use of blockchain. Similarly, Su et al. (2018) proposed Blockchain solution that utilizes the smart contracts concept to improve and secure the energy trades between the community members.

In terms of energy trading for electric vehicle domain, Liu et al. (2018) developed a blockchain-based solution for electric vehicle charging using smart grid platform. Through this solution, the driver publishes electricity charging and/or selling request. This

request is then entered in the blockchain trading platform once the request has been signed by the driver. Once the participating peers are matched based on their selling and buying requests, both peers confirm and agree on the trading constraints to finalize the transaction.

#### *2.2.2.2. Trade Supply Chain*

One of the challenges in trade supply chain is how to maintain food safety. In this essence, Feng Tian (2017) suggested multi-component framework using HACCP method. This method consists of several strategical tests that must be performed at several points during the supply chain cycle. The proposed framework combines the benefits of the Blockchain and IoT to improve the efficiency of the HACCP method. Blockchain is used to ensure the traceability and the immutability of the data. The framework employed IoT devices to collect critical and necessary information about the environment of the shipments. The sensor devices in the proposed use-case are used to monitor the lifecycle of the harvested planets. In addition, sensors are also deployed at the warehouse and the distribution points to monitor the temperature of the storage area and the shipment environment. As author did not describe the important technical information for this deployment and therefore his framework can be seen as a conceptual framework.

Vos et al. (2018) addressed the problem of traceability and security throughout the international trade supply chain using a blockchain-based solution. The authors assume that the economic operator trusts the customs agents at their country. In addition, the authors assume the customs agents across borders do not trust each other. In this system, the economic operators insert or remove packages. The authors assume that the participating entities are the economic operators, customs agencies, containers and packages. In this line, containers and packages are actually the physical assets that moves

between the entities. The shipment process starts by using the claim submission protocol. Using this protocol, operators include all of the required information about the shipment and identify its path. Then, the container claims validation protocol and validates this information before submitting them to be to the blockchain. The certification protocols allow the customs to register and invoke the participated operators.

Imeri and Khadraoui (2018) demonstrated conceptual framework to address the problem of transporting dangerous goods through the use of blockchain. The transportation of such products normally involves too many restrictions. For instance, the exact path and speed of the transportation vehicle must be monitored. In addition, other related information such as the temperature of the tank must be monitored as well. The authors addressed these problems through the use of smart contracts. In this direction, smart contracts are used to track the movement path of the vehicle and to monitor the status of the shipment continuously throughout the cycle. Smart contracts are also used as an emergency notification system to inform the authorities if any situation occurs. The provided solution covers the emergency goods transportation, and this solution addressed an important challenging problem. However, based on the type of goods, a real-time reaction could be vital, and therefore technical and detailed experiment to validate the efficiency is required.

Similarly, Wu et al. (2017) described a blockchain-based solution to facilitate the process of tracking the distribution of goods. This system consists of centralized server nodes, set of peer nodes, administrative node and external monitoring node. The server node has the address of all participants, and the set of peers represent the participating entities, namely supplier, carrier and customer. The administrative node is a special node

that has access to all of the information generated by the all participants. The external monitoring nodes validate different issues related to the shipment based on the type of distribution agreement. For instance, the geographic location of the truck can be validated on the movement pattern. This system supports three types of events: shipment initializing, custody event, and monitoring event. Custody event shows the current participating owner for the product. Monitoring event is used for the track movement according to the geographical location.

Bocek et al. (2017) developed modum.io, a pharmaceutical supply chain management system. This system aims to ensure that the regulation enforced by the health authority are met during the trade supply chain. This system is composed of three layers, front-end, backend and IoT layers. the backend layer consists of a blockchain and http server. The front-end layer consists of smart phones. The backend is connected to the front-end through the server. The sensor devices in this architecture are connected to the other layers through the server as well. The user connects to the sensor and start the temperature monitoring process via Bluetooth technology. Once these sensors are configured to start the monitoring process, they are placed inside the packages and the process of shipping starts. The user connects to the sensor devices and download the temperature reading when the shipment arrives to the destination.

Toyoda et al. (2017) employed blockchain to address the counterfeiting products issue in supply chain. In their solution, the authors implemented a full-fledged proof-of concept that qualifies all parties in the supply chain to transfer and prove the product's ownership. The product's ownership can be proven using the proposed Proof-of-Possession-of-Products concept using the unique product code that is written in the Radio Frequency

Identification (RFID). In addition, each party in this chain verifies the legitimateness of the RFID data and appends additional evidences into the RFID tags. This application is suitable for branded goods and not for cheap products due to the additional cost associated with it.

Alzahrani and Bulusu (2018) proposed the block-supply system, which aims to detect and avoid counterfeiting products in trade. The system is composed of two phases: initialization and verification. The initialization phase involves registering the product and the manufacturer on the Blockchain. Initialization is performed through the use of the public-private keys of the registered manufacturer. The registered manufacturer will be able to add all of his own products on the network. Throughout the life cycle, once a node receives an asset, two types of authentications are performed, local authentication and global authentication. Local authentication is used to verify that no manipulation occurred at the transaction level. Global authentication is used to verify the legitimacy of all participants in this transaction.

A pilot study by Kamath (2018) examined the idea of tree-to-shelf traceability, where fruits and vegetables are attached with sensors devices to monitor the environment. In addition, warehouses and trucks are equipped with IoT devices to monitor the stores conditions. The proposed solution provides the users with an interface to access the related historical data about the goods.

Mao et al. (2018) developed a food trading system based on blockchain technology. This system mainly consists of two components: user nodes and scheduling node. User node plays the role of seller or buyer. While, Scheduling node plays an authority role and is used to verify the transaction and perform other related administrative tasks. For

example, scheduling nodes contains a record of all transactions occurred in the system. In this system, consensus is performed through the use of improved version of Practical Byzantine Fault Tolerance (PBFT). This aims to increase the efficiency of the consensus method and reduce the used computational power. Similar solutions are usually used for public blockchain to ensure the integrity of the data.

#### *2.2.2.3. Ecommerce*

Ecommerce has revolutionized the way in which trade is performed between traders across the world. Ecommerce trade can be mainly categorized into Consumer-to-Consumer (C2C), B2B, B2C and Business-to-Government (B2G), where the former two types are bi-directional based on the flow of goods (Yin et al. 2018). The introduction of these e-commerce trading types has been established to capture various trading domain behaviors. For instance, B2B is expected to deal with large volume of shipments between companies. While, the trade volume in B2C is less and its items are normally bought for personal use.

These trading types facilitate the trade and business between parties through trading goods or services. For instance, C2C connects private individuals, B2B connects various businesses to other businesses rather than to private individuals (consumers), B2C connects businesses directly to the consumers, and B2G connects businesses to government agencies. Each of the trading types has in nature different characteristics and challenges. For instance, trading in C2C fashion raises traceability and security challenges to ensure the identity of the participating partners (Yin et al. 2018), whereas dealing with trusted entities such as government organizations and well-known companies (B2G and B2B) removes the security challenges and highlights the requirements of creating scalable architecture. In the government sector, the participating employees' identities in the trade

are known and pre-defined. The main requirement in this scenario is to ease the process of conducting trade. For example, multiple employees in a government entity have to agree on the terms and conditions of a particular trade. Thus, the business architecture must support the complex communication between various personnel representing the buyer. The scalability component is also important when the buyers are known and are represented as a business. Having participants in e-commerce domain with different acting roles denotes the necessity to consider their needs and requirements during the design of any e-commerce solution. These needs and requirements change dramatically according to the type of participant, whether it is an individual customer or an established business. Therefore, the value proposition of the e-commerce platform must consider the participants behavior during the design of the platform services.

As part of any ecommerce platform, the product delivery or transportation between the seller and the buyer should be dealt with in a careful manner to make sure the conditions they agreed on are satisfied. For instance, Hasan and Salah (2018) proposed a blockchain solution for proof-of-delivery between the buyer and the seller using multiple transporters. In this paper, the authors identified the main components in this system to be: seller, buyer, courier service, arbitrator, and smart contract attestation authority. These entities perform their action on Ethereum platform to ensure the completion of the transaction. Seller acts the initiator of the deal since he advertises the product using smart contracts. Buyer acts as a recipient for the goods once the transaction is completed. Courier service handles the transportation of the goods. Similar to any trading process, it is required to have an authority entity that handles and makes sure that all parties are satisfied. In this system, this is handled by the arbitrator. Once the current online purchase transaction for a specific

smart contract is completed, the attestation authority confirms that the conditions of these tasks are satisfied and adds the verified smart contract to the blockchain. The actual spending is done using Ether which utilizes the benefits of Ethereum platform. The proposed system utilized the benefits of blockchain technology in terms of traceability and security. However, the arbitrator plays an administrative task and acts as a third party who monitors the deal.

In a study which was set out to determine the scalability and security challenges in blockchain, Chen et al. (2017) proposed Permissioned Blockchain Framework (PBF). This framework aims to ensure the trust and security of ecommerce trading platform. The authors suggested special and conservative hierarchy for the participating nodes and their role in the Blockchain. The participants are divided in this framework into two sets: peer set and entity set. The peer set represents the trading platform such as TMALL and EBAY, while the entity set represents the buyers and sellers. In addition, the peers in this framework are divided into three types: the construction peer, the validation peer, and the regular peer. Only the construction peers can generate blocks and validation peers are the responsible to validate these blocks. Peers normally write their transactions on *micro-blocks*. These blocks are then sent to the validation peers to verify their genuineness. Then, the validated blocks are prepared as *peer-blocks*, which is also sent for validation. Once these blocks are validated through the voting mechanism, the *key-blocks* are constructed and added to the Blockchain. This multi-level of validation mechanism aims to increase the efficiency and security of the presented platform.

In the same vein, Xie et al. (2018) also proposed trust framework for ecommerce through the use of Blockchain. This framework also uses the idea of having multi-level

participants with different authorities in order to increase the trust level between the participants.

Ying, Jia and Du (2018) demonstrated a use-case study that shows the benefits of using a blockchain ecommerce platform to offer employees flexible benefits. In this platform, instead of regarding the employees of gift cards, they are giving cryptocurrency which they can use over the ecommerce platform to buy a product. The proposed platform provides the employees with flexibility and may further benefit the company, because it might construct deals with the suppliers.

## 2.3. Outlier Detection Approaches

Blockchain technology ensures the integrity of the data which expands the horizon of developing several data mining techniques that captures the trading behaviours. Such techniques are expected to employ outlier detection mechanisms to detect any abnormality in the companies trading behaviours. The use of outlier detection mechanism to detect the abnormal activities have been studied extensively in various domains (Olson 2007).

For instance, in (Malini & Pushpa 2017; Ceronmani Sharmila et al. 2019) the authors investigated the problem of detecting credit card fraudulent activities. Accordingly, in this study, k-nearest neighbour (K-NN) approach has been adopted to ensure the legitimacy of the credit cards transactions and detect anomalies

To address the problem of detecting the money laundering related activities, a distance-based approach was developed to capture the likelihood of money laundering behaviours (Gao 2009). The proposed approach groups the transactions based on the distances between the points within the group. Then, the points are analysed based on pre-defined threshold

to determine legitimacy of these transactions. In the same vein, a study that addresses the financial behaviours were conducted (Chen, Wang & Chen 2007). In this study, the authors focused on Taiwanese companies and investigated the benefits of employing outlier detection approach such as Local Outlier Factor (LOF) and K-NN.

In the cars insurance domain, a K-NN approach has been also adopted to detect fraud activities by (Badriyah, Rahmaniah & Syarif 2018). The proposed approach represents the transactions in multi-dimensional space, where the analysis is performed based on the actual distances between the points. The new transaction will be investigated with its k-nearest neighbours to confirm the status of the transaction. The value of *k* is determined by performing a sensitivity analysis and the entire process eventually uses majority voting mechanism to determine whether the new transaction is legitimate of not.

In the computer network traffic domain, a mechanism to capture the abnormality in network traffics has been proposed by (Gan & Zhou 2018). This mechanism represents the network traffic as points in multi-dimensional space. Then, the points are analysed using clustering techniques to determine the similarity between these points.

In the transportation domain, to improve the bus root planning in the transportation domain, two studies experimented several outlier detection algorithms. These algorithms attempt to determine the bus locations that must be removed from the root in order to optimize the operational cost. Accordingly, the proposed algorithms use the entire city map to rank the bus stops based on their locality. This ranking takes into consideration the number of available buses and the maximum allowed travelling time to each bus. Thus, the proposed algorithms aim to remove the bus locations such that the new bus roots cover the

entire city within the pre-defined travelling deadline for each bus (Almiani, Chawla & Viglas 2014; Almiani et al. 2018).

To optimize the distance-based outlier detection converging time, a multi-core clustering algorithm was proposed by (Bhaduri, Matthews & Giannella 2011). This algorithm uses pruning method to determine the outlier factor value for a given point based on pre-defined threshold value. In this algorithm, any point that reaches the pre-defined threshold will not be considered. A hierarchal sorting technique was employed to update the threshold value. Additionally, the proposed algorithm adopted the concept of leadership points that is used to monitor the transactions set and the process of determining the likelihood at each point being an outlier.

## 2.4. Discussion

This section presented a discussion of the findings from the literature review conducted. As such, proposals and arguments from the literatures investigated and reviewed about the application of blockchain in improving global trade supply chain and ecommerce are presented and discussed. Therefore, this section discusses the findings of this review.

Blockchain applications demonstrate a promising opportunity for improving processes and practices in the global trade supply chain and ecommerce fields. Interestingly, the reviewed researches revealed that the application of blockchain solutions and technologies play an important role in the advancement of trade supply chain and ecommerce. This means that there are various fields of applications of blockchain that can contribute to providing cost effective solutions for different types of industries (Petersson and Baur,

2018). Therefore, it is important to highlight the role of blockchain in application to trade supply chain and ecommerce.

In the general perspective, supply chains are facing three fundamental challenges in the modern era; data visibility, process optimization and demand management (Godbole n.d.). In order to address these inefficiencies, various emerging technologies are being adopted into supply chain management systems. In particular, the application of blockchain is growing in the area of global trade supply chain. According to Belu (2019), the application of blockchain technology in the field of international trade can contribute to reducing costs, providing more transparency and security and easier control and flow of goods among others. This suggests that the application of blockchain in trade supply chain can improve global trade supply chain by promoting immutability, trust and transparency thereby transforming the international trade sector.

Blockchain in international trade can be applied in supply chain and logistics in order to improve global trade practices. Blockchain application is a new way of organizing and controlling trade value chain and import and export operations (Belu 2019). In addition, blockchain technology also contributes to improving the global trade supply chain by making it more transparent. According to Norberg (2019, p. 4), “Blockchain will decrease the costs of trade, which will empower globalization, trade, and optimize the global value chains that the ICT revolution has made possible”. As explained by the same author, the application of blockchain in trade supply chain allows for improving agreements that are already in place and supporting the underlying characteristics of modern trade agreements. Through the adoption of Blockchain technology in the global trade supply chain enables new levels of trust, transparency and accountability towards revolutionizing and

reinventing international trade. Similarly, Cole, Stevenson and Aitken (2019) asserted that the adoption of blockchain technology in supply chain management perspective presents opportunities for enhancing product safety and security, improving quality management, reducing counterfeiting and improving the sustainability of supply chain management. This suggests that embracing blockchain in supply chain management can help transform the global nature of supply chains.

In ecommerce, blockchain technology presents promising opportunities towards improving scalability and addressing performance issues (Lim et al. 2019). Findings from this literature review revealed that the most popular Blockchain technology being adopted in the field of ecommerce include Bitcoin, Ethereum and Cryptography among others. As the scalability of ecommerce continues to increase, blockchain technology contributes to optimizing business processes, reducing operating costs and improving efficiency of collaboration and interaction (Zhu & Wang 2019). This suggests that the application of blockchain technologies can contribute to improving ecommerce.

The findings from the review also revealed that there are a number of applications of blockchain for ecommerce in different business scenarios. For example, blockchain network technique is being integrated in some ecommerce models in order to enhance consumer to consumer experience (Shorman, Allaymounq & Hamid 2019). This means that Blockchain technology can contribute to enhancing customers' ecommerce experience through the establishment of trust and credibility. According to Sheikh *et al.* (2019), blockchain technology provides support to ecommerce platforms to store highly secured data and provided high performance transactional processing. This is a major concern in immutable records for web payments and online processing for orders.

This means that one of the major advantages of blockchain application in ecommerce is promoting secure transactions particularly during online payments and purchases. As such, blockchain technology presents opportunities for revolutionizing the traditional ecommerce industry.

## 2.5. Conclusion

This review provided important insights into the solutions that are designed blockchain environments should deal with various challenges. The use of the blockchain technology is mainly influenced by the application domain. Ecommerce platforms are for public users, which represents a challenge on ensuring the effectiveness and efficiency of the architecture design. One of the major issues in such platforms is the consensus method. Various researches that have employed the well-known proof-of-work concept aim to guarantee the data security and integrity. The reason is the high intense computational mathematical puzzle is introduced to guarantee the data integrity. However, this method limits the scalability of the proposed approaches since it requires a high computational power which negatively impacts the performance. In this context, any ecommerce platform should create a balance between data security and scalability. This is challenging since improving one of them reduces the efficiency of the other.

In addition, some functionalities could be located outside the blockchain due to the complex business processes (Xu et al. 2016; Lu & Xu 2017). Thus, the architecture of the ecommerce platform should adopt loose coupling to avoid hindering performance of the chain and not to perceive blockchain as a one-stop solution where all functionalities reside. Furthermore, any architecture should also contain an administrative node playing the regulator role of the network in terms of registering users and defining activities. The

security for this node should be ensured to increase the trust level between the customers and the platform.

Lu and Xu (2017) presented origin-chain, which can be seen as a platform to ensure the traceability of the trading process. The authors integrated origin-chain functionality with several ecommerce platforms such as jd.com. In their platform, the traceability of the goods flow was addressed. However, there was no presence of any customs processes or reputation mechanism. Additionally, several works have integrated their Blockchain-based trading solutions with ecommerce platforms (Chen et al. 2017; Hasan & Salah 2018; Xie et al. 2018; Ying, Jia & Du 2018). However, these works focused on traceability issues and did not explore the benefits of having a single digital channel that monitors all of the ecommerce activities. Moreover, the customs processes were not targeted as part of their work. In this thesis, the aim is to build customs enabled blockchain-based framework to enhance trade facilitation and compliance in ecommerce domain as discussed in chapter 4.

The use of the blockchain technology will ensure the integrity of the exchanged information between the ecommerce trading parties. Therefore, the customs declaration application can be automatically generated by pulling the relevant information from the blockchain. Thus, this framework allows the shipping agent or the buyer to trigger the automatic generation of the customs declaration once the shipment leaves the exporter counter. This will significantly simplify the customs declaration process as well as contribute toward optimizing the trading experience.

In terms of abnormality detection, the discussed proposals apply distance-based or density-based measure to analyze the similarity between the points. In addition, these

proposals are not designed mainly based on the application scenario itself. The context of the data (application scenario) can help further in determining the actual similarity level between the points. Accordingly, this thesis presents hybrid approach that takes all of the mentioned points into consideration. It proposes multilevel outlier detection mechanisms that can be used to detect any abnormality in the trading behaviours. Compared to the rest of the literature, this study applies a divide and concur approach that helps in simplifying the problem by dividing the available data into groups based on its characteristics. As it will be discussed in chapter 5 and chapter 6, this reduction eventually results in establishing an adaptive outlier detection approach.

# 3. Chapter 3: Research Methodology

Cross-border trade is a critical vehicle for economic stability and revenue generation. This work seeks to streamline ecommerce trading processes using blockchain technology while ensuring the needs of all concerned parties. This research first aims to determine the factors that drive the adoption of Blockchain technology. To this aim, a literature review is thoroughly conducted to investigate the factors that influence the adoption of blockchain technology in various domains such as the trade supply chain, energy trade and food safety in general as well as in ecommerce domain in particular.

The second aim is to build customs enabled blockchain-based framework to enhance trade facilitation and increase compliance while eliminating risks in ecommerce domain. To achieve this objective, Ethereum platform is used to host the on-chain functionalities of this framework, where smart contracts are implemented using Solidity programming language. Ethereum is recognized as a public Blockchain where any user can interact with the network. The Software Development Lifecycle (SDLC) methodology is followed to build this framework. An integrated web application is developed as a proof-of-concept to mock up the end-to-end process in ecommerce. Then, a preliminary evaluation on the integrated web application is performed to verify the usability of the framework. In this phase, the system is demonstrated to three users' categories to assess the usability of the framework using the System Usability Scale (SUS). The data resulting from this framework will provide advance access to information from various sources that can be used as inputs in constructing the off-chain components.

To address the need to determine value manipulation in ecommerce transactions and increase the efficiency of Customs Audit process, two off-chain clustering algorithms are proposed using the obtained dataset from Dubai Customs. To this end, the CRISP-DM methodology is employed for modelling these algorithms. Accuracy and precision measures are used to evaluate the performance of the proposed clustering algorithms.

In this chapter, the methodology followed in this thesis is illustrated to answer the following research questions:

1. What factors will influence the adoption of blockchain technology in ecommerce platforms?
2. How can blockchain enhance trade facilitation and compliance in ecommerce trade domain?
3. Can clustering techniques be used to determine risks related to value manipulation in the ecommerce transactions through by utilizing the proposed blockchain-based framework?
4. Can clustering techniques be used to identify risks in shipments for Customs post clearance audit by utilizing the proposed blockchain-based framework?

## 3.1. Research Approach

In this work, two different development methodologies are adopted: The Software Development Life Cycle (SDLC) for establishing the framework foundation as well as building the application that will be used for data capture, and The Cross Industry Standard Process for Data Mining (CRISP-DM) to mine the data post capture to detect anomalies. For the SDLC, the System Usability Scale (SUS) is used to evaluate the proposed system

usability. Whereas for CRISP-DM, the accuracy metric is mainly used to evaluate the performance of the proposed clustering algorithms.

In the following sections, the approaches to establish the foundation of the proposed framework as a distributed ledger system and to build the anomalies detection algorithms are discussed.

### 3.1.1. The Software Development Lifecycle (SDLC)

The proposed framework is founded on the principles and conventions of the World Customs Organization (WCO) to enable a standardized approach and deliver a scalable framework that drives faster adoption supported by a solid and flexible backbone using the SDLC methodology.

Therefore, as a proof-of-concept, a web application is developed using the SDLC methodology. During the SDLC, the creation of an artefact provides the researcher or software developer with feedback that adds to the main body of knowledge (Munassar, Ali & Govardhan 2010). The SLDC comprises various distinct stages. Each stage has its own deliverables, time frame and weight according to the requirements. The main stages of the SDLC include planning, requirement analysis, designing, implementing, testing, and maintenance, as shown in Figure 1.

The SDLC process can be represented through various abstract models such as the Waterfall model, Iterative model, V-shaped model, Spiral model, and Extreme model (Munassar, Ali & Govardhan 2010). Each model has its own advantages and disadvantages. For the development of the blockchain framework, the Iterative model was adopted. The Iterative model will allow flexibility in members joining the network

whenever they are fully ready to come on board without disrupting existing iterations or implementations. In the Iterative model, each iteration results in a prototype with its own specifications related to the producer of the supplied information. Each iteration is considered a mini SDLC, with the feedback of each iteration informing the next one until the final product is delivered. The iterative model allows the developer or researcher to obtain valuable feedback and results in the early stages of building the overall solution (Ruparelia 2010).

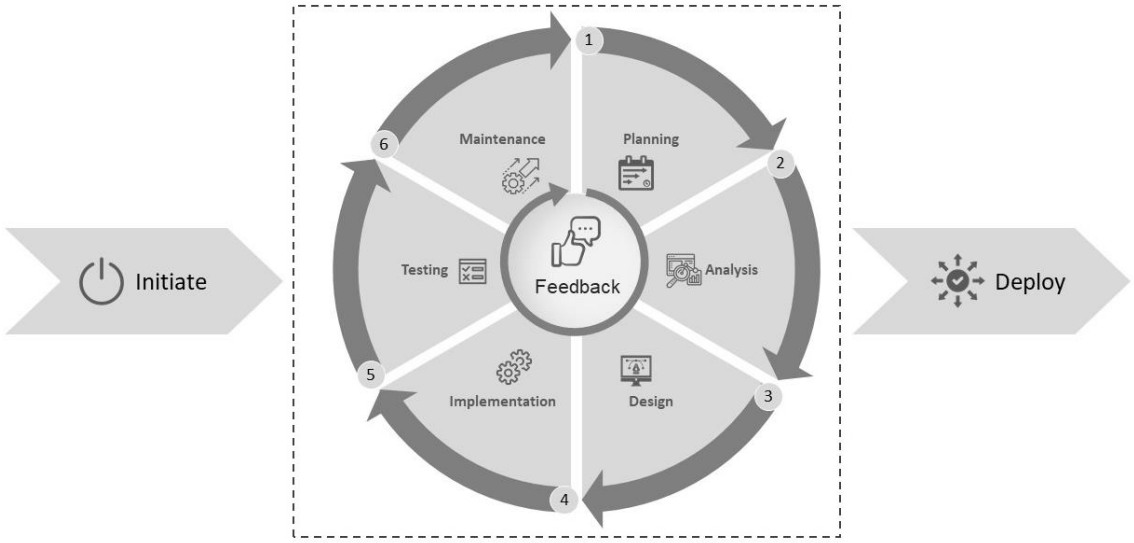


Figure 1: The Iterative Development Model

### 3.1.2. The Cross Industry Standard Process for Data Mining (CRISP-DM)

The second phase of the proposed solution is to develop detection methods indicating fraudulent or abnormal behaviours. The methods will help cover the risk classification and targeted auditing activities using the CRISP-DM methodology.

The SDLC is a good match for the software development and implementation since the scope, objectives and requirements are clearly defined with tangible deliverables (Saltz 2015). However, this methodology is not a good fit for the data science experiments due to

their broader scope, which ultimately aim to gain insights through prediction, segmentation, pattern recognition, information extraction, and clustering among other data mining tasks in various application domains.

There are various well-known methodologies for data mining such as Knowledge Discovery in Databases (KDD) (Fayyad, Piatetsky-Shapiro & Smyth 1996; Siyam, Alqaryouti & Abdallah 2020; Alkashri et al. 2021), SEMMA (Azevedo & Santos 2008b), and CRISP-DM (Pete et al. 2000; Altaheri & Shaalan 2020). Figure 2 illustrates these commonly used data mining methodologies and provides a comparison between them. KDD and SEMMA are similar in terms of structure and differ in terms if cyclic nature. In KDD, the cycle can be repeated triggering the corresponding stage relevant to the updated targets according to the evaluation results when all stages are completed. On the other hand, SEMMA has a cyclic nature that starts with “Sample”, “Explore”, “Modify”, “Model” and ends with “Assess” (Olson & Delen 2008).

The CRISP-DM methodology combines “Selection” and “Preprocessing” stages in KDD as well as “Sample” and “Explore” stages in SEMMA into one stage called “Understand Data”. CRISP-DM also includes two additional stages, namely “Understand Business” as a first stage and “Deployment” stage as a last stage, which make the development process, resembles a real-life solution development process. Furthermore, the reversal between stages is one of the advantages in CRISP-DM, which helps in working with real data.

These methodologies are used for data instead of software and provides more flexibility than the SDLC models. Furthermore, these methodologies share the objective to uncover

hidden knowledge and gain insights through data mining algorithms. In addition, these methodologies help the researchers in the analysis process of the available data.

In this work, CRISP-DM is adopted as it is industry neutral as well as technology and application neutral. It is an extremely popular data mining methodology that is considered a data mining *de facto* standard for developing knowledge discovery projects (Mariscal, Marbán & Fernández 2010).

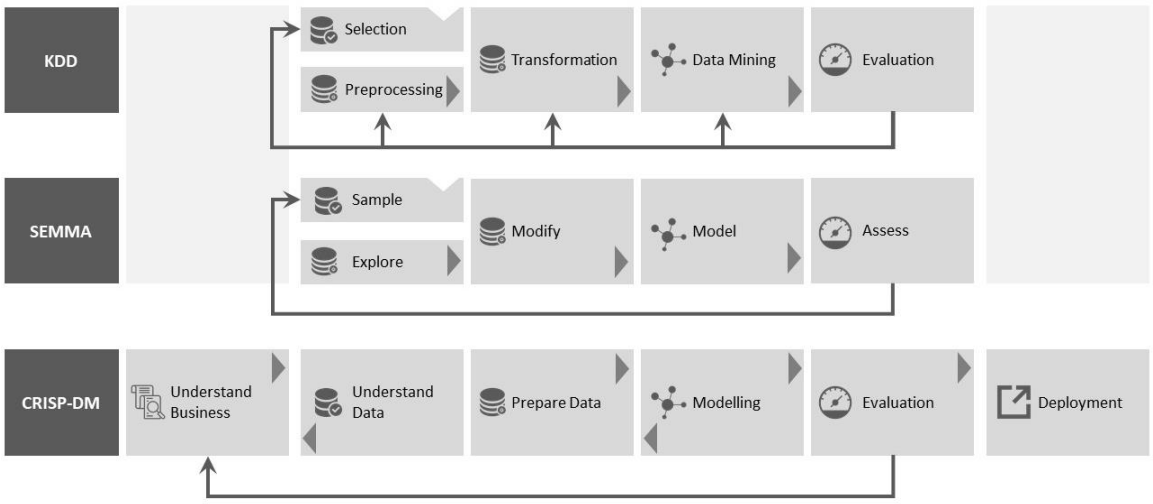


Figure 2: Comparison between Data Mining Methodologies

As illustrated in Figure 2, the CRISP-DM methodology comprises 6 stages, namely Understand Business, Understand Data, Prepare Data, Build Data Mining Model, Evaluate Performance and Deploy the Model. The first stage requires having clear understanding of current and desired business objectives along with gaps, challenges, and opportunities. The second stage deals with comprehending the available data in terms of shape, state, and quality. The third stage goes into data synthesis to provide a cleansed and harmonized dataset. The fourth stage takes the formatted data and begins the modelling exercise, which

is evaluated in the fifth stage to determine the most significant model. Finally, in stage six, the model is packaged for deployment to production to handle real-time data.

The modelling stage is concerned about building the Data Mining model. In this study, clustering approaches were adopted. The focus in this study is to address real-life scenarios related to customs post clearance audit and value manipulation. The obtained dataset is divided into 90% safe customs declarations and 10% risky customs declarations. This creates a challenge to reduce the learning bias through controlling the outliers and class imbalance. Traditional classification algorithms are biased toward the majority class. This may lead to degrade the classifier performance. According to the literature, using clustering techniques proved its efficiency in various domains (Olson 2007). Thus, this study adopts clustering techniques as a promising staring point, since converting the problem into geometrical space is expected to reduce the problem complexity and avoid bias. In our problem, the presence of the HS-Code hierarchy underlines the expected benefit of using clustering since the presence of such controlling parameters already divided the input data into isolated subspaces. By transforming it in a way where the relationship between the points is represented by using Euclidian distance and this simplifies the complexity of the problem and gives the ability on how to control performance and achieve the desired outcomes.

## 3.2. Data Collection

To evaluate the usability of the proposed framework, data is gathered through questionnaires. Questionnaires were administered after the users have the chance to interact with the web application during a workshop. The workshop aimed to present the participants with an overview of the proposed system and demonstrate how it works.

Several tasks were given to the users according to the type of participant. Participants were given the SUS questionnaire to evaluate the usability of the proposed framework and the level of satisfaction after completing the tasks that were assigned to them during the workshop. The SUS was developed by Brooke as a tool to measure the perceived usability of a system (Brooke 1996). Questionnaires are considered a method that is commonly used in data collection due to its efficiency in terms of cost and time. Furthermore, this method is free of bias of the researcher as it does not have any influence by the researcher. Thus, it produces reliable and dependable results (Kothari 2004).

To evaluate the performance of the proposed data mining clustering algorithms, data is obtained from Dubai Customs; a government entity in Dubai. This dataset was prepared for shipments conducted in the year of 2018 and includes 500,000 records. The data was masked to hide critical information related to shipments and customers. This is ensured to maintain privacy by preserving sensitive and personal data related to consumers, businesses and authorities. The records were labelled to identify the risky company declarations. The resulting dataset is used as the input data to address the third research question. This dataset was further prepared by the intelligence department in Dubai Customs to identify those risky shipments that include value manipulation, which is used as input data to address the fourth research question. Table 1 provides the details related to the data collection methods and the aims and research questions that they address.

Table 1: Data Collection Methods to Address Research Objectives and Questions

| # | Objective | Research Question | Data Collection Method | Chapter |
|---|---|---|---|---|
| 1 | Determine factors that influence the adoption of blockchain technology in ecommerce platforms | What factors will influence the adoption of blockchain technology in ecommerce platforms? | Literature Review | 2 |
| 2 | Build customs enabled blockchain-based framework to enhance trade facilitation and compliance in ecommerce domain. | How can blockchain enhance trade facilitation and compliance in ecommerce trade domain? | Questionnaire | 4 |
| 3 | Develop a clustering technique that utilizes the proposed blockchain-based framework to identify risks in shipments for Customs post clearance audit by utilizing the proposed blockchain-based framework | Can clustering techniques be used to identify risks in shipments for Customs post clearance audit by utilizing the proposed blockchain-based framework? | Dataset Obtained from Dubai Customs | 5 |
| 4 | Develop a clustering technique to determine risks related to value manipulation in the ecommerce transactions by utilizing the proposed blockchain-based framework | Can clustering techniques be used to determine risks related to value manipulation in the ecommerce transactions through by utilizing the proposed blockchain-based framework? | Dataset Obtained from Dubai Customs | 6 |

## 3.3. Data Analysis

### 3.3.1. System Usability

The evaluation of blockchain systems depends on its objective. For instance, various researches aim to measure the communication delay through the network latency (Xu et al.

2019). Other researches aim to measure the acceptance level, usefulness, usability, and satisfaction (Heimgartner 2017; Ramadhan & Iqbal 2018).

The evaluation of the proposed framework is mainly concerned about the usability of the Customs-based Trade Facilitation website according to the buyer, shipping agent and customs administration perspectives. Therefore, a web-based application that is connected to the blockchain trade facilitation framework for e-commerce platforms has been built to simulate the end-to-end process including e-ecommerce platforms. To evaluate the blockchain framework, usability tests were used by allowing stakeholders to interact directly with the system.

According to ISO 9241– 11 (Bevan & Carter 2016, p. 269), usability is defined as "the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". Therefore, this study adopts the three main aspects of usability according to ISO 9241– 11; effectiveness, efficiency and satisfaction. Effectiveness is defined as the ability of users to achieve the intended goal using the system while considering the needed degree of precision of the output and the avoidance of errors and risks. Efficiency is defined as the resources needed to achieve a specific task, such as time and mental effort. Lastly, satisfaction is defined as users' attitudes towards the use of the system (Bevan, Carter & Harker 2015).

Usability can be evaluated through objective measures, such as data logs, or through subjective measures, such as questionnaires. Both measures are considered essential for an accurate evaluation outcome (Bevan & Carter 2016). Different methods can be used to

evaluate usability depending on the objective of the usability study. Methods of usability evaluation include cognitive walkthroughs, heuristic evaluations, comparison against guidelines, and usability tests. Effectiveness and efficiency can be easily measured through objective measures such as error logs and time on task. On the other hand, satisfaction is considered a subjective measure of usability that can be evaluated through think aloud sessions, focus groups, interviews, or questionnaires. Different tools were developed to evaluate users' satisfaction with a system. For instance, the Technology Acceptance Model (TAM) (Davis, Bagozzi & Warshaw 1989) considered perceived ease of use and perceived usefulness as the two aspects impacting users' satisfaction, and therefore their intention to use the app.

In this study, a workshop is conducted where users interact directly with the system. Then, the System Usability Scale (SUS) (Brooke 1996) is conducted to measure the framework usability in general, with a focus on users' satisfaction.

The SUS questionnaire was developed as a "quick and dirty" measure of the perceived usability of a system (Brooke 1996). However, as a result of numerous studies conducted, SUS has become an industry standard due to its ease of administration to participants, the ability to be used on a small sample size (as small as twelve participants), and its validity and reliability as a measure that distinguishes between usable and unusable systems (Tullis & Stetson 2004; Bangor, Kortum & Miller 2008; Lewis & Sauro 2009b).

The SUS questionnaire comprises 10 questions with 5-point Likert scale to assess the agreement level with the SUS questions. The rate for each question varies from 1 for "Strongly Disagree" to 5 for "Strongly Agree". The questions are divided into 2 groups:

the odd numbered questions and the even numbered questions. The odd numbered questions (1,3,5,7 and 9) represent the positive questions. Whereas, the even numbered questions (2,4,6,8 and 10) represent negative questions. The purpose of this approach is to avoid bias in the participants answers (Lewis 2018).

The standard questions of the SUS proposed by (Brooke 1996) have been revised to reflect the purpose of the proposed solution as illustrated in Appendix A: System Usability Scale. The term "this/the system" from the original paper by Brooke (1996) was replaced with term "Customs-based Trade Facilitation website". Moreover, the word "cumbersome" in item 8 was replaced with the word "awkward" as suggested for non-English speakers (Finstad 2006; Lewis & Sauro 2009b).

The researcher was available during the workshop to assist participants while performing the tasks. The SUS questionnaire was administered electronically to the participants after they had the chance to interact with the system during the workshop. It is recommended to administer the SUS questionnaire directly after the participants interacted with the system and before any discussion takes place (Lewis & Sauro 2009b). Participants were instructed that all items should be answered. However, if the participant felt he cannot respond to a certain item, the centre point of the Likert scale shall be marked. Participants were also instructed to mark down their answer immediately rather than thinking about a question for a long time (Brooke 1996).

Responses were transferred to a spreadsheet for analysis. The overall score of the SUS is calculated in 2 stages. The first stage includes normalizing the user inputs. In case of odd numbered questions, the new value results from subtracting 1 from the user's selection

value (USV) (USV-1). In case of even numbered questions, the new value results from subtracting the user's selection value from 5 (5-USV). The user's final result represents the total sum of all new values multiplied by 2.5. And the final result of SUS represents the average of all user's final results, which represent the value of the system usability (Brooke 1996).

The SUS score ranges from 0 to 100, with 0 considered the absolute worst and 100 considered the absolute best. However, the SUS can be hard to be interpreted by itself as it is not a percentage. Thus, normative data collected through numerous studies situated the SUS scores as percentiles, which allowed for a more meaningful way to interpret SUS scores (Lewis & Sauro 2009b; Brooke 2013). The average score for SUS was found to be 68, representing the 5th percentile (Brooke 2013).

Another way to judge the SUS is the "university grade analogue" (Bangor, Kortum & Miller 2008). According to this method, grades range from A for high performing systems, to F for failing systems. While not scientifically validated, this grading method can be distributed to match the normalization of percentiles, where C is considered the average value.

Moreover, SUS scores can be described as "acceptable" or "not acceptable". For instance, scores above 68 (the average) are considered acceptable, scores below 50 are considered unacceptable, and scores between 50 and 70 are considered marginally acceptable (Bangor, Kortum & Miller 2008). Finally, Bangor, Kortum and Miller (2008) related 212 scores with an additional 7-point adjective rating question that asked participant to rate the user-friendliness of the system by choosing from the options "Worst

Imaginable", "Awful", "Poor", "Ok", "Good", "Excellent", and "Best Imaginable". For instance, it was found that the rate "Excellent" was associated with scores above 85, while the rate "Good" was associated with scores above 72.75. Table 2 and Figure 3 summarize the ratings proposed to describe the SUS scores (Bangor, Kortum & Miller 2008; Sauro 2018).

Table 2: Proposed Rating Correlating with SUS Scores (Bangor, Kortum & Miller 2008; Sauro 2018)

| Score | Percentile | Grades | Adjective | Acceptable |
|---|---|---|---|---|
| 84.1 – 100 | 96-100 | A | Best Imaginable | Acceptable |
| 80.8 – 84.0 | 90 – 95 | B | Excellent | Acceptable |
| 78.9 – 80.7 | 85 – 89 | B | Good | Acceptable |
| 77.2 – 78.8 | 80 – 84 | C | Good | Acceptable |
| 74.1 – 77.1 | 70 – 79 | C | Good | Acceptable |
| 72.6 – 74.0 | 65 – 69 | C | Good | Acceptable |
| 71.1 – 72.5 | 60 – 64 | C | Good | Acceptable |
| 65.0 – 71.0 | 41 – 59 | D | Ok | Marginal High |
| 62.7 – 64.9 | 35 – 40 | D | Ok | Marginal Low |
| 51.7 – 62.6 | 15 – 34 | F | Ok | Marginal Low |
| 25.1 – 51.6 | 2 – 14 | F | Poor | Not Acceptable |
| 0 – 25 | 0 – 1.9 | F | Worst Imaginable | Not Acceptable |

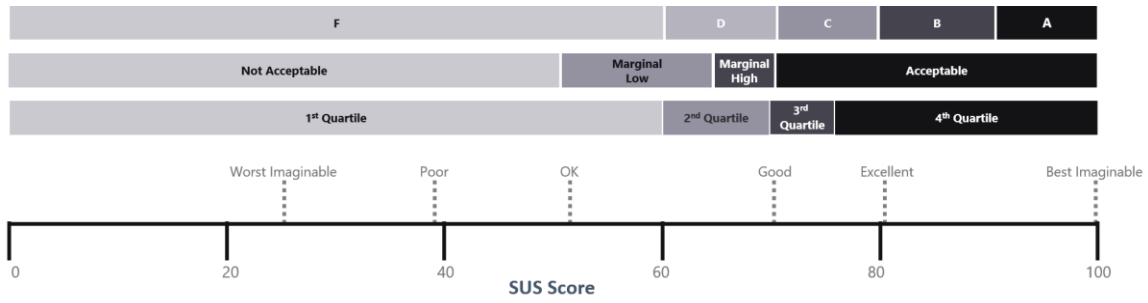


Figure 3: SUS Score and Rating (Bangor et al. 2009)

### 3.3.2. Algorithm Performance Evaluation

To evaluate the performance for each experiment settings, the accuracy and precision are considered. While accuracy refers to the correctly classified labels to the overall classifications, precision refers to the ratio of the correctly classified labels with certain value to the overall classification with this particular label value (Abualigah, Khader & Hanandeh 2018; Wu, Yang & Shen 2021). Accuracy measure is used to evaluate the performance of the proposed data mining algorithms that address the third and fourth questions. In addition, precision measure is also used as an additional measure to evaluate the performance of the proposed algorithm to answer the fourth question. To calculate these metrics, the following variables needs to be captured:

- **True Positives (TP)**: this variable denotes the correctly predicted shipments that have risk.
- **True Negatives (TN)**: this variable denotes the correctly predicted safe shipments not containing risks.
- **False Positives (FP)**: this variable denotes the falsely predicted shipments with risk.
- **False Negatives (FN)**: this variable denotes the falsely predicted shipments with no risk.

Accordingly, the performance evaluation metrics are calculated as follows:

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

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

## 3.4. Trustworthiness

Validity, reliability and objectivity are approaches used to establish trustworthiness of the data in quantitative research (Pitney 2004). This study adopts the SUS, which is a questionnaire that proved to be valid and reliable for different domains for sample sizes as small as 12 participants (Bangor, Kortum & Miller 2008; Lewis & Sauro 2009a). Moreover, objectivity is established by ensuring the researcher does not influence the results of the study with his personality, beliefs or values (Payne & Payne 2004). Therefore, questionnaires were completed anonymously to avoid influencing the participants opinion.

Finally, the obtained dataset includes customs declarations shipments information in standard format according to WCO ensuring that the proposed data mining clustering algorithms are applicable on other datasets following the same industry standard.

## 3.5. Scope of the Current Study

In this study, the aim is to eliminate cross-border ecommerce barriers to reach competitive advantage. The study is concerned with ecommerce mainly from a customs perspective. It addresses challenges relative to trade process communication, connectivity, operational risk and governance. The desire is to reduce reliance on the ineffective paper-based human driven activities and formalities while ensuring transparency and consensus among stakeholders. The selection and adoption of advanced technologies and methods stemmed from the need to establish a holistic single source of truth for cross-border trade activities. The proposed framework will greatly rid the trade supply chain from unnecessary fraud infested activities that lead to major financial losses for countries and organizations.

## 3.6. Role of the Researcher

To achieve the desired objectives of this study, a quantitative approach was adopted. The researcher started by making a formal request from the concerned business units to supply data relative to goods specifications, value and movement in accordance with the standard international customs declaration form. This data was supplied in raw format to avoid any forms of bias. With raw data in hand, the researcher performed some data preparation activities which may provide indication that fraud has occurred.

Then, a custom-enabled blockchain framework to enhance trade facilitation and compliance in ecommerce domain was developed. To test the usability of this framework, a web application was developed to mimic the ecommerce platforms and the interactions among participants. Then, a workshop was conducted to familiarize the users with the proposed solution. Therefore, the researcher conducted an overview of the proposed solution to the participants followed by interactive discussions. After that, a live demonstration of the application was conducted to illustrate the end-to-end process with all features. Participants were given the SUS questionnaire to provide their feedback on their experience after completing the workshop tasks. This was completed anonymously without the interference of the researcher to avoid influencing the participants' opinion.

After that, data mining techniques were developed to identify any value manipulation or fraudulent activities based on the available historical data. This is evaluated using scientific methods which uses datasets with random records for testing the proposed models for the measurement of performance metrics to eliminate any forms of data bias.

## 3.7. Ethical Considerations

During the development of this framework, it was required to collect the data for the analysis and study of the related ecommerce transactions from various sources. To gain access to these sources, a formal request was sent by the university to local customs authority. Maintaining the privacy was ensured by preserving sensitive and personal data related to consumers, businesses and authorities.

As part of the semi-structured interviews and surveys administration, the participants are asked to provide a consent for their participation (see Appendix B: Informed Consent Form). Participants information will be kept confidentially to ensure their anonymity (Creswell 2007).

## 3.8. Conclusion

This chapter illustrated the methodologies that are adopted in this work to manage the research process in a formal industry standard basis. The reliance on these methodologies helps eliminate gaps in the research and raise the acceptance level of findings. Constant engagement of stakeholders at every stage of the process ensured staying on track with the objectives of this research. The following chapters dive into the selected blockchain and data mining technologies, which will constitute the framework that should answer the research questions of this study.

# 4. Chapter 4: Trade Facilitation Framework for Ecommerce Platforms using Blockchain

The continuous growth in cross-border trade through electronic commerce (E-Commerce) has produced great opportunities for the global economy. The high volume of trade and the non-regular patterns in ecommerce trade suggest that the trade supply chain must be re-structured to ensure the continuous development of this domain. Based on the participating entities, ecommerce trade can be categorized into Consumer-to-Consumer (C2C), B2B, B2C and Business-to-Government (B2G) where the former two types are bi-directional based on the flow of goods. Each of these types has its own challenges and demands (Raadt & Xu 2021).

For instance, trading in C2C fashion raises traceability and security challenges to ensure the identity of the participating partners (Yin et al. 2018). Whereas, dealing with government entities removes the security challenges and highlights the requirements of creating scalable architecture. For example, it is expected that multiple employees in a government entity must agree on the terms and conditions of a trade before shipment. The scalability component is also important when the buyer is represented as a business. Having businesses as participants ease the identification process. The identity of the participating entity can be easily confirmed based on the business size and history.

One of the key stakeholders in this trading ecosystem is the customs administrations. The role of customs administrations is vital in society protection, legitimate trade facilitation and economic growth sustainability (Wulf & Sokol 2005). Customs administrations are expected to act as validators to a high number of expected transactions

in a daily basis, increasing the pressure on customs operations and reducing the trader satisfaction.

Furthermore, one of the customs responsibilities is the revenue collection (WCO 2018a). Duties are collected during the goods clearance process based on a pre-determined threshold. According to Chapter 3 from the Revised Kyoto Convention by World Customs Organization (WCO), the customer can request duty refund (also known as drawback) from the customs administration if the goods are to be re-exported or returned in a later stage (WCO 2008).

However, the duty drawback is not a straightforward process, since the shipment declaration that needs to be re-exported must match the imported shipment declaration. To address the presented challenges, any employed solution should be simple, allows traceability, and utilizes the advantages of having advance information as suggested by WCO cross-border ecommerce framework (WCO 2018a).

Blockchain can address the demanding requirements of simplifying the ecommerce trading domain and to ensure the traceability of the transactions. In this chapter, a blockchain-based conceptual framework is proposed. In this framework, the trade agreement and payment are performed at an ecommerce platform. Once the buyer and the seller negotiate and agree on the trading specifications, the customs administrations are consulted to perform an initial assessment. If the trade is confirmed by the trading parties, the trade details and the customs assessment are recorded in the blockchain. In this context, blockchain is used to trace the trade and to make sure that the customs requirements are fulfilled. This framework aims to increase the trust level between the buyer and the seller,

facilitate the trade by ensuring that the customs clearance is issued without any delay, and reduce the processing cost.

As discussed in chapter 2, more recent attention has focused on investigating the use of blockchain technology to improve the electronic trade supply chain. Furthermore, it provided a brief about blockchain technology aspects and discusses the related studies in the literature. These studies presented thus far focus on investigating the security and scalability challenges in ecommerce solutions. In this chapter, the focus is to optimize the process of ecommerce trade from customs perspective, and therefore the customs related processes are targeted.

The following section discusses the details of the proposed framework. Section 4.2 illustrates the details about the system evaluation. Finally, section 4.3 concludes the chapter and summarizes the key findings.

## 4.1. Customs-Based Ecommerce Framework

The presented framework in this work aims to simplify the overall trading process in line with the WCO ecommerce principles. In general, these principles target the use of advanced technologies to facilitate and streamline the trading process to fulfil customs responsibilities related to society protection and duty collection (WCO 2018a).

### 4.1.1. Overview

The proposed framework aims to emphasize the main components that can be used to improve the ecommerce trading business and comply with the desired WCO principles. Thus, this framework focuses on improving the traceability and simplicity of ecommerce trade. It also creates and establishes an adaptive structure that can be used to handle future

demands. The traceability feature is guaranteed through the use of the blockchain technology, where all transactions performed by the participating entities are recorded. In terms of simplicity, the framework supports layer-based architecture for the proposed functionalities. These functionalities are grouped into layers based on the provided services. Thus, future services or functionalities can be easily integrated into the framework. Additionally, the concept of duty drawback and e-wallet is used. Based on the agreement between the buyer and the customs administrations, the buyer might be allowed to re-export the imported goods within a predefined time deadline. In this case, the buyer is illegible to apply for duties drawback. This aims to reduce the pressure on the local traders and to support the e-hub concept. E-hub can be represented as a regional and global trade centre to import the goods from one location and re-export it to the rest of the world.

The proposed framework deals with two types of participants, namely traders and customs administrations. Traders refer to the seller and buyer entities, where the seller is a cross-border company. Customs administrations monitor the trade and ensures the fulfilment of the compliance of the government rules and regulations as well as the calculation of the required duties on the goods. These participants interact through the blockchain where all of their transactions are recorded and processed. The framework targets the B2B and B2C trading categories. B2B connects various businesses to other businesses in which trade is typically performed using two types of trading platforms, namely public and private. Using public ecommerce platform, businesses communicate and negotiate the deals with option to choose from various sellers. On the other hand, the private ecommerce platform is owned by the cross-border trader giving option to the buyers to access their platform to perform the trade. B2C connects businesses directly to the

consumers. The proposed platform can be expanded to cover other trade categories with slight modifications.

To perform a trade, each of the trading participants (buyer and seller) have to be registered on the blockchain. Then, each of these participants determine their role in this trade. In this context, once the seller, buyer and customs administrations agree on the trade specifications, the proposed system will be invoked to start the tracing mechanism. Moreover, the proposed framework attempts to simplify the process of customs declaration and ensures that all required data is available on the blockchain.

As it will be discussed in this section, this tracing mechanism ensures that the trade conditions imposed by all parties are satisfied.

### 4.1.2. Framework Architecture

As illustrated in Figure 4, the proposed architecture integrates the services of monitoring and tracing ecommerce trades into the seller's ecommerce platform. This framework deals with four types of participants, namely the customs administrations, the local trader (company or individual), the cross-border trader (seller), and the ecommerce platform.

Once the buyer and the seller agree on the trade conditions, the seller initiates the agreement registration and the trade tracing through the ecommerce platform. These services work to validate the compliance of the trade agreement with customs rules and regulations. The seller and the buyer nodes are notified if additional documents are required by the customs administration to complete the clearance process. During the trade process, all of the performed transactions are recorded on the blockchain. This includes the initial

trade agreement and the customs administration feedback. Therefore, once the goods arrive to the customs premises, they will be processed based on the original evaluation initiated by the customs administration.

The framework consists of four layers, namely ledger layer, enforcement layer, business logic layer, and service layer. In the ledger layer, the actual transactions are stored, represented as blocks. In the enforcement layer, all of the trade specifications and requirements are converted into smart contracts, which automate the trading processes. The monitoring process tracks the status of the trade agreements constraints and fulfilments. For instance, if the deadline threshold for submitting the drawback request is passed, the duty will not be refundable anymore. These smart contracts access the required information to perform their tasks from a local database. The presence of this database is important since the storage capabilities of the blockchain are limited (Muzammal, Qu & Nasrulin 2019). The enforcement layer also includes the authentication component, which makes sure that all transactions on the blockchain are legitimate such as validating and verifying if the seller or the sellers' country is within the permitted parties.

In the business logic layer, the customs' related processes are implemented to support the automatic clearance and duty calculation processes. This layer consists of three main functionalities; identify users, collect duties, and manage e-wallet. In the user identification process, the registration of trading parties is confirmed with the customs administration. Duty assessment involves the duty calculation based on the customs tariff codes for the commodities. The customs tariff code which is also known as the Harmonized System Code (HS-Code) is a product specific code maintained by the WCO. These codes are available for every product involved in the trade supply chain. The e-wallet process

manages the registered buyer's digital wallet. The e-wallet is used to handle the customs transactions in terms of customs duty payment or refund processes.

In the service layer, three main services are implemented; register agreement, track and update status, and request drawback. In the agreement registration, the actual trade conditions and trade parties are stored off-chain on the local database as a form of binding agreement. During this process, the agreements conditions are compared against the customs regulations to make sure that they comply with the customs rules and regulations. The trading parties are notified when further documents or information are required. In the status track and update service, all parties can use this service to determine the status of the trade and update any information related to the shipment. Finally, in the drawback request service, the buyer submits his interest for duties drawback due to re-exportation.

As illustrated in Figure 4, both the seller and the buyer interact with the framework through the agreement and tracking services. The buyer also interacts through the refund service. The framework achieves the desired objectives by simplifying the processes of registration, customs clearance, duties calculation, and drawback.

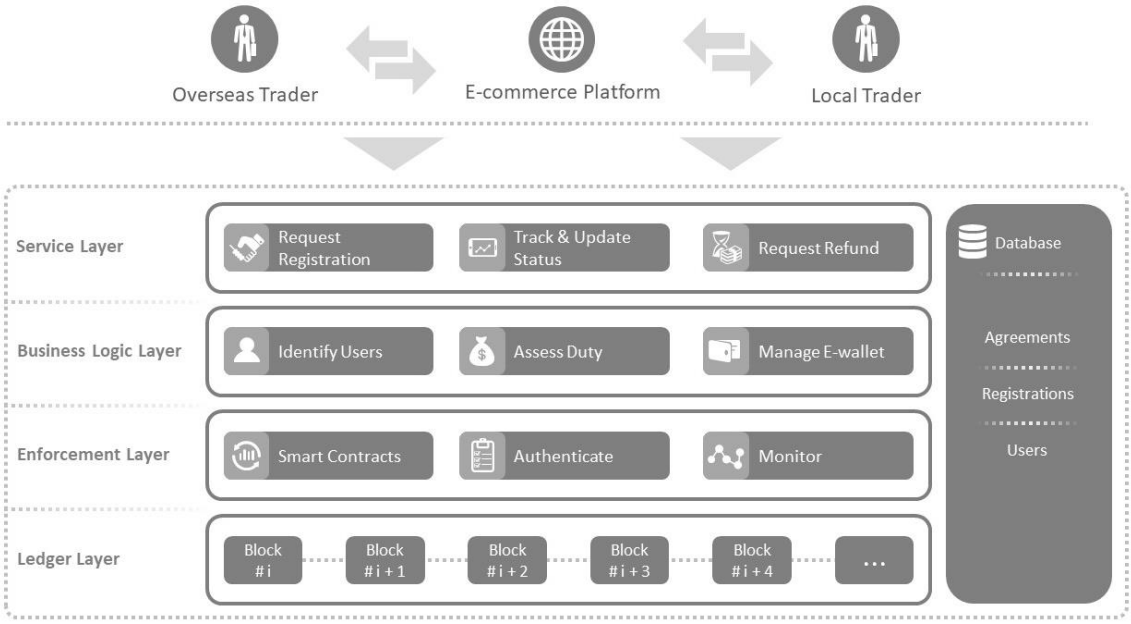


Figure 4: Ecommerce Service Architecture

### 4.1.3. Framework Key Processes

The ground concept of the proposed framework is to improve the trading experience, and in this section describe the interaction mechanism between the processes of this framework.

#### *4.1.3.1. Registration*

In the presented framework, registration is performed at three levels; namely the ecommerce platform, buyer and agreement levels in the framework. To interact with the framework, the ecommerce platform and the buyer have to be registered users with the framework. This is established by contacting the customs administration to gain access to the framework. In situations where the buyer represents a local trader, before gaining access to the framework, the local trader must obtain a business code, which is normally obtained from the economic department. In situations where the buyer represents an

individual, the framework uses the information provided by the ecommerce platform to identify and register his activities. Once the registration is performed, the registration details are stored in the local database as shown in Figure 4. This information is typically used by the identified users process to validate the identity of the traders and the ecommerce platform.

The agreement registration is considered as one of the primary processes in the presented framework since it relatively triggers a large number of processes. This process aims to perform a trade as shown in Figure 5. This process is called to either examine trade requirements or to perform trade registration. For any trade, the examination part is performed first. Once the buyer initiates the trading process, the examination starts by validating the identity of the buyer and the ecommerce platform. The validation is performed by calling the identity users service in the business logic layer. Then, as part of the examination, using the HS-Codes of the items included in the shipment, this process proceeds by determining if permits are required to be submitted. Additionally, using HS-Code, the process checks if any of the shipped items are forbidden in the country, and in this case, the seller and buyer will be notified to cancel the trade. The last step of the examination is to determine the required duties, and then send all of the obtained information to the buyer.

If the buyer chooses to accept the trading requirement, the registration part of this process starts by reserving the required duties from the buyer's e-wallet. Then, the trade agreement information will be stored in the database. To monitor the fulfilment of the trade agreement, the smart contract process in the enforcement layer is used to convert the trade

agreement into a smart contract. Using this contract, the fulfilment of several requirements such as permits submissions and delivery conditions can be monitored.

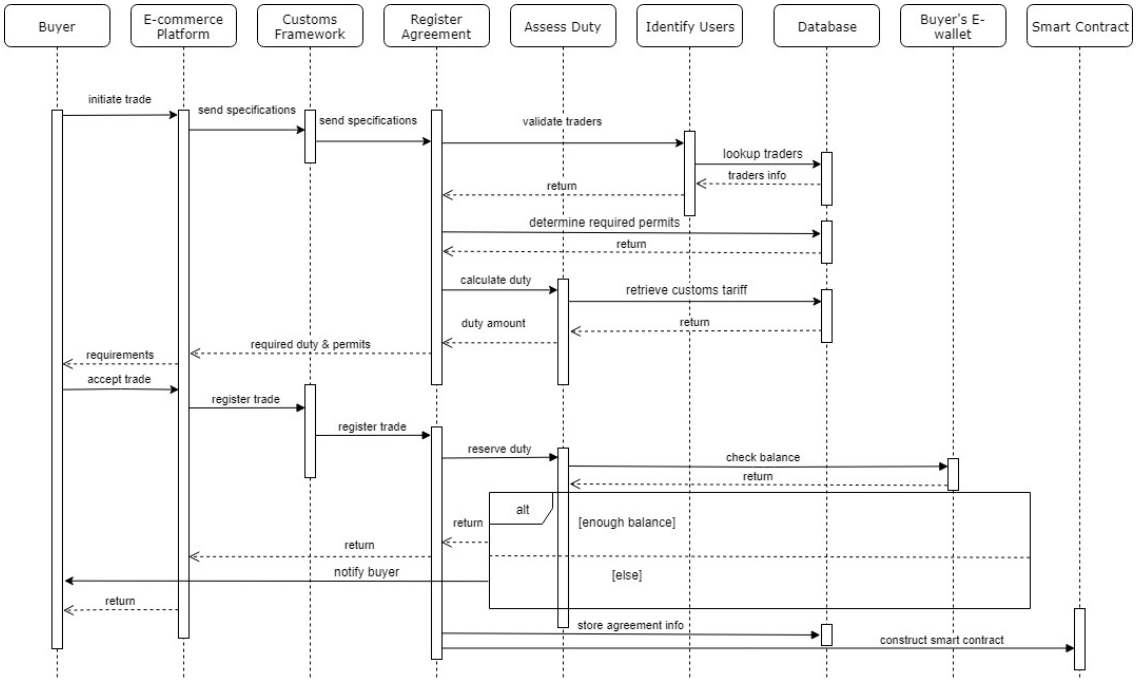


Figure 5: The Trade Registration Sequence Diagram

#### *4.1.3.2. Customs Clearance*

Once the shipment arrives at its destination, the customs administration will check the shipment information against the shipment information available from the framework. If the information match and all of the shipment requirements are met (i.e., permits requirements), customs will clear the shipment to be collected by the buyer. In other scenarios, the customs declaration management department will be informed to process the shipment. Figure 6 shows the main steps of the clearance process. Customs agent will use Track and Update Status service to retrieve the shipment identification information from the local database. Using this information, the framework will confirm with the smart contract component whether a permit request was initially requested by the customs

administration. In situation where no permits are required, the customs agents will proceed with the clearance process. Otherwise, the framework will be used to trace whether the buyer provided the required permits. If any of the permits are still not submitted, the buyer will be notified to complete the submission process. If all documents are submitted, the process continue to clear the goods. In cases where the shipment requires physical inspection, the declaration will be marked to be inspected by the inspection officers.

Once the shipment is cleared, the reserved duties from the buyer e-wallet will be transferred to the customs administrative account. The customs clearance process is performed using the Track and Update Status service in the service layer. In cases where local companies do not comply with the rules and regulations, fines will be applied, and the fine amount will be deducted from the company's e-wallet.

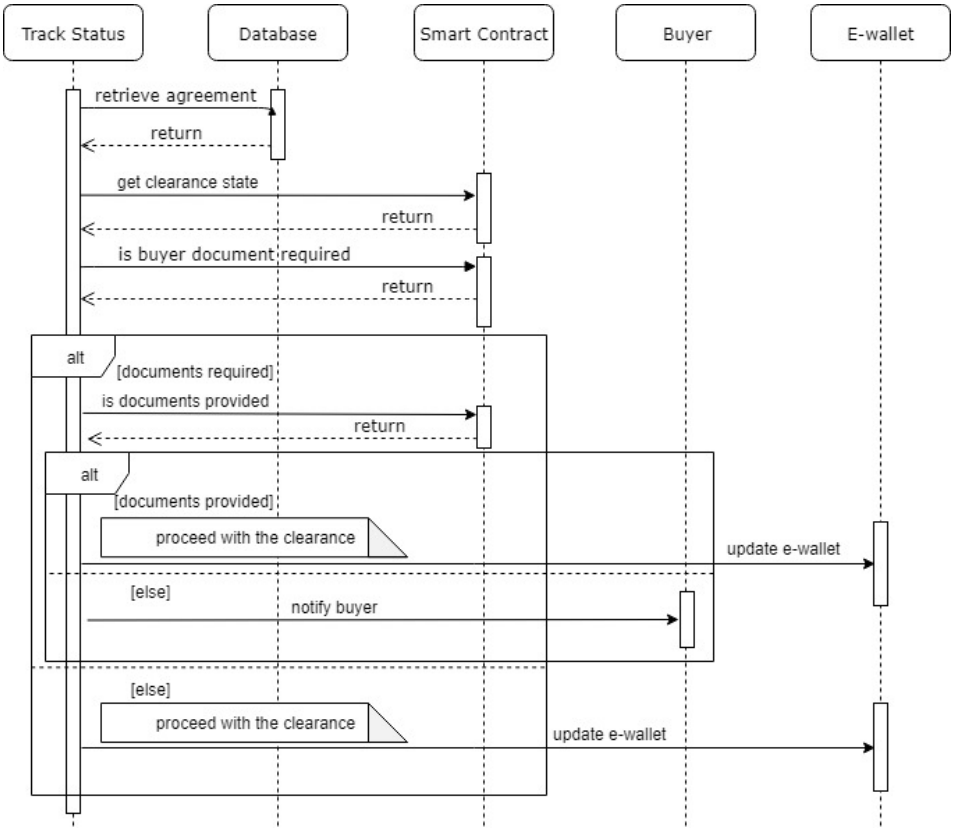


Figure 6: The Customs Clearance Sequence Diagram

#### *4.1.3.3. Drawback (Refund)*

The payer may request duty drawback service from customs administrations in case of product return by the consumer within a specified allotted time. Upon acceptance of the duty drawback request, customs administration will authorise the re-export of goods and initiate the refund of duty to the payer's e-wallet or account.

In the smart contracts layer, the monitoring process determines whether the refund service is permitted for a particular trade or not. This is decided based on the pre-defined time deadline. For instance, the maximum allowed time interval can be set to six months from the clearance date.

#### *4.1.3.4. Track and Update Shipment Information*

Tracking and updating shipment information deal with the processes that perform validation on the shipment status as well as updating the current status of the shipment. In the service layer, traders and customs administration can request the current status of a particular shipment using the track and update service. This service verifies the user permissions by contacting the "Identify Users" process in the business logic layer. Then, the required information will be given to the user using the smart contracts in the enforcement layer. Additionally, the ecommerce platforms, traders and the customs administration can update the status of the shipment using this service at any point of time. This update includes information about the shipment such as the current location, shipper information, expected arrival date and clearance status.

The presented processes cover all aspect of a trade, starting from the order creation until the shipment arrives at the destination. Hosting the framework at the customs administration helps in monitoring the growth of the ecommerce trading domain. This is established by using the information to analyse the ecommerce market behaviour.

#### *4.1.3.5. Network Design and Smart Contracts*

The proposed network architecture consists of two peers, namely the customs administration and the ecommerce platform. The owner of the ecommerce platform hosts the latter peer. In the proposed network, Ethereum platform is used to host the on-chain functionalities of this framework, where smart contracts are implemented using Solidity programming language. Ethereum is recognized as a public Blockchain where any user can interact with the network. Thus, the employed identification process will ensure that no unauthorized users can trigger any deployed smart contract. In solidity, the "address" data

type is used to store Ethereum accounts information. The proposed solution uses the e-wallet addresses that are provided by the Ethereum platform to identify the permitted users. Therefore, the permissions will be assigned to the users after confirming the corresponding e-wallet's address with the pre-determined set of addresses. At any stage during the processing life cycle of a trade, the seller, buyer, and the customs administration participants will use these addresses to contact the framework to change the status of any shipment

For each trade in the Blockchain, several smart contracts are constructed during the processing of the trade. Each trade is represented by several variables, which are summarized in Table 3. These variables represent the characteristics of each trade and their initial values. In Solidity, there is no presence for the float and double data types. To overcome this limitation, "uint" data type has been used where the fractions have been limited to two decimals. Therefore, the stored numbers will be multiplied by 100. For instance, if (price=22,888.75) needs to be stored, then the stored value will be (price=2,288,875). The type of the trade (tradeType) can be either B2B (value 0), or B2C (value 1). The permitted period of the drawback (drawbackPermittedPeriod) is only applicable in the B2B trade types. The actual documents in the proposed Ethereum Blockchain are stored on an external database (Off-Chain).

Table 3: Trade Variables

| Variable Name | Description | Data Type |
|---|---|---|
| tradeType | This field represents the type of trade. The possible values are:<br>• 0: B2B<br>• 1: B2C | uint8 |
| commoditiesIds | This field represents the commodity id as per the invoice. | uint |
| commoditiesNames | This field represents the commodity name. | string |
| commoditiesDescriptions | This field represents the descriptive details of commodity. | string |
| commodityTypes | This field represents the type of commodity such as electronics, medical among others. | string |
| commodityOrigins | This field represents the country of origin for particular commodity. | uint |
| unitePrices | This field stores the unit price for each of the commodities. | uint |
| shipmentMethod | This field represents the shipment method for a particular trade. | uint |
| customsDuties | This field stores the corresponding customs duty for each of the commodities in a particular trade. | uint |
| drawbackPermittedPeriod | This field stores the permitted time period for applying for duty drawback. | uint16 |
| requiredDocuments | This field stores an array of the required permits from various authorities to proceed with the shipment. | bool |

During the trade life cycle, certain status variables are maintained through the smart contract to capture the different status changes of the shipment under consideration. These variables capture various information such as documents requirements and clearance status. Table 4 summarizes the descriptive information of these status variables. For instance, if the HS-Code for the imported goods define that the buyer must submit certain permits from government authorities, the smart contract automatically blocks the clearance

of the shipment until the documents are provided as shown in Function 1 ($requireBuyerDoc()$). Once these documents are received, the status of the document's requirements will be changed to indicate the completeness of the requirements to continue with processing the clearance. At any stage during the processing time of the shipment, both of the trading parties and the customs administration can inquire about the current state of the shipment via calling the check Trade Status service as shown in Function 2.

Table 4: Trade Status Variables

| **Status Name** | **Description** | **Data Type** |
|---|---|---|
| isDocRequiredByBuyer | This status represents if there are documents required by the buyer. True value means that the buyer is required to provide document certain documents. | Bool |
| isBuyerReqDocsProvided | This status represents if the required documents required by the buyer are provided or not. | Bool |
| isTradeAccept | This status represents if the corresponding trade is accepted or not. | Bool |
| clearanceStatus | This status represents the clearance decision regarding particular trade. The possible clearance values are:<br>• 0: not cleared,<br>• 1: clear goods,<br>• 2: clear with inspection,<br>• 3: further investigation required.<br>• 4: clear with permit | uint8 |
| drawbackRequest | This status represents if the B2B buyer request for a duty drawback corresponding trade. | bool |
| drawbackStatus | This status represents if the B2B buyer request has been approved by the customs administration or not. | bool |
| clearanceDetails | This field store the clearance detailed information such as the permits required, documents required, suspicious information among others. | string |

Eventually, the shipment details such as type, weight and the quantity will be verified upon the shipment arrival to the customs premises. If information stored on the blockchain does not match the received items, the clearance process of the shipment will be suspended. On the other hand, if the shipment details match the information stored on the blockchain, the clear goods process, which is shown in Function 3, will be invoked to clear the shipment after deducting the duty amount from the buyer's e-wallet through the ***deductDuty***() function.

Once the shipment processing is commenced, the traders and the customs administration can change and inquire about the current shipment location through the setLocation and getLocation functions respectively. For instance, during the processing of the shipment at a transit location, the setLocation function can be invoked by the ecommerce platform to reflect the current location on the blockchain.

Eventually, the shipment details such as type, weight and the quantity will be verified upon the shipment arrival to the customs premises. If information stored on the blockchain does not match the received items, the clearance process of the shipment will be suspended. On the other hand, if the shipment details match the information stored on the blockchain, the clear goods process, which is shown in

Function 3, will be invoked to clear the shipment after deducting the duty amount from the buyer's e-wallet through the ***deductDuty***() function.

Function 1: The Buyer's Document Requirements

```
Function requireBuyerDoc (bool require, address _user) public
1   require (_user == admin_address, "Not Authorized User");
2   if (require == true)
3       isDocRequiredByBuyer = true;
4       isBuyerReqDocsProvided = false;
5   else
6       isDocRequiredByBuyer = false;
7       isBuyerReqDocsProvided = true;
```

Function 2: Trade Status Verification

```
Function checkTradeStatus (address _user) public returns (string)
1   String out="::"+ clearenceState;
2   require((_user==admin_address ||
            _user==buyer_address)
            "Not Authorized User");
3   if (isDocRequiredByBuyer)
4       out+="buyer documents required::"
5       if (isBuyerReqDocsProvided)
6           out+="documents provided::"
7       Else
8           out+="documents not provided::"
9   If (tradeType==0)
10      If(drawbackRequest)
11          out+="drawback requested";
12      else
13          out+="drawback not requested";
14  out+= clearanceDetails;
15  return out;
```

Function 3: The Clear Goods Function

| **Function** clearGoods (uint8 _clearence_state, address _user) **public returns** (bool) | |
|---|---|
| 1 | require (_user == admin_address, “Not Authorized User”); |
| 2 | **if** (deductDuty()) |
| 3 | clearence_state = _clearence_state; |
| 4 | **returns** (true); |
| 6 | **return** (false); |

## 4.2. System Evaluation

This section illustrates the activities taken to evaluate the proposed framework.

### 4.2.1. Web Application

To simulate the end-to-end process, a web application was developed that hooks into the blockchain framework to get up to date information related to the specific products shared by the customs administration of the intended destination. The blockchain trade facilitation framework represents the core of the solution and the web component represents the front-end interface. The web application components are implemented using Model-View-Controller (MVC) in ASP.NET 4.51 and HTML5. The connection between the framework and front-end interface is established through the use of jQuery Asynchronous JavaScript and XML (AJAX) calls. AJAX is an effective way of exchanging data between various components. These jQuery AJAX calls will connect the web application with the ledger which is stored in an Ethereum blockchain.

The web application comprises three different areas that simulate the buyer journey, shipping agent representative relevant tasks, and the customs declaration officer relevant

tasks. The buyer journey starts with choosing the ecommerce platform. Then, the buyer can explore the ecommerce platform and add the desired products and quantities to the shopping basket. Once the buyer completes shopping, he may proceed to the checkout process. There are some slight differences in the checkout process from the normal one that are represented by the seamless integration with the customs administration through the customs blockchain framework. In the checkout process, a message will be sent with the trade specifications to the framework through the jQuery AJAX call for trade registration with verification parameter. An immediate response from the framework will appear in the checkout page with customs details and requirements in case of any duties or required permits or any forbidden items. The buyer then can proceed with the transaction and go through the checkout process on the ecommerce platform. Upon the completion of the checkout process, the complete transaction details will be also sent to the customs framework through the jQuery AJAX call for trade registration with registration parameter. As stated in the Registration process, a smart contract will be created with the trade specifications and also these details will be stored off-chain in the customs administration with reference to the smart contract. This makes all shipment information available for the customs administration internal systems for risk assessment. After that, the buyer may track the status of the shipment through the ecommerce platform track shipment page. This page pulls the shipment status updates using the jQuery AJAX call to the “Track Status” service from the customs framework.

The shipping agent representative’s journey starts when the buyer completes the checkout process. The shipping agent is responsible to deliver the purchased products to the destination country. Throughout the shipping process, the shipping agent representative

will update the status of the shipment by selecting the current status in the “Update Shipment Status” page. Upon any status update, the “Update Status” service in the customs framework is called using the JQuery AJAX to reflect this update in the blockchain. Furthermore, the shipping agent representative can create the customs declaration through the “Create Declaration” page. This page includes of all shipments pulled from the customs framework through the jQuery AJAX calls. The shipping agent representative will select the desired shipment, update the information if required and proceed with creating the declaration. All declaration information will be automatically populated with the most recent updated information pulled from the blockchain.

Finally, customs declarations that are required to be mitigated are sent to the customs declaration management. The customs declaration officer reviews and verifies the received declarations and takes the decision to clear the shipment or clear with inspection or decline it.

### 4.2.2. Workshop

As part of the evaluation process, a workshop has been conducted. The workshop was managed by three advisors. The participants in this workshop represent three categories, namely buyers, shipping agents’ representatives and customs clearance officers. Accordingly, 10 buyers, 5 shipping agents’ representatives and 5 customs clearance officers were invited to attend this workshop. The buyers comprise 5 users who represent several businesses and 5 users who represents individual users. The shipping agents’ representatives were invited from 2 different companies. Experts in customs domain were invited to play the customs clearance officers role. Prior to the workshop, the participants were asked to read and sign the consent form.

At the beginning of the workshop, an overview of the proposed solution was presented to the participants followed by interactive discussion. After that, a live demonstration for the application was conducted to illustrate the end-to-end process with all features. The participants were given the chance to freely ask questions about the application. After that, several tasks were given to the users, as illustrated in Table 5. Participants were asked to report any error or fault in the web application. Moreover, each type of participants was given a specific time to complete the tasks according to their complexity.

Table 5: Users Tasks in the Workshop

| Task | Task Description | User Category |
|---|---|---|
| 1 | Select the Ecommerce platform | Buyer |
| 2 | Search for the desired items or products throughout the application | Buyer |
| 3 | Add the desired item to the shopping cart | Buyer |
| 4 | Specify the required quantity | Buyer |
| 5 | Repeat 2-3 if you need to buy other items | Buyer |
| 6 | Open the shopping cart and proceed with the purchase | Buyer |
| 7 | Review the Customs Administration response with the customs requirements | Buyer |
| 8 | Accept or reject the trade | Buyer |
| 9 | Open the shipping agent page | Shipping Agent |
| 10 | Locate the desired shipment | Shipping Agent |
| 11 | Update the status to “Picked by the Shipper” | Shipping Agent |
| 12 | Update the status to “Package Shipped” | Shipping Agent |
| 13 | Update the status to “Arrived Destination Country” | Shipping Agent |
| 14 | Initiate the Customs Declaration | Shipping Agent |
| 15 | Open the customs clearance page | Clearance Officer |
| 16 | Retrieve shipment details | Clearance Officer |
| 17 | Verify shipment details | Clearance Officer |
| 18 | Select the desired decision | Clearance Officer |

| | | |
|---|---|---|
| 19 | Verify the declaration status | Buyer |
| 20 | Verify the declaration status | Shipping Agent |
| 21 | Upload the required Permits if required | Buyer |

Efficiency of the web application is defined as the resources consumed when performing a task. In this study, the time on task is considered as a measure of efficiency. All users were able to complete the tasks within the time limit as expected. For effectiveness, reported errors were considered. During the workshop, no errors or bugs were reported.

### 4.2.3. The System Usability Scale

Participants were given the SUS questionnaire to seek their feedback on their experience after completing the workshop tasks. The SUS was developed by Brooke as a tool to measure the perceived usability of a system (Brooke 1996). The SUS questionnaire comprises 10 questions with 5-point Likert scale to assess the agreement level with the SUS questions. The rate for each question varies from 1 for “Strongly Disagree” to 5 for “Strongly Agree”. The questions are divided into 2 groups, namely the odd numbered questions and the even numbered questions. The odd numbered questions (1,3,5,7 and 9) represent the positive questions. Whereas, the even numbered questions (2,4,6,8 and 10) represent negative questions. The purpose of this approach is to avoid bias in the participants answers (Lewis 2018). The standard questions of the SUS proposed by Brooke (1996) have been revised to reflect the purpose of the proposed solution as illustrated in Table 6. The overall score of the SUS is calculated in 2 stages. The first stage includes normalizing the user inputs. In case of odd numbered questions, the new value results from subtracting 1 from the user’s selection value (USV) (USV-1). In case of even numbered

questions, the new value results from subtracting the user's selection value from 5 (5-USV). The user's final result represents the total sum of all new values multiplied by 2.5. And the final result of SUS represents the average of all user's final results.

Table 6: The System Usability Scale Revised Questions

| # | **Question** |
|---|---|
| 1 | I think I would like to use the Customs-based Trade Facilitation website frequently. |
| 2 | I found the Customs-based Trade Facilitation website unnecessarily complex. |
| 3 | I thought the Customs-based Trade Facilitation website was easy to use. |
| 4 | I think that I would need the support of a technical person to be able to use the Customs-based Trade Facilitation website. |
| 5 | I found the various functions in the Customs-based Trade Facilitation website were well integrated. |
| 6 | I thought there was too much inconsistency in the Customs-based Trade Facilitation website. |
| 7 | I would imagine that most people would learn to use the Customs-based Trade Facilitation website very quickly. |
| 8 | I found the Customs-based Trade Facilitation website very awkward to use. |
| 9 | I felt very confident using the Customs-based Trade Facilitation website. |
| 10 | I needed to learn a lot of things before I could get going with the Customs-based Trade Facilitation website. |

Table 7 and Figure 7 present the SUS resulting from the workshop for each of the user's categories as well as the overall result. The user's category with the highest acceptability from the proposed system is the customs declaration officer. This indicates that the proposed solution significantly reduces the complexity of the processes typically performed by customs officers. Accordingly, the customs declaration officer is no longer required to manually verify the details of the ecommerce trade. The buyer achieves excellent usability rating according to Bangor et al. (2008). Typically, the buyer has no clear idea about the customs requirements until the goods reaches the destination country.

By adopting the proposed solution, the buyer will know beforehand all of the requirements before initiating the trade. In addition, the proposed solution grants the ability to the shipping agent representative to perform and track all of the shipments in timely manner. And this is the main reason behind the reported SUS significant results. Overall, the result shows that the system is acceptable by all users' categories.

Table 7: The System Usability Scale Results

| **Users Category** | **SUS Result** |
|---|---|
| Buyers | 85.25 |
| Shipping Agents' Representatives | 82.50 |
| Customs Declaration Officer | 86.00 |
| Overall SUS Average | 84.75 |

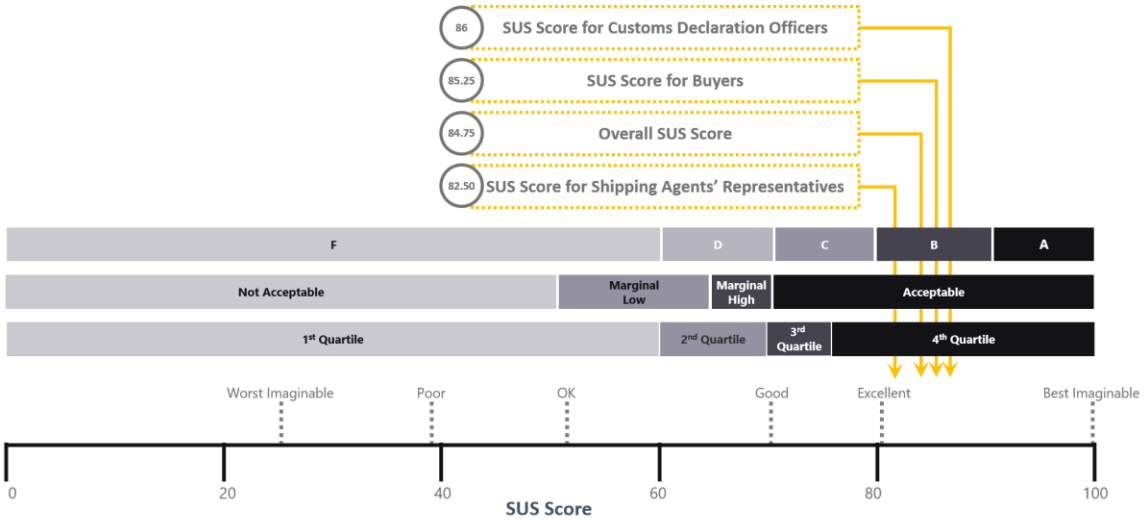


Figure 7: The System Usability Scale Results Representation

## 4.3. Conclusion

In this chapter, a framework to address the challenges from customs perspectives related to ecommerce trade was proposed. This framework focuses on simplifying the trading process to ensure traceability and data integrity. This is achieved by integrating the

framework with blockchain technology. The requirements of simplifying the trading process are addressed in order to resolve the duty collection and drawback challenges. The proposed framework addresses B2B and B2C trade categories, where the traders can use this framework to further ensure the trade simplicity without complications.

This framework represents the foundation, which ensures visibility across the chain for all participating parties and guarantee data integrity, as well as determine where goods are originated using built-in capabilities of blockchain technology. The following two chapters illustrates how data mining clustering algorithms can be used to identify risks to further support legitimate trade.

# 5. Chapter 5: Outlier Detection for Customs Post Clearance Audit Using Convex Space Representation

Customs administration has the responsibility of protecting the society and sustaining the economic development. This task is performed by making sure that the trade is following the country legislation and rules. The task of protection involves auditing and assessing the risk factors for each company trading behaviour to identify any potential risk. This auditing task involves analysing all of companies' shipments to identify whether any shipment poses a security risk. Performing this task is a major challenge since a large number of attributes that describe each shipment are all required to be investigated and analysed.

By law, the companies must retain the original documents for all shipments for a defined period based on the company's trade license. Moreover, the customs administrations perform a post clearance audit on the company's historical transactions and may request for any of the original documents during this period (WCO 2018b).

Furthermore, as part of the declaration processing, customs administration collects the duty amount according to the tariff code of the commodities. This duty is associated with the commodity type and the duty value varies based on the commodity nature. Thus, in addition to the typical risk factors such as smuggling, the trader may manipulate the commodity values to avoid paying duties or reduce the duty amount. The valuation manipulation is a major concern to the customs administrations since it may result in loss of revenue.

The post clearance audit takes into consideration the trader's behaviour in order to correctly analyse any risk factors that rise from the historical transactions of the company. This analysis is a process that is typically performed by auditors at the organization's premises. This process is a time consuming and may not be efficient since the selection process of the companies may not consider the entire company's history. Therefore, this indicates the need for building a solution that optimizes the process of selecting the company that is under consideration for customs audit.

In this chapter, the aim is to develop a *convex-based algorithm* that provides an efficient solution that helps in determining if a company shipment poses any risk. This solution can be adopted by customs administrations to simplify the overall auditing process. The proposed solution takes into account the trading pattern of each company during the analysis of the historical transactions. Using this pattern, the proposed solution is able to determine shipments that poses any security issues. For instance, Figure 8 illustrates the representation of the shipments in three-dimensional space where goods value, country of origin and duty value represent each of the shipments.

Each shipment in the proposed approach is considered as a point in multi-dimensional space, where the dimensions in this space represent the shipment attributes such as cost, country of origin, type among others. Accordingly, the entire history of all shipments is represented as points in this space. Whereas, the locations of the shipments in this space highlights the similarity between them. Using such representation, the problem of detecting risky shipments can be seen as an outlier detection problem. Once the trader historical transactions are represented in the multi-dimensional space, the analysis is performed by establishing a convex hull for the represented points. Convex hull results in a convex space.

Convex space is defined as a space representation where any line that connects two points fall into the current subspace (Blelloch et al. 2020). By using convex subspace, the aim is to extract the behavioural attribute for shipments that are most likely to be in the safe and risky shipments.

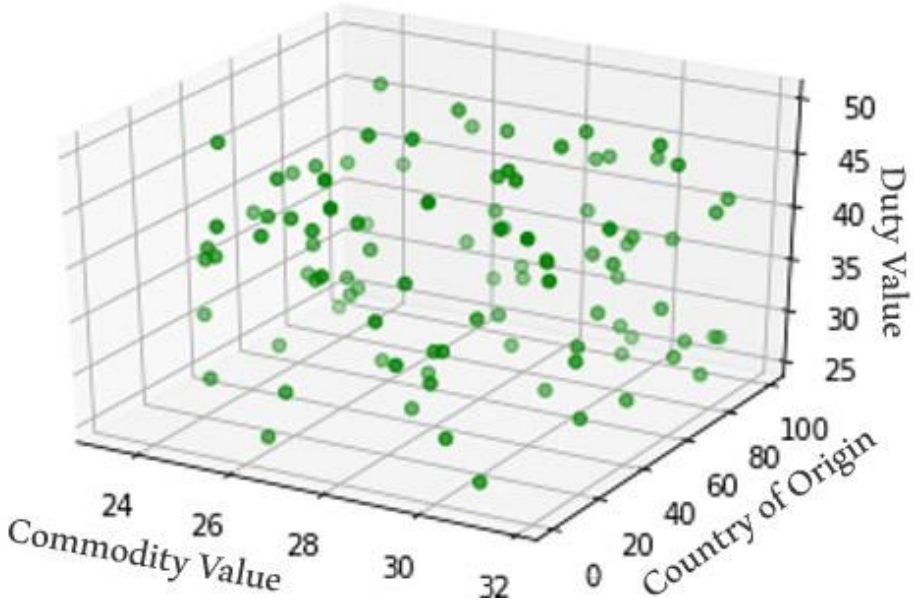


Figure 8: Multi-Dimensional Space Representation Example

The problem of detecting abnormal activities has been extensively studied in the literature, as discussed in chapter 2. To address this problem, the use of outlier detection algorithms has result in significant improvement in different domains. To address the problem of detecting the abnormality behaviour in the companies trading behaviour from customs perspective, this study aims is to construct convex space representation for the companies trading transactions. In such space, the closer the transaction to the border, the more likely that this transaction is considered an outlier. Using such approach is beneficial through simplifying the detection mechanism since at any stage the process of assessing a

new transaction works by identifying the distances between the points inside the convex space.

This chapter is composed of four sections. The next section is concerned with the methodology that is adopted to develop the convex-based approach. The second section illustrates the proposed algorithm and the different stages that are followed to achieve the aim of this study. The third section discusses the experimental sceneries and illustrates the research findings. Finally, the fourth section provides brief summary of the study and presents the key findings.

## 5.1. Methodology for Outlier Detection in Customs Post Clearance Audit

This work aims to determine whether a certain company shipment poses any risk according to the company trading behaviour. This section illustrates the proposed convex-based approach that is performed through representing the company's historical shipments during the analysis to simplify the analysis process. This approach consists of three steps, namely space representation, convex space construction and riskiness identification.

Various methodologies were adopted in the literature to perform data mining tasks. The following methodologies are well-known and have been used in various application domains: KDD (Fayyad, Piatetsky-Shapiro & Smyth 1996; Siyam, Alqaryouti & Abdallah 2020; Alkashri et al. 2021), SEMMA (Azevedo & Santos 2008a) and CRISP-DM (Pete et al. 2000; Altaheri & Shaalan 2020). These methodologies aim to establish proper holistic understanding prior to performing the data mining tasks in order to gain insights and discover hidden knowledge. CRISP-DM methodology is adopted in this study as it is an

industrial and technological neutral. Furthermore, it is one of the commonly used methodologies in the development of knowledge discovery projects data mining methodology (Mariscal, Marbán & Fernández 2010).

Figure 9 illustrates the six stages that represents the CRISP-DM methodology. In the first stage, a proper understanding of current and desired business requirements is essential. In the second stage, the data collection activities take place followed by studying and understanding the nature and the characteristics of the data. The third stage is concerned about data preparation and cleansing to produce harmonized dataset. In the fourth stage, the modelling exercise starts by taking the input prepared dataset and building the convex-based approach. In the fifth stage, the results of the modelling exercise are evaluated. Finally, the output model is deployed to production in order to handle real-time data.

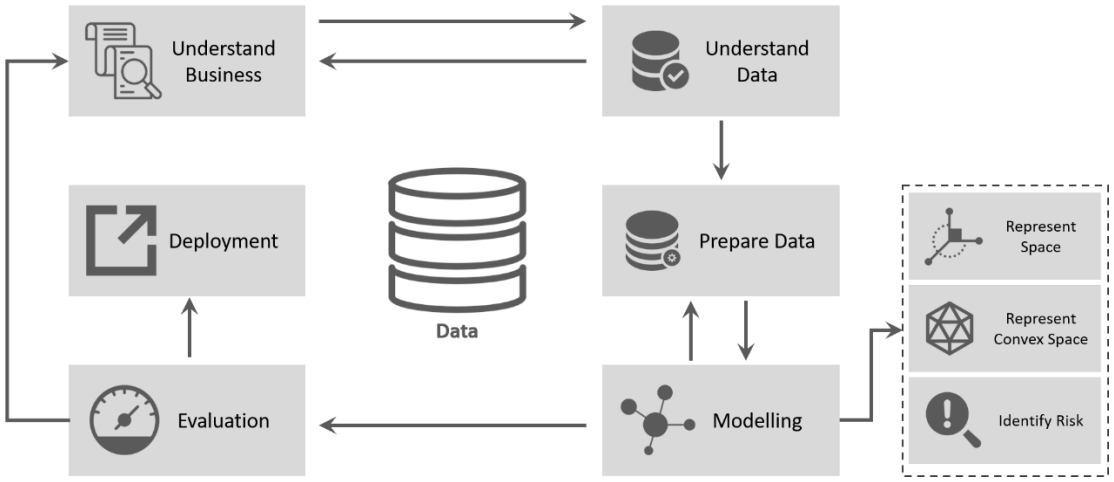


Figure 9: Methodology for Outlier Detection in Customs Post Clearance Audit

The data used in this study is obtained from Dubai Customs which basically represents random company shipments information. This dataset was prepared for shipments conducted in the year of 2018 and includes 500,000 records. The records are labelled to

identify the risky company declarations. As part of data preparation, all shipments for importers who have duty exceptions are removed from consideration. These declarations have been removed since the importer has no duty to pay. The following sections discusses the details of the proposed convex-based approach and the various experiment settings to evaluate the performance of this approach.

## 5.2. Convex-based Approach

In the proposed approach, the companies trading shipments are represented as points in multi-dimensional space. Based on the commodity type, a convex space is constructed to represent the available shipments information. In the convex space, connecting any two points inside this space will result in having the entire lines fall into the same space. The objective of such representation is to simplify the process of monitoring the company trading behaviour. The adoption of such mechanism aims to establish "Confidence" space. This space represents shipments with distinct safe characteristics. Accordingly, to construct the convex space, the available historical shipments data is analysed to include only those shipments with normal behaviour.

The available historical shipments information is divided into safe and risky to establish this space for each trade commodity type. Each one of these shipment types will be processed separately. The representation step works to represent each shipment as point in three-dimensional space, where the dimensions of this space are the country of origin, commodity value and duty value.

Once all available historical data are represented in this three-dimensional space, the convex space representation works to determine the two convex spaces to cover the safe

and risky shipments separately. The point of constructing this convex space is to establish a relationship between the shipment locality in the space and the risk factors for this particular shipment.

### 5.2.1. Space Representation

The objective in this step is to use the shipments attributes in representing the shipments in multi-dimensional space. Accordingly, the attributes that are used have to help in clarifying the relationship between the shipments. In this work, the attributes used for space representation are the country of origin, commodity value and duty value. These attributes are selected because they can be used to determine the similarity in terms of shipment path and commodity value.

The country of origin as well as the other attributes are selected to help in predicting the variation of the value between the countries. For instance, electronics from Japan are most likely to have more value than electronic from China. The duty value as well as the value of the commodity can together highlight the potential of smuggling or suspicious behaviours specially when these two values do not follow predictable pattern. This representation will be conducted for each commodity type using the entire available historical shipments information. In addition, this representation will be conducted to the safe and risky shipments.

### 5.2.2. Convex Space Representation

This step aims to represent the safe and risky shipments into two convex spaces. Such representation aims to simplify the process of identifying whether the shipment under consideration is safe or risky. In this vein, the location of any point helps in determining

whether this point can be identified as outlier (safe) or not. Accordingly, points closer to the border of the risky convex space are more likely to be outliers (safe) compared to points located inside the convex space. Figure 10 illustrates an example of this representation, where the represented shipment point in green colour is expected to be an outlier since it is closed to the convex border.

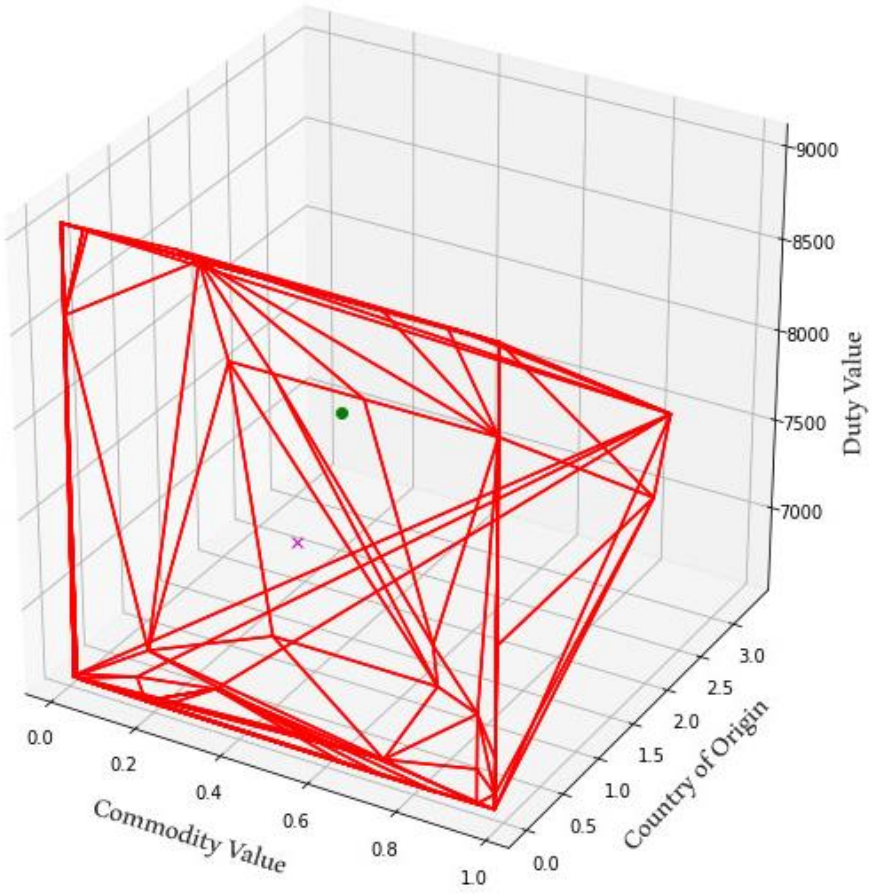


Figure 10: Convex Space Representation Example

To represent the risky space as a convex space, this study employs the Quickhull algorithm proposed by (Barber, Dobkin & Huhdanpaa 1996). The convex hull for a space (safe or risky) is the smallest convex set that contains all the space points.

Given the number of dimensions ($d$), this algorithm starts by selecting $d + 1$ points, which do not share a plane or hyperplane, and these points are used to establish the initial hull. Then, for each facet ($f_i \in F$) of the constructed hull, the algorithm proceeds by constructing the *outside* set of the facet. This set represents all unassigned points that are located above the facet. Each point can be assigned to only one facet. Now, for each facet with non-empty outside set, the process of expansion starts by selecting the farthest point ($p$) from this set, and initialize the visibility set ($V$) to be $F$. This visability set is expanded by adding all neighbouring facets, which also located below $p$. The boundary of the visiability set forms the set of horizon ridges ($H$). Then, for each ridge, facets are created by combining these ridges $H$ to $p$. The created facets are then linked together based on their locality to establish a new hull. This process then repeated by determining the outside set for each new facet ($F'$), and perform the expansion process. This process stops when the constructed (new) facets have empty outside sets.

In the risk identification step, this convex representation is used at several stages to analyse the shipments from different perspectives.

### 5.2.3. Riskiness Identification

In this step, the main idea is to use the two convex space representations to identify whether a certain shipment poses any risk. The identification of this risk suggests that the company who is responsible for this shipment is a target for auditing. The use of the two convex spaces aim to filter the analysis criteria. The risky convex space is expected to be significantly smaller than the safe convex space. However, the overlapping between the safe and risky convex spaces is considered as unknown space that does not belong to the

safe convex space. Accordingly, the focus will be on the resultant safe convex space once this overlapping is eliminated.

Then, the actual risk identification is performed using the safe convex space for validation using the HS-Code. The HS-Code is an eight digits number that is used to uniquely identify the traded commodities. Accordingly, the performed analysis takes the trader behaviour in terms of his previous transactions.

The objective of this step is to assign a risk score for the shipment based on its location in the convex space. A shipment could have single or multiple items, and accordingly each shipment will be represented by several points equal to the number of commodity types in this shipment.

As a pre-processing step, the convex space representation is established and the overlapping areas between the two spaces are eliminated. The calculation of the risk score for a given point takes into consideration two factors, namely distance to border and the distance to the neighbours. The points that are closer to the border of the convex space are more likely to be outliers (risky shipments). Whereas, the points that are closer to the centre of the convex space are more likely to represent safe shipments. In addition, the distance of the point to its neighbours also influences the risky score. For instance, assume that a single point that is located inside the convex space where the average distance between this point and its neighbours is relatively small. This point is more likely to be safe compare to points located in faraway distance from the neighbouring points.

The calculation of the points risk factors starts by determining the set of points that must examined. Once these points are identified, the process proceeds by analysing the

identified node in sequential order. For each point, the algorithm calculates the distance between this point and the centre of the convex $d_c$ as well as the border of the convex $d_b$. Points that are closer to the centre of the convex are more likely to be safe. These two values are used to calculate the border factor ($f_b$) which is calculated by dividing the distance to the centre over the distance to the border ($\frac{d_c}{d_b}$). When the value of this factor is equal to one, the distance between the point and the border is equal to the distance to the centre.

Next, the distance factor is calculated by taking into consideration the average distance between this point and its k-neighbours. The distance factor of a point ($p$) is calculated as follows:

$$f_d(p) = \frac{avgK - d(p)}{avgK - d(c)}$$

Where $avgK - d(p)$ refers to the average distance between point $p$ and its k-neighbors. $avgK$-$d(c)$ is the average distance between the centre of the convex and it its k-neighbours. The value of this factor indicates whether the point is located in dense or sparse area of the convex. Once these two factors are calculated, the riskiness score of the HS-Code is calculated as follows:

$$HS - score(p) = f_b(p) \times f_d(p)$$

Using this score, points located near the border of the space or points in areas with low density will have relatively high riskiness score. In particular, any risk score value that is higher than one ($> 1$) is worth investigating. This is due to the fact that points with score

higher than one are either near the border of the space or are located in less density areas. Function 4 describes the proposed convex-based algorithm in this study.

Function 4: The Proposed Convex-based Algorithm

RiskScore (p, k, level)

**input** p: the new shipment point
***level***: 0 $HSCode$, 1: $HSCode\ header$
***output***: $score(p)$

$Convex \leftarrow retriveConvex(p\ HS-code)$

If $p$ inside $Convex$ **then**

    $d_c \leftarrow DistanceToCenter\ (Convex, p)$

    $d_b \leftarrow DistanceToBorder\ (Convex, p)$

    $f_b \leftarrow \frac{d_c}{d_b}$

    $f_d(p) = \frac{avgK - d(P)}{avgK - d(C)}$

    $score(p) = f_b(p) \times f_d(p)$

Else

    $score(p) = 0$

**Return** $score(p)$

## 5.3. Experiments and Discussion of Results

This part illustrates the various set of experiments that have been conducted to evaluate the performance of the proposed convex-based outlier detection approach.

In this section, several sets of experiments are conducted to evaluate the accuracy of the proposed approach. The input dataset represents 500,000 shipments declarations occurred during the year of 2018 obtained from Dubai Customs. As a pre-processing step,

the declarations applications are grouped based on the HS-Code. Then, the lowest 10% of these declarations' applications are removed. The objective of this step is to increase the accuracy since as mentioned, low number of declarations for a given HS-Code might result in reducing the anticipated performance. Once this filtration step is performed, the declarations are grouped based on the risk factor into safe and risky groups. To determine the points that will be used for each convex space representation, the *k-means--* (*k-means minus minus*) to remove the points that are on the border with respect to the safe space and risky space is employed. The *k-means--* is an extension to the well-known *k-means* algorithm where the Farthest $n\%$ of the nodes is removed from consideration. In the proposed approach, the removed $n\%$ of the nodes represents the borders of the convex space. This percentage is eventually converted to absolute value that represents the *k* value.

The HS-Codes are grouped based on the number of shipments that belong to each HS-Code. Accordingly, in the presented experiments, the percentage of HS-Code groups selected is varied and the impact of the $k$ value in the *k-means--* is studied. In addition, the testing sample was prepared by selecting a random 10% from each resulted HS-Code convex clusters.

To illustrate the impact of the percentage of the removed points in *k-means--* on the overall accuracy of the proposed approach, several experiments were executed while varying the percentage while the *k* value set to 5 and percentage of the selected HS-Code groups set to 25%. Using *k-means--* as a pre-processing step aims to reduce the possibility of having noisy data around the convex cluster. The use of percentage measure aims to simplify the experiment. This measure was eventually converted to *k* value as input parameter for the *k-means--* algorithm. As illustrated in Figure 11, increasing the

percentage of the eliminated points to over 15% results in reducing the accuracy of the proposed approach. Therefore, the impact of this percentage in this pre-processing step on the approach performance needs to be taken into consideration. Thus, increasing the percentage of this input parameter results in reducing the number of the actual shipments that are used in the convex space representation. Therefore, increasing the percentage of this parameter without careful analysis will impact performance since it will reduce the number of available shipments for each convex space representation. On the other hand, having low percentage (<=15%) will improve the performance since it will result in removing the points that are located at the border of the space.

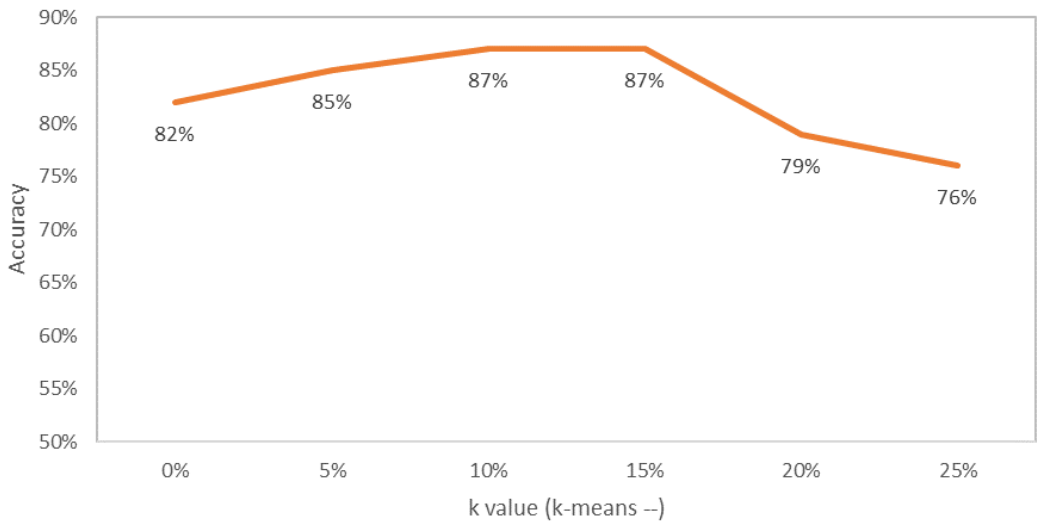


Figure 11: Performance Evaluation when changing the k Value in k-means--

Figure 12 illustrates the impact of changing the percentage of the selected HS-Code groups on the accuracy of the proposed algorithm while the *k* value is set to 5. In this experiment, the assumption is that shipments with riskiness score greater than one has the potential being labelled as outlier and therefore it requires further investigation. From the figure, it is evident that increasing the percentage of the selected HS-Code groups has a

stable impact on the accuracy until the percentage reaches to 55%. After this percentage, the accuracy started to drop. This is attributed to the fact that large groups will obtain high accuracy due to the number of shipments in these groups. On the other hand, using any outlier mechanism on groups with low number of shipments will obtain less accuracy.

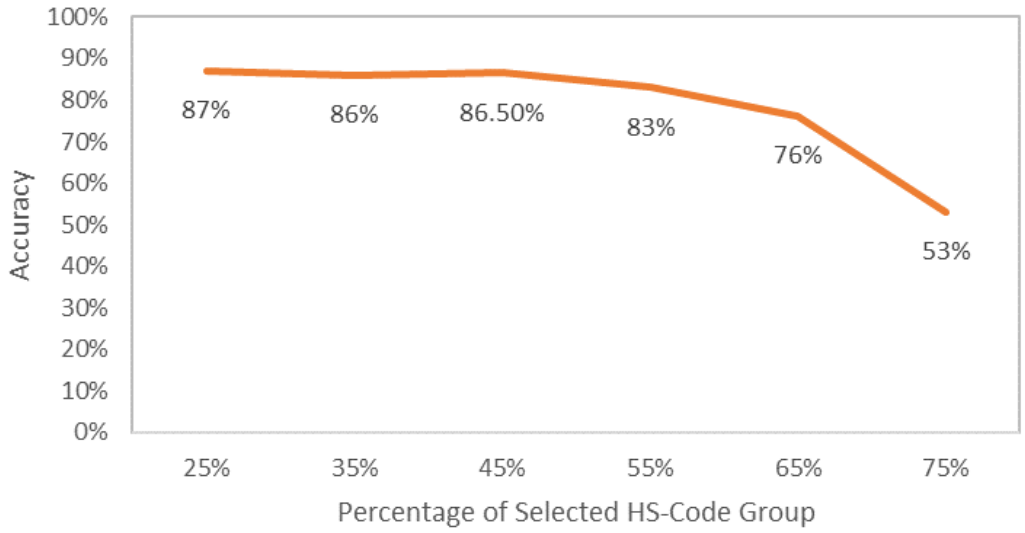


Figure 12: Performance Evaluation when Changing the Percentage of the Selected HS-Code Groups

Finally, to investigate the relationship between the k-nearest neighbor and the accuracy of the proposed approach with the percentage of the selected HS-Code groups is set to 25%, several experiments were executed while varying the value of $k$ in the risk identification step. Figure 13 shows the performance results for these experiments. As illustrated in Figure 13, increasing the value of $k$ to 5 results in increasing the accuracy of the proposed algorithm. Whereas, changing the value of $k$ to over 8 results in reducing the performance of the proposed algorithm. Therefore, the impact of the $k$ value on the definition of point neighborhood needs to be taken into consideration. Hence, increasing the $k$ value will increase the number of points that belong to each point's neighborhood

group. And this will improve the possibility of detecting density-based outliers. However, increasing this value without performing sensitivity analysis will result in reducing the performance since eventually having large neighborhood points reduces the possibility of detecting local outliers.

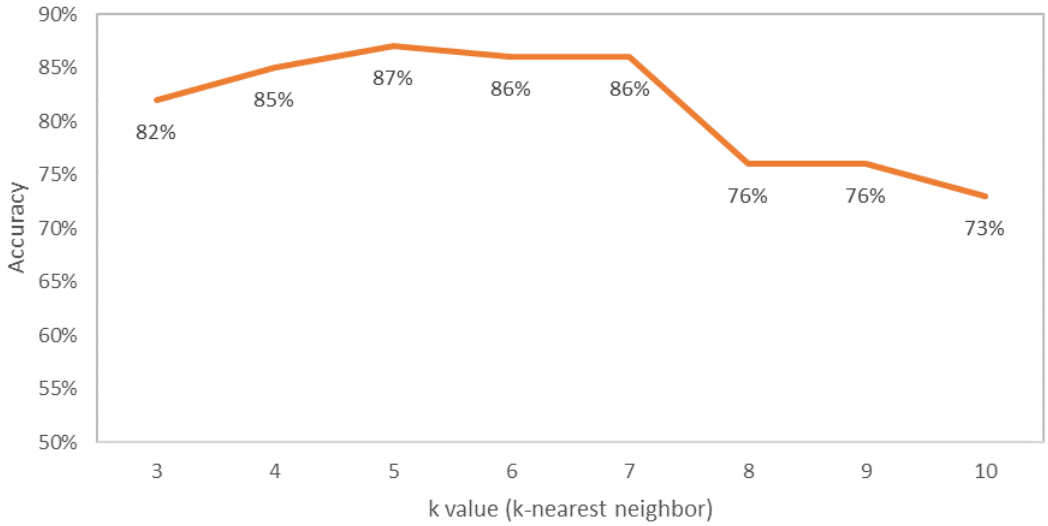


Figure 13: Performance Evaluation when Changing the k Value in k-nearest Neighbor

## 5.4. Conclusion

Trade supply chain is characterized by operating silos. Information is shared among parties using paper documents which makes them vulnerable to manipulation and forgery. The role of customs administration is safeguarding society and local businesses from frauds and criminal activities. As part of the customs core functions, post clearance audit plays an important role in which the past transactions of a company are examined for non-conformance.

This study aims to provide mechanism that helps in the process of detecting abnormal activities in companies trading behaviour. To achieve this objective, a convex-based approach was proposed. This approach works by creating a relationship between shipments

in order to clearly interpret the distances between these shipments. Accordingly, in the proposed approach, the shipments are represented as points in multi-dimensional space. This approach was able to significantly determine shipments that poses security issues. The performance results indicated the potential of adopting it in customs administrations as it achieved an accuracy of 87%. This approach will significantly increase the audit accuracy and provide better resource management thus reducing operating cost for both customs and the entity being audited. This chapter tackled risks associated with shipment data in general. The next chapter looks at these risks from customs valuation perspective to identify value manipulation attempts that fall outside the acceptable range.

# 6. Chapter 6: Customs Valuation Assessment Using Cluster-Based Approach

The role of customs Administrations is vital in protecting the country borders from hazardous people and materials. Customs Administrations encounter challenges in keeping balance to sustain the economic growth through facilitating the movement of cargo and passengers while applying controls to detect customs fraud and other offences and therefore protect the society. The controls that are applied ensure that the trade and travel movement is legitimate. One of the essential controls at the customs administrations is the Customs Valuation.

The Customs Valuation function is responsible for validating and determining the correct declared values in the customs declaration. This includes the process of estimating the values of the declared imported goods. This aims at protecting the revenue and establishing effective collection mechanisms and identifying any fraud activities. Additionally, value manipulation highlights the potential of smuggling goods. Therefore, determining whether the declared value for the goods is correct has an important impact on protecting the local society and economy.

The customs valuation process is triggered whenever there is a doubt on the declared value of goods. This can be identified through the established mechanisms based on the pricelist that the customs administrations maintain along with the tariff and the origin information. Upon the trigger of the valuation process, the valuation officers verify the declared value using any of the following six methods; namely; Transaction Value, Transaction Value of Identical Goods, Transaction Value of Similar Goods, Deductive

Method of Valuation, Computed Method of Valuation, and Derivative Method of Valuation. The transaction value method is the most commonly used method among other methods. The transaction value method takes into consideration the showing cost provided in the invoice in addition to any proof of sale to verify the authenticity of the provided cost (Rosenow & O'Shea 2010).

The customs valuation process also handles the follow-ups related to the customs valuation developments regarding the implementation of the World Trade Organization (WTO) Valuation Agreement. This agreement aims for establishing a comprehensive system for determining and validating the customs value in a fair and neutral manner (WCO 2019).

The process of validating the goods value is normally performed manually and is dependent on the valuation officer experience. It also requires tremendous human efforts to determine the correct value. The result of this process is either there are issues found with respect to the declared value or there are no issues found. The issues are usually related to fraud activities, human errors while submitting the goods value, legitimate trade agreements, or seasonal trades that cause variations in the values. The importance of this process in protecting the local society underlines the demand of establishing accurate mechanisms that can be used to verify the declared goods value.

In this chapter, a cluster-based approach that aims to identify whether the declared goods value raises any suspicious behaviour is proposed. The proposed approach consists of two stages. In the first stage, the investigation takes into account all declared shipments that share the same goods HS-Code. In this stage, the shipments are represented as points

in multidimensional clusters. Each cluster consists of customs declarations that share the same HS-Code. Figure 14 illustrates the representation in multidimensional space. In this representation, the price for the new shipment can be analysed by using statistical measures that capture the locations of the new shipment inside the cluster. This stage labels the shipment as either normal or outlier. When the new shipment is labelled as normal, the status of this shipment will be confirmed in terms of values. The outlier shipments will be sent to the second stage for further analysis. In the second stage, an outlier detection mechanism is employed for the clustered points to determine whether the new shipment can be considered as an outlier or not. In this case, the results of the second stage override the results of the first stage. This mechanism is adopted to identify whether a new shipment can be considered as an outlier. It is required to have an agreement between both stages. Both of these stages take into consideration the distribution of the points to determine whether a shipment should be considered as an outlier or not. However, the first stage validates based on distance measures while the second stage validates based on density.

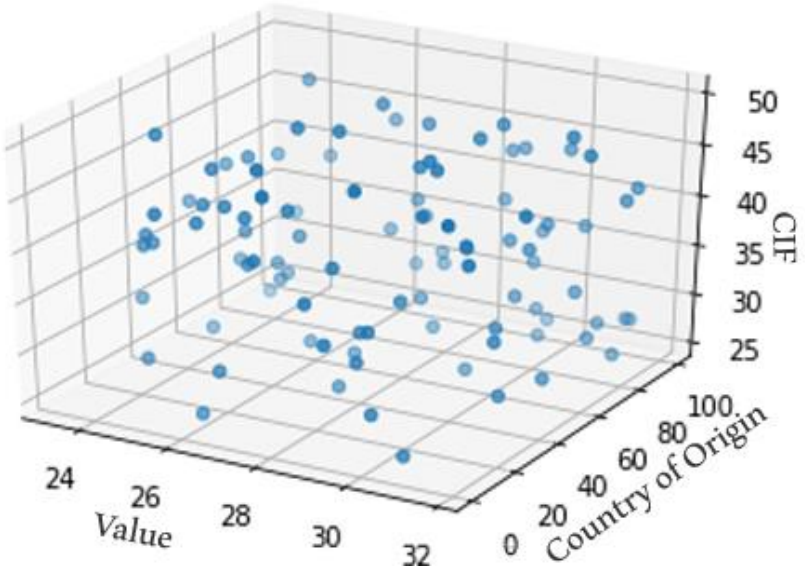


Figure 14: Multi-dimensional Space Representation Example

In several application domains, the use of outlier detection mechanism to detect the abnormal activities have been studied extensively as discussed in chapter 2. In this line, this study proposes multi-level outlier detection approach to determine any value manipulation from customs perspective. The main components of the proposed approach combine the LOF algorithm with the distance-based approach. This combination is selected to capture the differences between the transactions in terms of density and distance.

This chapter comprises four sections. The next section is concerned with the methodology used for this research. The second section illustrates the proposed algorithm and the different stages that are adopted. The third section discusses the various experimental settings and presents the research findings. Finally, the fourth section provides brief summary of the proposed work and the research findings.

## 6.1. Methodology for Customs Valuation Assessment

The problem of ensuring the accuracy of the declared goods is a challenging problem. This study aims to study how clustering techniques can be used to detect abnormal declared goods values. This section illustrates the proposed cluster-based approach that is performed on two stages of analysis namely; distance-based and density-based analysis. These stages aim to determine whether any manipulation has occurred in terms of the goods declared value.

There are several methodologies that are used to perform data mining tasks such as KDD (Fayyad, Piatetsky-Shapiro & Smyth 1996; Siyam, Alqaryouti & Abdallah 2020; Alkashri et al. 2021), SEMMA (Azevedo & Santos 2008a) and CRISP-DM (Pete et al.

2000; Altaheri & Shaalan 2020). These methodologies have common objectives to uncover hidden knowledge and gain insights.

In this work, the CRISP-DM is adopted. CRISP-DM is neutral in terms of industry, technology and application. It is one of the most popular data mining methodologies that is considered as a data mining *de facto* standard for developing knowledge discovery projects (Mariscal, Marbán & Fernández 2010).

As illustrated in Figure 15, The CRISP-DM methodology comprises six stages. The first stage entails obtaining clear understanding of business requirements, challenges and opportunities. The second stage deals with studying and comprehending the collected data. The third stage goes into preparing the data in order to provide a cleansed and harmonized dataset. The fourth stage takes the prepared data and starts the modelling exercise represented in the proposed cluster-based approach. The outcome of the modelling exercise is evaluated in the fifth stage. Finally, in the sixth stage, the model is packaged for deployment to production to handle real-time data.

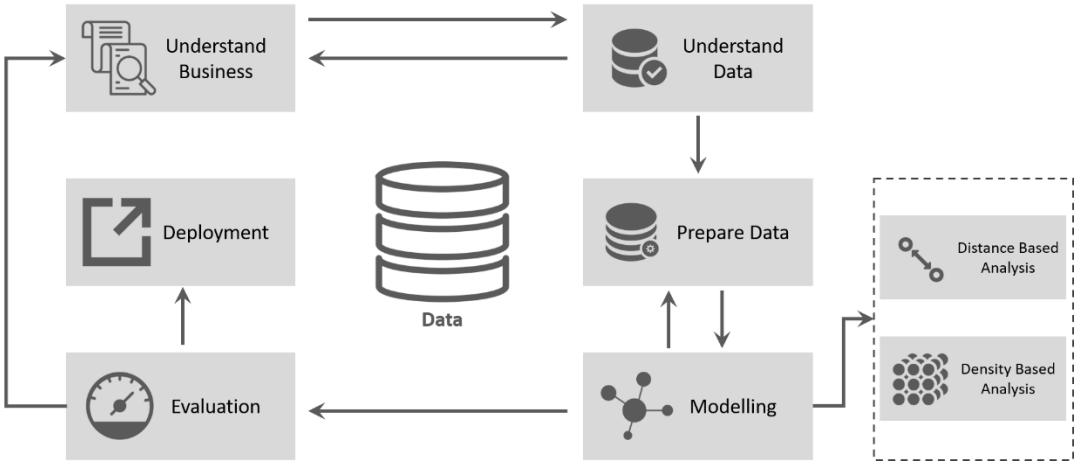


Figure 15: Methodology for Customs Valuation Assessment

The data that is used in this work is obtained from Dubai Customs; a government entity in Dubai. The dataset represents the shipments declarations information. The dataset comprises 500,000 shipment declaration records. These records were performed in the year of 2018 and labelled to identify goods value manipulations. This research captures the corresponding COO, Commodity Type and CIF value (Cost, Insurance-Freight Forwarder) features from declaration information that have the most significant impact on the goods declared value. The following section discusses the details of the proposed cluster-based with the various experiment settings to capture the performance of this approach.

## 6.2. Cluster-based Approach

The proposed approach considers each goods type as an isolated cluster. This isolation aims to simplify the analysis process since the attention is to eventually investigate the goods values that share the same commodity type. The proposed approach analyses the shipment in terms of behaviours and patterns. Behaviour analysis is performed during the first stage by taking into consideration the *Euclidian* distance between the points. On the other hand, pattern analysis is performed by analysing the distribution of the points inside each cluster. Once the first stage completes its execution, the new point is classified into either normal or abnormal. If a shipment is classified as abnormal, the first stage has concluded that based on the locality of the new shipment in its cluster, this shipment has abnormal features. This will trigger the second stage. If the first stage tags a shipment as normal, this result will be confirmed and the second stage will not be triggered. In both of these stages, the following features are used to represent the shipments: HS-Code, Commodity Value (Cost), COO, Declared CIF Value. HS-Code will be used for clustering purposes where other three attributes have a clear relationship that can be used to determine

if a value manipulation has occurred. For instance, there a strong relationship between the COO and the other two values Commodity value and CIF value. COO is considered as significant indicator about the commodity value and CIF value.

Once the clusters are constructed, the distance-based analysis starts processing the entered goods value. During this stage, a *similarity* value for the new entered goods value will be calculated. This value determines whether the new goods value can be considered as normal or not. The calculation of this value takes into consideration the distribution of the points inside the clusters and the locality of the new price value inside the cluster. If the calculated value denotes a normal price, the status of the new entered goods value will be confirmed as normal. On the other hand, when the similarity value highlights that the entered goods value is abnormal, the density-based analysis stage will be triggered to confirm the status of the new price. The density-based analysis stage examines the new goods value against all available historical approved declarations that have been processed.

The core of the density-based analysis is to use an outlier detection mechanism to either confirm the abnormality of the entered goods value or not. Using this mechanism, the status of the new goods value will be confirmed as *normal* when the density-based analysis determines an active goods value pattern.

### 6.2.1. The Distance-based Analysis Stage

In this stage, each goods type that shares the same HS-Code is represented as a single cluster (partition). The objective of this step is to determine whether the declared goods price can be considered normal. To achieve this objective, this step takes into consideration the distribution of the processed declaration values that are represented as points in each

cluster. This step uses a distance-based mechanism to detect the abnormality of the new declaration entered value.

For each cluster, this stage starts by determining the cluster centroid, which is defined as the closest point to all other points within the cluster. The distance between any point and the cluster centroid can be considered as a reasonable indicator whether the new point (goods value) can be considered normal. However, the distance between the same cluster points must be also considered in the process of determining the normality of any given point.

Function 5 illustrates the process of this stage. Once the clusters representation is constructed. The distance (*dc*) between the new point (goods price), and the cluster centroid is calculated. This distance alone represents how far the new point from the rest of the cluster points. Beside this distance, it is also required to determine the distance between the new point and its neighbouring points. This can be calculated by determining the average distance between the new point and all other points inside the cluster (*dn*). These two distances values are used to calculate the similarity values as follows:

$$Similarity = \left(\frac{dc}{mc}\right) \times (\frac{dn}{md})$$

Where *md* represents the maximum average distances between the cluster points, and *mc* represents the maximum average distances between the cluster's points and the cluster centroid. For any new shipment point, if the similarity value of this point is less than one (<1), the declared goods value point will be considered as normal. In case that the similarity value is higher than one (>1), the declared goods value will be considered as an abnormal and this will trigger the density-based analysis stage.

Function 5: The Proposed Cluster-based Algorithm

```
    DistanceBasedAnalysisStage (p)
    Input p: New Shipment Point Information
    Output status
1   c = LocatePointCluster(p)
2   dc = FindDistancetoCentroid(p, c)
3   dn = AvgDistancetoNeighbors(p, c)
4   md = 0
5   mc = 0
6   For each point (j) in c
7       If (FindDistancetoCentroid(j, c) > mc) then
8           mc = FindDistancetoCentroid(j, c)
9       If (AvgDistancetoNeighbors(j, c) > md) then
10          md = AvgDistancetoNeighbors(j, c)
11  End For
12  s = (dc/mc) × (dn/md)
13  If (s <= 1) then
14      Return "Normal"
15  Else
16      Return DensityBasedAnalysisStage(p)
```

Figure 16 illustrates an example about the distance-based analysis stage. From the figure, the point that is highlighted in orange colour is more likely to be considered as an outlier compared to the point highlighted in green colour.

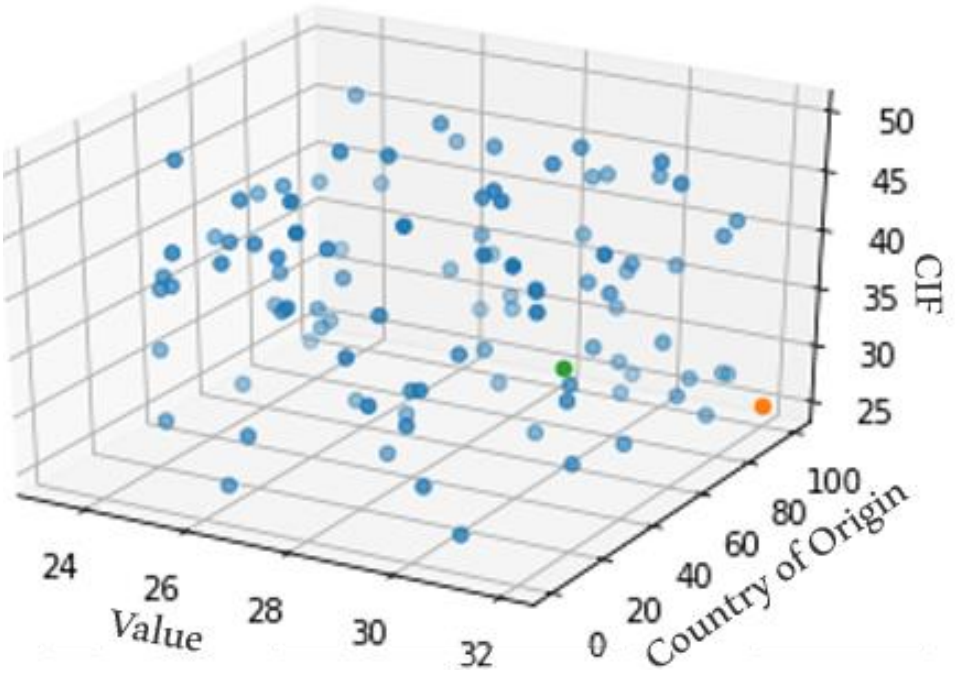


Figure 16: Example of Distance-based Analysis Stage

### 6.2.2. The Density-based Analysis Stage

Once the distance-based analysis stage has labelled a shipment as an outlier, the density-based analysis stage will be triggered to examine the new shipment in terms of density. The distance-based analysis uses distance measures to determine whether any shipment is normal or an outlier. Accordingly, in order to confirm the outlier label for any shipment in the density-based analysis stage, the density of the cluster is examined beside the new shipment point location. In this stage, the LOF algorithm is employed. The LOF algorithm determines whether any given point is an outlier based on the distribution of the points and the density of the points across the entire cluster. For each point, the LOF algorithm returns an outlier factor value greater than or equal to zero (>=0). In an instance when the outlier factor value for a new given point is less than one (<1), this point will be considered as an inlier (normal). However, the new point will be considered as an outlier (abnormal) when the outlier factor value is greater than one (>1).

In this stage, once the new shipment cluster is identified, all of the cluster points in addition to the new points will be used as inputs for the LOF algorithm. This algorithm will analyse the shipment information in order to determine whether any pattern that validates the declared value is present in the historical data.

## 6.3. Experiments and Discussion of Results

To evaluate the performance of the proposed approach, various experiments with different settings have been conducted. The dataset used in this work represents 500,000 declarations applications obtained from Dubai Customs for the year of 2018. These declarations are labelled to normal and abnormal with respect to the declared goods value. These declarations are pre-processed to eliminate all records that have duty exemptions. In addition, the least 10% of the clusters with minimum number of declarations are eliminated. The normal declarations are considered as the historical available shipments' declarations, and these declarations will be used to construct the clusters. The testing sample was prepared formulating random 10 groups, where each groups includes 200 items divided into 100 abnormal items and 100 normal items. Through these experiments, the following experiment parameters are used:

- **Number of declarations**: the number of used declaration testing samples groups for evaluation purposes.
- **The size of used clusters**: the percentage of HS-Code clusters used for evaluation purposes.

To evaluate the performance for each experiment settings, the accuracy and the precision are calculated. The accuracy is the ratio of the correctly predicted shipments value manipulation status to overall predictions. While, the precision is the ratio of

correctly predicted shipments with value manipulation to the total shipment with value manipulation.

Various experiments were conducted with varying the size of the used clusters using the full testing sample (10 groups). Figure 17 shows the accuracy and precision of the presented approach against the size of the used clusters. In these experiments, the clusters are grouped in ascending order based on the number of available declarations for each cluster. Accordingly, the percentage of used cluster is varied with highest number of points as recorded in the experiment. According to the results, reducing this percentage results in reducing the gap between the accuracy and the precision. In addition, reducing this percentage will have significant impact on the overall performance of the proposed approach. This is due to several factors; reducing the percentage of the selected clusters results in selecting the clusters with the largest sizes. Having clusters with large sizes improve the classification performance of the proposed approach. The reason is that the size of each cluster plays a major role in determining the authenticity of the obtained result. The size plays a major role in the first stage since it is a distance-based mechanism and impacts the result of the second stage since it will influence the density of the resulted clusters. This explains the seen performance and clarifies the gap between the accuracy and the precision.

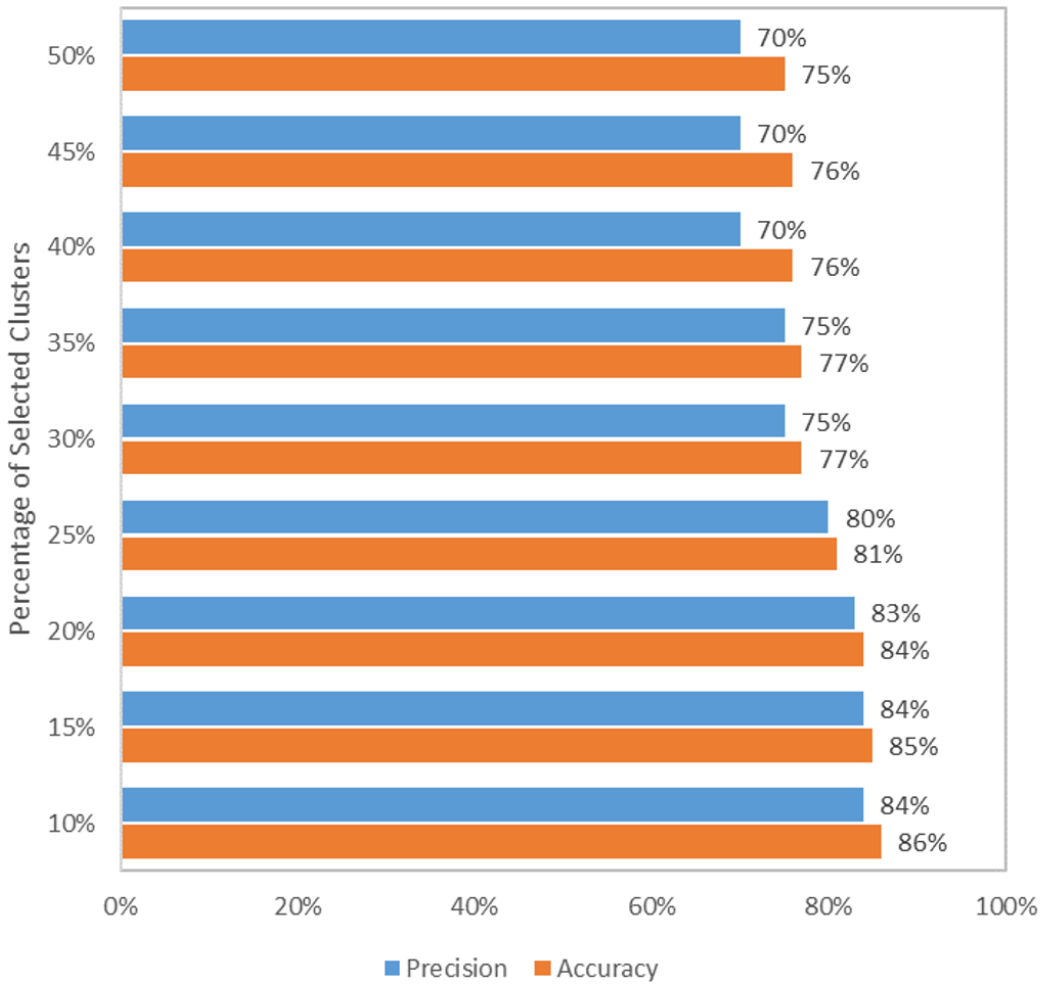


Figure 17: Performance Results against the Size of the Used Clusters

Figure 18 illustrates the accuracy and precision of the proposed approach against the number of used testing sample groups using the largest 10% of the clusters. The figure shows that increasing the number of used testing samples results in improving the overall performance of the proposed algorithm. In addition, reducing the number of used testing samples increases the gap between the accuracy and the precision.

In other words, the number of declarations affects the performance of the proposed approach. For instance, increasing the number of used declarations stabilizes the performance of the proposed approach since this will increase the likelihood to capture

several experimental scenarios. This is evident in the results since increasing the number of used declarations to over eight groups have no clear influence on the behaviour.

In addition, as shown in Figure 18 at lower number of groups with less than or equal to three, the performance of the proposed algorithm was not stable. Regarding the gap between the accuracy and the precision, performing experiments using low number of declarations will increase the probability of inaccurate classifications. This explains the reason behind the seen gap between these two metrics.

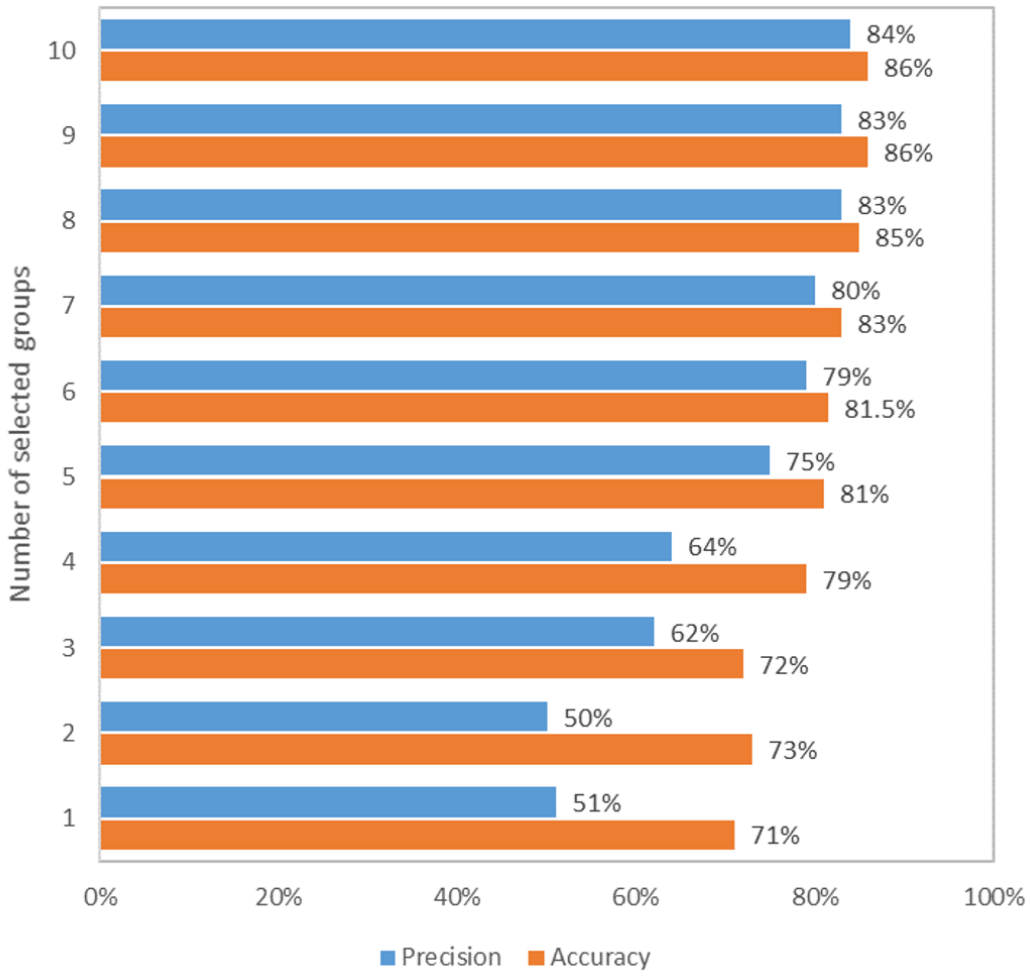


Figure 18: Performance Results against the Number of Used Declarations

## 6.4. Conclusion

The customs valuation is a critical function in any customs administration that serves the objectives of society protection and trade facilitation. Furthermore, it protects the revenue by identifying and preventing the leakage caused due to fraud activities by providing undervalued goods declarations. In this chapter, a cluster-based approach to detect customs value manipulation is proposed. The proposed approach consists of two stages; the distance-based analysis stage and the density-based analysis stage. These two stages work to capture the abnormality of the shipments in terms of declared goods value. The first stage focuses in its analysis on the distances between the points that represent the shipments. While, the second stage focuses on the density aspects of these points to confirm result of the first stage. The results achieved an accuracy of 86%, which shows the importance of combining both behavioural and pattern analysis to process shipments declaration. This also shows the potential of the proposed algorithm to simplify and automate the process of detecting the value manipulation as part of the declaration process at customs domain. The next chapter summarizes the thesis and its key findings as well as discusses the theoretical and practical implications, limitations of this study and recommendations for further study.

# 7. Chapter 7: Conclusion

Cross-border trade is a critical vehicle for economic stability and revenue generation. This fact is tainted by many old and lengthy formalities along with strict regulations. Nations depend on local customs organizations for trade facilitation and cross-border security. Being competitive in today's economy mandates creating smoother trade flow by adopting effective and efficient procedures. As the volume of trade increases year after year and demands for goods are on the rise, the customs role in the cross-border trade supply chain becomes more crucial. This work introduces a new model of operation that leverages the power of advances in technology to achieve the desired goals.

The trade supply chain consists of many parties spanning multiple industries each of which follows industry specific standards and formalities. All would benefit from gaining visibility to prior and post actions to raise the level of their efficiency and eliminate gaps. The natural traditional way is to establish direct integration among the various participants. Since data will be sourced from multiple sources, this will require creating multiple integration layers. However, this method has proven to be costly and hard to maintain. At the same time, it will not provide the assurance that data will not be wrongly manipulated. The difficulty stems from dealing with multiple data models and technologies. Therefore, it becomes quite essential to work on establishing a full supply chain records system in which data is recorded once after consensus is reached among all participants. This type of system will raise the level of integrity and will only require one integration point.

Understanding these requirements points toward a blockchain-based solution offers the needed features that ensure immutability, consensus, traceability and integrity. Blockchain

technology has the ability to satisfy the needs of stakeholders to prevent any value manipulation while eliminating manual activities and human intervention.

Another key challenge for the trade supply chain is fraud detection. The risk of fraud can occur at any stage within the trade supply chain. Discovery of the risk may occur at a later stage which may cause the risk to be intensified and have ripple effects. The financial losses from these risks can lead to economic decay. Trade supply chain participants particularly customs authorities aspire to identify risks early on. Early warnings give authorities the ability to deter criminal individuals and networks. Current practices are not all effective as they mainly rely on human perception and lack scientific proofs.

By combining the capabilities of blockchain technology to establish a harmonized network and supporting that with science to identify risky behaviours, the research is able to assist trade supply chain participants achieve their goals and objectives in a methodical approach. The long-desired balance between trade facilitation and border security is easily achievable with the proposed framework.

## 7.1. Summary of the Study

Cross-border trade supply chain parties for quite some time have had to deal with a mounting number of obstacles and challenges. Some of these are strategic in nature and result in a multifaceted financial loss. However, the future seems to carry some promising opportunities for overcoming a lot of these traditional barriers. Advances in technology are leading the way toward reform and modernization in various sectors. This work addresses issues associated with ecommerce activities from the perspective of customs organizations.

The proposed framework is based on blockchain technology complemented by artificial intelligence to streamline procedures and provide a cross-border trade community with novel capabilities. This study started with blockchain technology as a base to connect all concerned parties to a single consensus-based ledger system. This solution will enable visibility for the full cross-border trade lifecycle which in turn will allow the respective parties to chart the proper course of action.

Additionally, the proposed work attempts to eliminate many of the decisions and actions that are driven by human perception and shift to an automated mechanism based on scientific proofs and techniques. One of the critical areas that is targeted in the study is the risk assessment techniques and transform these strategic functions into an algorithmic formula for the detection of fraudulent activities.

Today, supply chain parties are resorting to point-to-point integration, which is heavy in terms of implementation, administration and maintenance cost. The proposed framework will drastically reduce this continuous cost stream and only require a onetime initial investment. The adoption of this framework will have major benefits for the trade supply chain; attributed to having accurate advance information sharing which allows parties to be proactive and provides greater insight into trade activities. This will eliminate or reduce the number of checks and balances required today, which will result in the reduction of unnecessary resource allocation and elimination of certain business processes. In essence, trade parties will have stronger capabilities that help them make better-informed fact-based decisions. The result of this work will significantly reduce operating cost while increasing the efficiency and effectiveness that drive economic value.

## 7.2. Key Findings

The objective of this work was to look into the technology capabilities that could revolutionize this global economic pillar (cross-border trade) and ensure it creates shared value for all participants. The findings are presented by responding to the research questions in this section.

### 7.2.1. RQ1: What factors will influence the adoption of Blockchain technology in ecommerce platforms?

The volume of ecommerce transactions continues to increase and create demand for the fast delivery of goods and products. With today's model of ecommerce domain, this seems extremely hard to accomplish given the weak visibility and authenticity practices in the various stages. As discussed thoroughly in chapter 2, there were several factors that influence the adoption of blockchain technology in the ecommerce domain. These factors include providing visibility across concerned parties while ensuring adequate security to improve the speed of trade activities in a flexible and scalable manner.

Blockchain technology can contribute to providing cost-effective solutions for different types of industries (Petersson and Baur, 2018). Blockchain applications demonstrate a promising opportunity for improving processes and practices in the global trade supply chain and ecommerce fields. The trade supply chains are facing major fundamental challenges namely; data visibility, process optimization and demand management (Godbole n.d.).

According to Belu (2019), the application of blockchain technology in the field of international trade can contribute to reducing costs, providing more transparency and

security and easier control and flow of goods among others. This highlights blockchain as a promising technology in trade supply chain that can improve global trade. This technology promotes unique feature such as immutability, and trust and transparency, thereby transforming the international trade sector.

Blockchain in international trade can be applied in supply chain and logistics in order to improve global trade practices. Blockchain application is a new way of organizing and controlling trade value chain and import and export operations (Belu 2019). In addition, blockchain technology also contributes to improving the global trade supply chain by making it more transparent. According to Norberg (2019, p. 4), "Blockchain will decrease the costs of trade, which will empower globalization, trade, and optimize the global value chains that the ICT revolution has made possible". As explained by the same author, the application of blockchain in trade supply chain allows for improving agreements that are already in place and supporting the underlying characteristics of modern trade agreements. The adoption of blockchain technology in the global trade supply chain enables new levels of trust, transparency and accountability towards revolutionizing and reinventing international trade. Similarly, Cole, Stevenson and Aitken (2019) asserted that the adoption of blockchain technology in supply chain management perspective presents opportunities for enhancing product safety and security, improving quality management, reducing counterfeiting and improving the sustainability of supply chain management. This suggests that embracing Blockchain in supply chain management can help transform the global nature of supply chains.

In ecommerce, blockchain technology presents promising opportunities towards improving scalability and addressing performance issues (Lim et al. 2019). As the

scalability of ecommerce continues to increase, blockchain technology contributes to optimizing business processes, reducing operating costs and improving efficiency of collaboration and interaction (Zhu & Wang 2019). These features state that the application of blockchain technologies can contribute to improving ecommerce.

Blockchain network technique is being integrated in some ecommerce models to enhance consumer to consumer experience (Shorman, Allaymounq & Hamid 2019). This means that Blockchain technology can contribute to enhancing customers' ecommerce experience through the establishment of trust and credibility. According to Sheikh *et al.* (2019, p. 777), blockchain technology provides support to ecommerce platforms to store highly secured data and provided high performance transactional processing. This is a major concern in immutable records for web payments and online processing for orders.

This means that one of the major advantages of Blockchain application in ecommerce is promoting secure transactions particularly during online payments and purchases. As such, Blockchain technology presents opportunities for revolutionizing the traditional ecommerce industry.

### 7.2.2. RQ2: How can blockchain enhance trade facilitation and compliance in ecommerce trade domain?

As discussed in Chapter 4, a blockchain-based framework was proposed. This framework aimed at streamlining ecommerce trading processes using blockchain technology while ensuring the needs of all concerned parties. The proposed framework will revolutionize the way trade supply chain is handled and create a shift from working in reactive limited visibility to proactive full visibility mode through facilitating access to

advance information sharing. This framework will enhance the governance of trade activities and establish a trust level among stakeholders. This framework was developed by adopting the SDLC methodology using the iterative approach.

Many researches in literature investigated the security and scalability challenges in trade supply chain in general and ecommerce in particular. For instance, Feng Tian (2017) adopted blockchain technology in food safety in international trade. In the studies conducted by Toyoda et al. (2017) and Alzahrani and Bulusu (2018), the authors presented a blockchain solution to address the counterfeiting challenges. In addition, Lu and Xu (2017) developed a blockchain-based solution to provide tamper-proof data traceability and regulation validation management system. While, Kamath (2018) examined the idea of tree-to-shelf traceability. In another study, Chen et al. (2017) introduced special and conservative hierarchy for the participating nodes and their role in the blockchain. Similarly, Xie et al. (2018) proposed a trust framework for ecommerce through the use of blockchain. In another domain, Ying, Jia and Du (2018) presented a use-case study that shows the benefits of using a blockchain ecommerce platform to offer employees flexible benefits. The capabilities of blockchain cover many of the desired techniques needed as part of transforming the trade supply chain; shared ledger capability helps synchronize data among all participating parties, immutability guarantees integrity of data, provenance helps determine origin of goods and products to prevent fraud and establish traceability and smart contracts capability ensures the consensus to establish cross-party workflow. These capabilities give blockchain technology a leading edge over other offerings that do not have these needed capabilities out of the box and will require larger investment.

In this study, the proposed blockchain-based framework aimed at optimizing the process of ecommerce trade from customs perspective, and therefore the customs related processes were targeted. The proposed blockchain framework was evaluated using the SUS which shows overall high acceptability levels by all user categories. The proposed solution illustrated how blockchain can cover the factors under consideration, namely integrity, provenance and traceability. This can be complemented with off-chain scientific data mining capabilities for risk and fraud detection to provide the necessary risk detection measures.

This framework introduces many benefits such as reducing the paper dependency and providing a digitized trade platform that guarantees faster to market delivery at reduced cost while ensuring compliance with rules and regulations.

### 7.2.3. RQ3: Can clustering techniques be used to identify risks in shipments for Customs post clearance audit by utilizing the proposed blockchain-based framework?

According to the results discussed in chapter 5, clustering techniques can be used to simplify the post clearance audit in customs. To achieve this objective, a convex-based approach was proposed. This approach works by creating a relationship between shipments in order to clearly interpret the distances between these shipments. Accordingly, in the proposed approach, the shipments were represented as points in multi-dimensional clusters by employing the CRISP-DM methodology. This clustering technique analyses the blockchain transactions of an entity with its peers based on past conduct to determine any behavioural deviations from the norm. Thus, any company related transactions can be analysed to determine abnormal behaviours between this company and similar companies.

The performance results indicated the potential of adopting it in customs administrations as it achieved an accuracy of 87%.

The reported performance highlights the significance of the proposed convex-based approach compared to other outlier detection studies in different domains (Almiani, Chawla & Viglas 2014; Almiani et al. 2018; Badriyah, Rahmaniah & Syarif 2018; Ceronmani Sharmila et al. 2019).

Customs administrations are empowered to assess companies' behaviours and detect outliers using Convex-based clustering approach to establish legitimacy. This approach will significantly enhance customs post clearance audit performance and provide better resource management, thus reducing operating cost for both customs and the entity being audited.

### 7.2.4. RQ4: Can clustering techniques be used to determine risks related to value manipulation in the ecommerce transactions by utilizing the proposed blockchain-based framework?

According to the results discussed in chapter 6, clustering techniques can be used to determine value manipulation in the ecommerce transactions. To achieve this objective, a cluster-based approach to detect customs value manipulation was proposed. The proposed approach consists of two stages. The analysis of the first stage focuses on the distances between the points that represent the shipments, while in the second stage it focuses on the density aspects of these points to confirm result of the first stage. This approach will enhance the operational security and governance of ecommerce transactions and allow parties to focus on risks they are concerned with.

This representation simplifies the process by identifying relationships among the various shipment features. Consequently, correlations are established to detect anomalies in declared values. The proposed framework is based on blockchain technology which will enable the use of real-time information to identify value manipulation. Employing the CRISP-DM methodology to represent data in clusters using distance- and density-based techniques allows customs administrations to analyse data instantly and determine if there is any manipulation in the declared goods values. The results achieved an accuracy of 86% which shows the importance of combining both behavioural and pattern analysis to process shipments declaration. In addition, the reported performance highlights the significance of the proposed approach compared to other studies in different domains (Almiani, Chawla & Viglas 2014; Almiani et al. 2018; Badriyah, Rahmaniah & Syarif 2018; Ceronmani Sharmila et al. 2019).

The results achieved through the proposed approach prove that employing clustering techniques using outlier detection algorithms to identify and eliminate risks of value manipulation as well as it will significantly transform trade supply chain.

## 7.3. Implications

At first glance, the trade supply chain gives the impression that it is well-connected and secure. However, digging deeper, the fragmented reality that lies at its core is discovered. Fragmentation has led to a painstaking experience across the various stages. Furthermore, a disconnect is noticed between strategic trade supply chain partners which further hampers efforts for harmonization. Despite attempts to facilitate trade activities in the 21$^{st}$ century, there was never a serious effort to look at challenges from a holistic and comprehensive point of view.

What is being proposed in this work is a framework that leverages the technical capabilities of advance technologies and establish synergy among participants. Key features include providing full visibility at every step of the trade flow. At the same time, it will eliminate integrity related issues coupled with secure transactions throughout the trade lifecycle. As a principle, protecting existing investments is key for increased adoption complimented by an orchestration mechanism that can deliver a holistic, efficient and effective approach in line with ISO 9241– 11 requirements to tackle the complexity of the various formalities and handle the scale of the trade global challenge.

The proposed framework covers the aforementioned needs of ecommerce cross-border trade activities in an economically viable approach. The solution provides the required scalability to support the various needs of the trade supply chain by providing streamlined procedures and delivering the needed traceability that takes competitive advantage to a whole new level. This is important to improve the cross-border ecommerce efficiency and create immutable audit trail transactions that improves transparency for all parties, reduce transactional cost and decrease interaction and clearance times.

One of the scenarios in which the world has experienced and led to some disruption in cross-border ecommerce is the COVID-19 pandemic. The dependencies on hard hit industries resulted in the near collapse of ecommerce activities worldwide. Additionally, consumer fears for lack of traceability, transparency and provenance reduced the number of activities even further. The lack of preparedness of organizations and governments to handle business as usual during emergencies and crisis conditions has led to supply chain closures. This could have been prevented by guaranteeing integrity, traceability, transparency for products and supplies through the careful adoption of technologies such

as blockchain. The proposed framework shows how ecommerce would greatly reduce business disruptions and raise trust in cross borders trade activities and maintain economics sustainability.

One of the success factors of any blockchain implementation is getting all concerned parties to work collaboratively to address present and future needs early on and celebrate small successes while working out the disputes.

Previous methods and techniques used by industry practitioners were limited in capabilities and had lower accuracy. The proposed framework provides a holistic trade supply chain management model and has the advantage of increasing the integrity of activities using full data driven approach to identify fact-based implications a human will not think to connect and look for.

## 7.4. Limitations

There are many existing implementations, consortiums and networks that are using blockchain to handle various trade related activities. However, each is having their own blockchain fabric which lacks cross-fabric integration. The industry as a whole has to work on developing this capability to enable adopters to have the freedom to work on the fabric of their choice while maintaining proper integration. Major blockchain vendors have addressed industry specific needs however, some of these players need to work on developing Application Program Interfaces (APIs) or standards to establish interoperability between the different fabrics. These APIs and standards must guarantee and maintain data integrity and auditability. Until these challenges are addressed, early adopters can still benefit from working within specific consortium and enjoy early competitive advantage.

Another challenge relates to regulatory risks. Regulations vary from industry to industry, country to country and sometimes state or city level laws and regulations. A typical blockchain implementation will span parties in various geopolitical locations. Addressing the participating countries regulatory requirements will be laborious and complex. This will require parties to agree on standards to address the binding applicable laws and regulations and the governance of legal considerations including restrictions and dispute resolutions while ensuring privacy and means digital identity.

Until the maturity of the blockchain technology reaches higher level, concerns withstand with the throughput of each blockchain fabric and the mechanisms used to interface with legacy systems.

Data quality on the blockchain is a concern since multiple parties injecting the data elements, they are responsible for providing while each uses different quality approach. Harmonizing the quality of data is not only critical for blockchain implementations but also for ensuring the efficiency of the subsequent data mining and risk detection algorithms as well as audit activities.

## 7.5. Scope for Further Study

Future work to include performing a pilot study to investigate the performance of the presented framework in chapter 4. The focus is mainly on determining the cost and time saving as well as the traceability capabilities of the proposed framework. In addition, the framework can be extended to cover other industries involved in the trade supply chain.

The proposed off-chain framework components are currently based on supervised learning techniques and would benefit from shifting to unsupervised learning techniques

that could explore potential associations due to either discovery of weak or indirect relationships between different transactions. Future work should test various unsupervised learning techniques to produce accurate predictions and allow for a larger set of parameters to be tested leading to more dynamic assessment techniques. Later, by using both techniques with weight and dependencies as built-in features to provide scores for each trader, a hybrid solution can be built. This score helps consumers accurately and based on actual figures establish the trustworthiness of suppliers or exporters.

In addition, investigating the use of other mechanisms in shaping the convex space for the nodes such as LOF may significantly improve the performance of the proposed approach. Additionally, there is potential to expand the proposed approach by providing a hierarchal scheme that covers the situation where the number of shipments for a specific HS-Code is not enough to perform an accurate analysis. This can be established by running an additional level where the nodes represent groups in the HS-Code header.

In chapter 6, LOF algorithm has been used to capture the relationship between the points in terms of density. Exploring the use of different outlier detection algorithms may have significant improvement on the performance of the proposed algorithm. In addition, there is an area of research to expand the scope of the proposed approach in this work to also cover the problem of goods smuggling. This problem has a direct relationship with the value manipulation mechanism since traders who try to smuggle undeclared goods will most likely to falsify the declared goods values.

Trust and reputation are the most substantial factors to be regarded in any ecommerce platform. Reputation-based models can help the buyers in various ecommerce platforms to

choose the best seller through the seller's reputation score. The reputation score of the seller is calculated based on the buyer's ratings on the sellers' services and products.

Reputation-based systems help users decide which sellers can be trusted and which cannot. This gives both buyers and sellers some insurance to complete transactions online. One example of determining the attribution of user's reputation is eBay's reputation system, called Feedback Forum. Feedback Forum constructs a reliable history of the seller based on the feedback received by the buyers. In some cases, buyers may pay a premium price for the sake of comfort and security of high-quality services (Resnick et al. 2000).

Different reputation-based models exist in ecommerce platforms such as the individual model, the system model and the reputation model. In the individual model, the buyers themselves choose the seller from whom they will do a transaction. In the system model, a seller is selected depending on his trustworthiness. In the case of the reputation model, the sellers are chosen depending on their high rank by using the trust score (Zhang, Cui & Wang 2014).

The success of ecommerce applications is mainly accomplished by the users' interaction. Reputation score allows the potential buyers to identify the service quality of a seller. Most of the reputation systems allow users to write comments and express their opinions on the services, whether positively or negatively. This results in the generation of abundant feedback comments, facilitating the generation of an actual trusted reputation score for a seller (Hijikata et al. 2007). However, users' feedback and comments are usually unstructured, making their analysis a challenging task.

However, reputation-based systems are faced with some challenges such as prompting, distributing, and aggregating feedback. For instance, people may not bother to provide feedback since no incentives are provided for their time filling out a form. Furthermore, it is difficult to ensure legitimate reports since there may be cases of blackmailing or group collaboration ratings to artificially inflate reputation.

## 7.6. Concluding Note

Trade supply chain participants are in dire need to raise the level of trust and reduce risks throughout the full trade cycle. Data manipulation and falsification is a prevalent act hindering efforts of facilitation and protection. Silo attempts to mitigate these risks have not produced the desired results. A more encompassing solution is required to cover all participants in a standardized fashion. The proposed framework covers documentary fraud and risk assessment centred around a blockchain environment coupled with a scientific approach to risk identification. This will lead to a more coordinated and unified mechanism that injects integrity, traceability and transparency in the trade life cycle while reducing the cost and time of trade movement. Adopting this framework will ensure faster delivery of goods to consumers which in turn will lead to higher consumer satisfaction and steady increase of volume for ecommerce trade.

# References

Abualigah, L. M., Khader, A. T. & Hanandeh, E. S. (2018). Hybrid clustering analysis using improved krill herd algorithm. *Applied Intelligence*, vol. 48(11), pp. 4047–4071.

Ahmad, R. W., Hasan, H., Jayaraman, R., Salah, K. & Omar, M. (2021). Blockchain applications and architectures for port operations and logistics management. *Research in Transportation Business & Management*. Elsevier Ltd, p. 100620.

Alcarria, R., Bordel, B., Robles, T., Martín, D. & Manso-Callejo, M.-Á. (2018). A Blockchain-Based Authorization System for Trustworthy Resource Monitoring and Trading in Smart Communities. *Sensors*, vol. 18(10), p. 3561.

Alkashri, Z., Alqaryouti, O., Siyam, N. & Shaalan, K. (2021). 'Mining dubai government tweets to analyze citizens' engagement'. , in *Studies in Systems, Decision and Control*. Springer, pp. 615–638.

Almiani, K., Chawla, S. & Viglas, A. (2014). The Effect of Outliers in the Design of Data Gathering Tours. *2014 Sixth International Symposium on Parallel Architectures, Algorithms and Programming*. IEEE, pp. 209–214.

Almiani, K., Viglas, A., Lee, Y. & Abrishambaf, R. (2018). Peripheral nodes and their effect in path planning in networks. *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 27(3), pp. 157–170.

Alqaryouti, O. & Shaalan, K. (n.d. a). Customs Valuation Assessment Using Cluster-Based Approach. *SN Computer Science*. Springer.

Alqaryouti, O. & Shaalan, K. (n.d. b). Outlier Detection for Customs Post Clearance Audit Using Convex Space Representation. *Technology in Society*. Elsevier.

Alqaryouti, O. & Shallan, K. (2020). Trade Facilitation Framework for E-commerce Platforms using Blockchain. *International Journal of Business Information Systems*.

Alqaryouti, O., Siyam, N., Alkashri, Z. & Shaalan, K. (2019a). Cryptocurrency usage impact on perceived benefits and users' behaviour. *European, Mediterranean, and Middle Eastern Conference on Information Systems*, pp. 123–136.

Alqaryouti, O., Siyam, N., Alkashri, Z. & Shaalan, K. (2019b). Users' knowledge and motivation on using cryptocurrency. *European, Mediterranean, and Middle Eastern Conference on Information Systems*, pp. 113–122.

Altaheri, F. & Shaalan, K. (2020). Exploring machine learning models to predict harmonized system code. *Lecture Notes in Business Information Processing*. Springer, pp. 291–303.

Alzahrani, N. & Bulusu, N. (2018). Block-Supply Chain. *Proceedings of the 1st Workshop on Cryptocurrencies and Blockchains for Distributed Systems - CryBlock'18*, (June), pp. 30–35.

Azevedo, A. & Santos, M. F. (2008a). KDD, semma and CRISP-DM: A parallel overview. *MCCSIS'08 - IADIS Multi Conference on Computer Science and Information Systems; Proceedings of Informatics 2008 and Data Mining 2008*, (January 2008), pp. 182–185.

Azevedo, A. & Santos, M. F. (2008b). KDD , SEMMA AND CRISP-DM : A PARALLEL

OVERVIEW Ana Azevedo and M . F . Santos. *IADIS European Conference Data Mining*, pp. 182–185 [online].Available at: http://recipp.ipp.pt/handle/10400.22/136%0Ahttp://recipp.ipp.pt/bitstream/10400.22/136/3/KDD-CRISP-SEMMA.pdf.

Badriyah, T., Rahmaniah, L. & Syarif, I. (2018). Nearest Neighbour and Statistics Method based for Detecting Fraud in Auto Insurance. *2018 International Conference on Applied Engineering (ICAE)*. IEEE, pp. 1–5.

Bangor, A., Kortum, P. T. & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale Usability Scale. *International Journal of Human–Computer Interaction*, vol. 7318.

Bangor, A., Staff, T., Kortum, P., Miller, J. & Staff, T. (2009). Determining what individual SUS scores mean: adding an adjective rating scale. *Determining what individual SUS scores mean: adding an adjective rating scale*, vol. 4(3), pp. 114–123.

Barber, C. B., Dobkin, D. P. & Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, vol. 22(4), pp. 469–483.

Belu, M. G. (2019). Application of Blockchain in International Trade: An Overview. *Romanian Economic Journal*, vol. XXII(71), pp. 2–16.

Bevan, N. & Carter, J. (2016). Human-Computer Interaction. Theory, Design, Development and Practice. *In International Conference on Human-Computer Interaction (pp. 268-278). Springer, Cham.*, vol. 9731(July), pp. 268–278.

Bevan, N., Carter, J. & Harker, S. (2015). 'ISO 9241-11 revised: what have we learnt about

usability since 1998?' , in *Human-Computer Interaction*, pp. 143–151.

Bhaduri, K., Matthews, B. L. & Giannella, C. R. (2011). Algorithms for speeding up distance-based outlier detection. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11*. New York, New York, USA: ACM Press, p. 859.

Blelloch, G. E., Gu, Y., Shun, J. & Sun, Y. (2020). Randomized Incremental Convex Hull is Highly Parallel. *Annual ACM Symposium on Parallelism in Algorithms and Architectures*, pp. 103–115.

Blossey, G., Eisenhardt, J. & Hahn, G. (2019). Blockchain Technology in Supply Chain Management: An Application Perspective. *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

Bocek, T., Rodrigues, B. B., Strasser, T. & Stiller, B. (2017). Blockchains everywhere - a use-case of blockchains in the pharma supply-chain. *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*. IEEE, pp. 772–777.

Brooke, J. (1996). *SUS-A quick and dirty usability scale*. *Usability Evaluation In Industry*. Edited by P. W. Jordan, B. Lyall, and T. Ian. McClelland, Bernard Weerdmeester.

Brooke, J. (2013). IUPS: a retrospective. *Journal of Usability Studies*, vol. 8(2), pp. 29–40.

Casino, F., Dasaklis, T. K. & Patsakis, C. (2019). A systematic literature review of blockchain-based applications: Current status, classification and open issues. *Telematics and Informatics*.

Ceronmani Sharmila, V., R., K. K., R., S., D., S. & R., H. (2019). Credit Card Fraud Detection Using Anomaly Techniques. *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*. IEEE, pp. 1–6.

Chang, S. E. & Chen, Y. (2020). When blockchain meets supply chain: A systematic literature review on current development and potential applications. *IEEE Access*.

Chang, S. E., Luo, H. L. & Chen, Y. (2020). Blockchain-enabled trade finance innovation: A potential paradigm shift on using letter of credit. *Sustainability (Switzerland)*.

Chang, Y., Iakovou, E. & Shi, W. (2020). Blockchain in global supply chains and cross border trade: a critical synthesis of the state-of-the-art, challenges and opportunities. *International Journal of Production Research*.

Chen, M.-C., Wang, R.-J. & Chen, A.-P. (2007). An Empirical Study for the Detection of Corporate Financial Anomaly Using Outlier Mining Techniques. *2007 International Conference on Convergence Information Technology (ICCIT 2007)*. IEEE, pp. 612–617.

Chen, S., Shi, R., Ren, Z., Yan, J., Shi, Y. & Zhang, J. (2017). A Blockchain-Based Supply Chain Quality Management Framework. *2017 IEEE 14th International Conference on e-Business Engineering (ICEBE)*. IEEE, pp. 172–176.

Cole, R., Stevenson, M. & Aitken, J. (2019). Blockchain technology: implications for operations and supply chain management. *Supply Chain Management*.

Creswell, J. W. (2007). An Introduction to Mixed Methods Office of Qualitative and Mixed

Methods Research, pp. 1–43.

David, B., Gaži, P., Kiayias, A. & Russell, A. (2018). 'Ouroboros praos: an adaptively-secure, semi-synchronous proof-of-stake blockchain'. , in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 66–98.

Davis, F. D., Bagozzi, R. P. & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, vol. 35(8), pp. 982–1003.

Derindag, O. F., Yarygina, I. Z. & Tsarev, R. Y. (2020). International trade and blockchain technologies: Implications for practice and policy. *IOP Conference Series: Earth and Environmental Science*.

Ding, F., Huo, J. & Campos, J. K. (2017). The Development of Cross Border E-commerce. *Proceedings of the International Conference on Transformations and Innovations in Management*. Paris, France: Atlantis Press, pp. 370–383.

Efanov, D. & Roschin, P. (2018). The all-pervasiveness of the blockchain technology. *Procedia Computer Science*. Elsevier B.V., vol. 123, pp. 116–121.

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *American Association for Artificial Intelligence*, vol. 17(3), pp. 37–54.

Feng Tian. (2017). A supply chain traceability system for food safety based on HACCP, blockchain & Internet of things. *2017 International Conference on Service*

*Systems and Service Management*. IEEE, pp. 1–6.

Finstad, K. (2006). The system usability scale and non-native English speakers. *Journal of Usability Studies*, vol. 1(4), pp. 185–188.

Gan, Z. & Zhou, X. (2018). Abnormal Network Traffic Detection Based on Improved LOF Algorithm. *2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*. IEEE, pp. 142–145.

Ganne, E. (2018). *Can Blockchain Revolutionize International Trade? Can Blockchain Revolutionize International Trade?* World Trade Organization Publications.

Gao, Z. (2009). Application of Cluster-Based Local Outlier Factor Algorithm in Anti-Money Laundering. *2009 International Conference on Management and Service Science*. IEEE, pp. 1–4.

Godbole, S. (n.d.). No Title. IBM Academy of Technology.

Hasan, H. R. & Salah, K. (2018). Blockchain-Based Proof of Delivery of Physical Assets with Single and Multiple Transporters. *IEEE Access*. IEEE, vol. 6, pp. 46781–46793.

Heimgartner, M. (2017). *Improving the Blockchain System for Temperature Monitoring*. University od Zurich.

Hijikata, Y., Ohno, H., Kusumura, Y. & Nishida, S. (2007). Social summarization of text feedback for online auctions and interactive presentation of the summary. *Knowledge-Based Systems*, vol. 20(6), pp. 527–41.

Hooper, A. & Holtbrügge, D. (2020). Blockchain technology in international business:

changing the agenda for global governance. *Review of International Business and Strategy*.

Imeri, A. & Khadraoui, D. (2018). The Security and Traceability of Shared Information in the Process of Transportation of Dangerous Goods. *2018 9th IFIP International Conference on New Technologies, Mobility and Security, NTMS 2018 - Proceedings*, vol. 2018-Janua(February), pp. 1–5.

Jiang, S., Cao, J., Wu, H., Yang, Y., Ma, M. & He, J. (2018). Blochie: A blockchain-based platform for healthcare information exchange. *Proceedings - 2018 IEEE International Conference on Smart Computing, SMARTCOMP 2018*, (April), pp. 49–56.

Kamath, R. (2018). Food Traceability on Blockchain: Walmart's Pork and Mango Pilots with IBM. *The Journal of the British Blockchain Association*, vol. 1(1), pp. 1–12.

Kothari, C. R. (2004). *Research Methodology: Methods and Techniques*. New Age International [online].Available at: https://books.google.ae/books?hl=en&lr=&id=hZ9wSHysQDYC&oi=fnd&pg=PA2&dq=Kothari+2004&ots=1tUdpDg5I2&sig=IO10ksTrsDlv5Nqw2JVgahONEpM&redir_esc=y#v=onepage&q=Kothari 2004&f=false.

Lewis, J. R. (2018). The System Usability Scale : Past , Present , and Future The. *International Journal of Human–Computer Interaction*. Taylor & Francis, pp. 1–14.

Lewis, J. R. & Sauro, J. (2009a). The factor structure of the system usability scale. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial*

*Intelligence and Lecture Notes in Bioinformatics)*. Springer, Berlin, Heidelberg, pp. 94–103.

Lewis, J. R. & Sauro, J. (2009b). 'The factor structure of the system usability scale'. , in *International conference on human centered design*, pp. 94–103.

Li, B. & Wang, Y. (2018). IEEE International Conference On Big Data Science And Engineering RZKPB : A Privacy-preserving Blockchain-Based Fair Transaction Method for Sharing Economy. *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*. IEEE, (3), pp. 1164–1169.

Li, Z., Kang, J., Yu, R., Ye, D. & Deng, Q. (2017). Consortium Blockchain for Secure Energy Trading in Industrial Internet of Things, vol. 3203(c), pp. 1–10.

Lim, Y. H., Hashim, H., Poo, N., Poo, D. C. C. & Nguyen, H. D. (2019). Blockchain technologies in e-commerce: Social shopping and loyalty program applications. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.

Lin, I. C. & Liao, T. C. (2017). A survey of blockchain security issues and challenges. *International Journal of Network Security*, vol. 19(5), pp. 653–659.

Lipsman, A. (2019). *Ecommerce set to cause major disruption - KPMG Ireland*. *eMarketer* [online]. [Accessed 25 January 2020]. Available at: https://home.kpmg/ie/en/home/insights/2019/10/disruptive-companies-and-

business-models.html.

Liu, C., Chai, K. K., Zhang, X., Lau, E. T. & Chen, Y. (2018). Adaptive Blockchain-Based Electric Vehicle Participation Scheme in Smart Grid Platform. *IEEE Access*. IEEE, vol. 6, pp. 25657–25665.

Liu, J., Li, W., Karame, G. O. & Asokan, N. (2018). Toward Fairness of Cryptocurrency Payments. *IEEE Security & Privacy*, pp. 81–89.

Lu, Q. & Xu, X. (2017). Adaptable Blockchain-Based Systems: A Case Study for Product Traceability. *IEEE Software*, vol. 34(6), pp. 21–27.

Malini, N. & Pushpa, M. (2017). Analysis on credit card fraud identification techniques based on KNN and outlier detection. *2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)*. IEEE, pp. 255–258.

Mao, D., Hao, Z., Wang, F. & Li, H. (2018). Innovative Blockchain-Based Approach for Sustainable and Credible Environment in Food Trade : A Case Study in Shandong Province , China.

Mariscal, G., Marbán, Ó. & Fernández, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *Knowledge Engineering Review*, vol. 25(2), pp. 137–166.

McDaniel, C. A. & Norberg, H. C. (2019). Can Blockchain Technology Facilitate International Trade? *SSRN Electronic Journal*.

Munassar, Ali, N. M. & Govardhan, A. (2010). Comparative Study on Agile software

development methodologies. *International Journal of Computer Science Issues (IJCSI)*, vol. 7(5) [online].Available at: http://arxiv.org/abs/1307.3356.

Muzammal, M., Qu, Q. & Nasrulin, B. (2019). Renovating blockchain with distributed databases: An open source system. *Future Generation Computer Systems*. Elsevier B.V., vol. 90, pp. 105–117.

Norberg, H. C. (2019). Unblocking the Bottlenecks and Making the Global Supply Chain Transparent: How Blockchain Technology Can Update Global Trade. *The School of Public Policy SPP Research Papers*.

Olson, D. L. (2007). Data Mining in Business Services. *Service Business*, vol. 1(3), pp. 181–193.

Olson, D. L. & Delen, D. (2008). *Advanced data mining techniques [electronic resource]*. Springer Science & Business Media.

Payne, G. & Payne, J. (2004). *Key Concepts in Social Research*. 1 Oliver's Yard, 55 City Road, London England EC1Y 1SP United Kingdom: SAGE Publications, Ltd.

Pete, C., Julian, C., Randy, K., Thomas, K., Thomas, R., Colin, S. & Wirth, R. (2000). Crisp-Dm 1.0. *CRISP-DM Consortium*, p. 76.

Petersson, Evelina and Baur, K. (2018). *Impacts of Blockchain Technology on Supply Chain Collaboration*. Jonkoping University.

Pitney, W. A. (2004). Strategies for Establishing Trustworthiness in Qualitative Research. *Athletic Therapy Today*. Human Kinetics Publishers Inc., vol. 9(1), pp. 26–28.

Puthal, D., Malik, N., Mohanty, S. P., Kougianos, E. & Yang, C. (2018). Blockchain as a Decentralized Security Framework. *IEEE Consumer Electronics Magazine*, vol. 7(2), pp. 18–21.

Raadt, S. & Xu, J. (2021). 'Improving cross-border ecommerce through digitalization'. , in *The Digital Transformation of Logistics*. Wiley, pp. 209–222.

Ramadhan, B. A. & Iqbal, B. (2018). User Experience Evaluation on the Cryptocurrency Website by Trust Aspect. *2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*. IEEE, vol. 3, pp. 274–279.

Resnick, P., Zeckhauser, R., Friedman, E. & Kuwabara, K. (2000). Reputation systems. *Communications of the ACM*, vol. 43(12), pp. 45–48.

Rosenow, S. & O'Shea, B. J. (2010). *A handbook on the WTO customs valuation agreement*. Cambridge University Press.

Ruparelia, N. B. (2010). Software development lifecycle models. *ACM SIGSOFT Software Engineering Notes*, vol. 35(3), pp. 8–13.

Saltz, J. S. (2015). The need for new processes, methodologies and tools to support big data teams and improve big data project effectiveness. *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, pp. 2066–2071.

Sauro, J. (2018). *MeasuringU: 5 Ways to Interpret a SUS Score* [online]. [Accessed 24 July 2020]. Available at: https://measuringu.com/interpret-sus-score/.

Scally, A. (2019). *Global Ecommerce 2019 - Insider Intelligence Trends, Forecasts & Statistics*. *KPMG* [online]. [Accessed 3 February 2020]. Available at:

https://www.emarketer.com/content/global-ecommerce-2019.

Seebacher, S. & Schüritz, R. (2017). Blockchain Technology as an Enabler of Service Systems: A Structured Literature Review. *International Conference on Exploring Services Science*. Springer, pp. 12–23.

Segers, L., Ubacht, J., Rukanova, B. & Tan, Y. H. (2019). The use of a blockchain-based smart import declaration to reduce the need for manual cross-validation by customs authorities. *ACM International Conference Proceeding Series*.

Sheikh, H., Azmathullah, R. M., Rizwan, F., Sattam, P. & Abdalaziz, B. (2019). A Blockchain-Based Platform Transforms E-Commerce Perspective into a Decentralized Marketplace. *International Journal of Management, Technology and Engineering*.

Shorman, S. M., Allaymounq, M. & Hamid, O. (2019). Developing the E-Commerce Model a Consumer to Consumer Using Blockchain Network Technique. *SSRN Electronic Journal*.

Siyam, N., Alqaryouti, O. & Abdallah, S. (2020). Mining government tweets to identify and predict citizens engagement. *Technology in Society*. Elsevier Ltd, vol. 60(November), p. 101211.

Su, Z., Wang, Y., Xu, Q., Fei, M., Tian, Y. & Zhang, N. (2018). A Secure Charging Scheme for Electric Vehicles with Smart Communities in Energy Blockchain. *IEEE Internet of Things Journal*. IEEE, vol. PP(c), p. 1.

Toyoda, K., Mathiopoulos, P. T., Sasase, I. & Ohtsuki, T. (2017). A Novel Blockchain-

Based Product Ownership Management System (POMS) for Anti-Counterfeits in the Post Supply Chain. *IEEE Access*, vol. 5(XXX), pp. 17465–17477.

Tullis, T. S. & Stetson, J. N. (2004). A Comparison of Questionnaires for Assessing Website Usability ABSTRACT : Introduction. *Usability Professional Association Conference* [online].Available at: http://home.comcast.net/~tomtullis/publications/UPA2004TullisStetson.pdf.

Turkanović, M., Hölbl, M. & Košič, K. (2018). EduCTX : A Blockchain-Based Higher Education Credit Platform, vol. 6.

Vos, D., Overweel, L., Raateland, W., Vos, J., Bijman, M., Pigmans, M. & Erkin, Z. (2018). DEFenD: A Secure and Privacy-Preserving Decentralized System for Freight Declaration. *Cryptography and Security*.

Wang, H., Zheng, Z., Xie, S., Dai, H. N. & Chen, X. (2018). Blockchain challenges and opportunities: a survey. *International Journal of Web and Grid Services*.

WCO, W. C. O. (2008). *International Convention on the Simplification and Harmonization of Customs Procedures* [online]. [Accessed 13 May 2019]. Available at: http://www.wcoomd.org/en/topics/facilitation/instrument-and-tools/conventions/pf_revised_kyoto_conv/kyoto_new.aspx.

WCO, W. C. O. (2018a). *Cross-Border E-Commerce Framework of Standards* [online]. [Accessed 13 May 2019]. Available at: http://www.wcoomd.org/en/media/newsroom/2018/july/wco-publishes-global-standards-on-ecommerce.aspx.

WCO, W. C. O. (2018b). *Post-Clearance Audit (PCA)*.

WCO, W. T. O. (2019). *Customs Valuation* [online]. [Accessed 25 November 2019]. Available at: https://www.wto.org/english/tratop_e/cusval_e/cusval_e.htm.

Wu, D., Yang, R. & Shen, C. (2021). Sentiment word co-occurrence and knowledge pair feature extraction based LDA short text clustering algorithm. *Journal of Intelligent Information Systems*. Springer, vol. 56(1), pp. 1–23.

Wu, H., Li, Z., King, B., Miled, Z. Ben, Wassick, J. & Tazelaar, J. (2017). A distributed ledger for supply chain physical distribution visibility. *Information (Switzerland)*, vol. 8(4), pp. 1–18.

Wulf, L. De & Sokol, J. B. (2005). *Customs modernization handbook*. Washington, DC: The World Bank.

Xia, Q., Sifah, E. B., Asamoah, K. O., Gao, J., Du, X. & Guizani, M. (2017). MeDShare: Trust-Less Medical Data Sharing among Cloud Service Providers via Blockchain. *IEEE Access*, vol. 5, pp. 14757–14767.

Xie, W., Zhou, W., Kong, L., Zhang, X., Min, X., Xiao, Z. & Li, Q. (2018). ETTF : A Trusted Trading Framework Using Blockchain in E-commerce. *2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design ((CSCWD))*. IEEE, pp. 612–617.

Xu, X., Lu, Q., Liu, Y., Zhu, L. & Yao, H. (2019). Designing blockchain-based applications a case study for imported product traceability. *Future Generation Computer Systems*. Elsevier B.V., vol. 92, pp. 399–406.

Xu, X., Pautasso, C., Zhu, L., Gramoli, V., Ponomarev, A., Tran, A. B. & Chen, S. (2016). The Blockchain as a Software Connector. *2016 13th Working IEEE/IFIP Conference on Software Architecture (WICSA)*. IEEE, pp. 182–191.

Yin, C., Wang, X., Rong, W., Wang, T. & David, B. (2018). A System Framework for Sharing Economy. *2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design ((CSCWD))*. IEEE, pp. 808–813.

Ying, W., Jia, S. & Du, W. (2018). International Journal of Information Management Digital enablement of blockchain : Evidence from HNA group. *International Journal of Information Management*. Elsevier, vol. 39(October 2017), pp. 1–4.

Zhang, X., Cui, L. & Wang, Y. (2014). Computing multi-dimensional trust by mining e-commerce feedback comments. *IEEE Transactions on Knowledge and Data Engineering*, vol. 26(7), pp. 1631–1643.

Zheng, Z., Xie, S., Dai, H.-N., Chen, X. & Wang, H. (2018). Blockchain Challenges and Opportunities : A Survey Shaoan Xie Hong-Ning Dai Huaimin Wang. *International Journal of Web and Grid Services*, vol. 14(4), pp. 1–24.

Zheng, Z., Xie, S., Dai, H., Chen, X. & Wang, H. (2016). Blockchain technology: principles and applications. *Research Handbook on Digital Transformations*. Edward Elgar Publishing, pp. 225–253.

Zheng, Z., Xie, S., Dai, H., Chen, X. & Wang, H. (2017). An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. *Proceedings - 2017 IEEE 6th International Congress on Big Data, BigData Congress 2017*, pp. 557–564.

Zhu, X. & Wang, D. (2019). Research on Blockchain Application for E-Commerce, Finance and Energy. *IOP Conference Series: Earth and Environmental Science*.

# Appendixes

## Appendix A: System Usability Scale

This is a standard questionnaire that measures the overall usability of a system. Please select the answer that best expresses how you feel about each statement after using the Customs-based Trade Facilitation website. The results of the study may be published but your identity will remain confidential and anonymous.

The research aims to design a customs-based solution to optimize the e-commerce trade supply chain by employing blockchain technology. In this research, the blockchain capabilities will be employed to keep track of the flow of goods. Moreover, the proposed solution will be utilized to construct a reputation engine as a single channel that eventually seeks to quantify the reputation score for each overseas seller and the traded goods.

This research seeks to address the following objectives:

1. Improve traceability and shipments tracking.
2. simplify the duty drawback process of customs clearance.
3. Simplify the current process of e-commerce trade in customs administration.
4. Provide a centralized reputation engine for the overseas traders' behaviour to help the buyers with the selection of the reputed sellers.

| | Strongly Disagree | Somewhat Disagree | Neutral | Somewhat Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1. I think I would like to use the Customs-based Trade Facilitation website frequently. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. I found the Customs-based Trade Facilitation website unnecessarily complex. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. I thought the Customs-based Trade Facilitation website was easy to use. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. I think that I would need the support of a technical person to be able to use the Customs-based Trade Facilitation website. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5. I found the various functions in the Customs-based Trade Facilitation website were well integrated. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. I thought there was too much inconsistency in the Customs-based Trade Facilitation website. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 7. I would imagine that most people would learn to use the Customs-based Trade Facilitation website very quickly. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 8. I found the Customs-based Trade Facilitation website very cumbersome to use. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 9. I felt very confident using the Customs-based Trade Facilitation website. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 10. I needed to learn a lot of things before I could get going with the Customs-based Trade Facilitation website. | ☐ | ☐ | ☐ | ☐ | ☐ |

**How likely are you to recommend this website to others? (Please circle your answer)**

**Not at all likely** 0 1 2 3 4 5 6 7 8 9 10

**Extremely likely**

## Appendix B: Informed Consent Form

**Date**:

**Researcher Details**:

**Name**: Omar Alqaryouti
**Address**: The British University in Dubai
**Contact**: +971557441442
**Email**: omar.alqaryouti@gmail.com

**Participant Details:**

**Name:**

**Contact:**

**Email:**

**Study Name**: Customs Trade Facilitation and Compliance for Ecommerce using Blockchain and Data Mining

**Purpose of the Research**

The research aims to design a customs-based solution to optimize the e-commerce trade supply chain by employing blockchain technology. In this research, the blockchain capabilities will be employed to keep track of the flow of goods. Moreover, the proposed solution will be utilized to construct a reputation engine as a single channel that eventually seeks to quantify the reputation score for each overseas seller and the traded goods.

This research seeks to address the following objectives:

1. Improve traceability and shipments tracking.
2. simplify the duty drawback process of customs clearance.
3. Simplify the current process of e-commerce trade in customs administration.

4. Provide a centralized reputation engine for the overseas traders' behavior to help the buyers with the selection of the reputed sellers.

**What you will be Asked to do in the Research**?

If you choose to participate, you will be asked to join a two hours workshop. During the workshop, you will be briefed about the application and the new journey as well as you will be trained to use this application. After that, you will be asked to use the application buy goods from various e-commerce platforms as well as you will also be asked to participate in a face-to-face interview with the researcher as well as to fill in a short questionnaire.

**Risks and Discomforts**

We do not foresee any risks or discomfort from your participation in the research. You have the right to not answer any questions.

**Benefits of the Research and Benefits to you**

Even though there may be no direct benefit to you from this study, a possible benefit of your participation is identifying possible ways of improving various processes involved in the e-commerce trade. This can benefit various stakeholders simplifying the various processes from different perspectives including the clearance process. This may result in reducing the time required for the daily e-commerce purchases.

**Voluntary Participation**

Your participation in the study is completely voluntary and you may choose to stop participating at any time. Your decision not to volunteer will not influence nature of your

relationship with the researcher or with The British University in Dubai either now, or in the future.

**Withdrawal from the Study**

You can stop participating in the study at any time, for any reason, if you so decide. Your decision to stop participating, or to refuse to answer particular questions, will not affect your relationship with the researchers or with The British University in Dubai. In the event you withdraw from the study, all associated data collected will be immediately destroyed wherever possible.

**Questions about the Research?**

If you have questions about the research in general or about your role in the study, please contact me. If you choose to participate in this study, you should consider the following:

**Research Results and Confidentiality**

The results of the study may be published but your identity will remain confidential and anonymous. For instance, we might refer to a trader as Trader A, and a shipping agent as Agent A. Interviews might be audio-recorded, but all recordings will be deleted once transcribed. You may also request that no audio recording shall be used. Your answers to the questionnaire will be also anonymous. The research results will be used for publication.

**Consent**

I, _____, consent to participate in ***Customs Trade Facilitation and Compliance for Ecommerce using Blockchain and Data Mining*** conducted by ***Omar Alqaryouti***. I have understood the nature of this project and wish to participate. My signature below indicates my consent.

**Signature** _____ **Date** _____

Participant

Use this section if imagery (photographs or video) will be taken of participants and used in teaching or dissemination of research.

I,_____, agree to allow video and/or digital images or photographs in which I appear to be used in teaching, scientific presentations and/or publications with the understanding that I will not be identified by name. I am aware that I may withdraw this consent at any time without penalty.

**Signature** _____ **Date** _____

Participant