Empowering Customs Agencies with ML/DL techniques

[1]Oumaima BENLAMHAIRA, [2]Hind LAMHARHAR,

[1]Jaafar ABOUCHABAKA.

[1]University Ibn Tofail, Faculty of Science, Department of Computer Science, Kenitra, Morocco,

[2]Moroccan Customs and Indirect Tax Administration, Department of computer Science, Rabat, Morocco.

Abstract:

This research paper addresses the increasing problem of fraudulent activities in the globalized world of international trade, particularly within the electronic customs clearance system. It points out the shortcomings of traditional manual inspection methods and highlights the necessity for advanced technological solutions. The paper focuses on the potential of AI, including machine learning, and deep learning, to efficiently detect and prevent fraud in the Moroccan customs administration. It also explores the applicability of the "BACUDA"project, a collaborative initiative aimed at developing tailored data analysis methodologies and algorithms. The paper offers a proof-of-concept for AI's role in improving customs fraud detection, introducing customized solutions specific to Moroccan customs, such as an AI-driven electronic clearance system, feature engineering techniques, data preparation, and anomaly detection. Additionally, it discusses future directions, emphasizing AI's ongoing advancements, seamless integration with existing customs procedures, and the significance of training and capacity-building programs to enhance fraud detection efforts within Moroccan customs.

Keywords: ML, DL, BACUDA, Fraud, AI, Customs.

I. Introduction

In today's globalized world, the increase in imports and exports has given rise to a growing concern regarding fraudulent activities within the e-clearance system. Instances of misclassified products, manipulated origin countries, and undervaluation of imported goods have become more prevalent, posing significant challenges for customs administrations.

Misclassification of products: This involves intentionally categorizing imported goods under incorrect tariff codes or product descriptions to take advantage of lower import duties or to bypass certain regulations. For instance, a company might label a luxury item as a less expensive product to avoid higher tax rates.

Manipulation of origin countries: Fraudsters may attempt to manipulate the declared country of origin of goods to exploit preferential trade agreements or to evade import restrictions or tariffs. They may falsely declare a different country as the origin to benefit from more favorable trade conditions or to bypass specific regulatory requirements.

Undervaluation of imported goods: This fraudulent practice entails intentionally declaring a lower value for imported goods than their actual market value. By undervaluing goods, importers can reduce the amount of customs duties, taxes, and

fees payable, resulting in significant financial losses for the government.

Traditional methods of customs inspections are labor-intensive and often rely on manual processes, resulting in inefficiencies and potential errors. Here's a summarized overview of these methods:

Physical Inspection: Customs officers manually examine shipments and containers to ensure compliance with regulations. This process is time-consuming and challenging for large volumes of goods.

Paper-Based Documentation Review: Officers review paper documents like invoices and certificates of origin to check for accuracy and consistency. This method is error-prone due to the need for meticulous examination of numerous documents.

Profiling and Risk Analysis: Customs agencies use manual profiling techniques based on risk indicators and patterns, relying on the knowledge and experience of officers. It may not effectively detect new or evolving fraud patterns.

Random Inspections: Customs administrations randomly inspect a percentage of shipments or transactions, which may not be targeted and could miss specific fraud types or patterns.

These traditional inspection methods are labor-intensive, time-consuming, and can be inefficient in detecting sophisticated fraudulent activities. The increasing volume of international trade and the complexities of global supply chains require more advanced and efficient approaches to fraud detection, such as leveraging artificial intelligence and machine learning algorithms.

In this paper, we focus on one type of fraud which is the undervaluation of imported goods.For example, a shipment of electronic devices with a true market value of $10,000 may be declared as $5,000 to reduce the customs duty payable.It can have significant negative impacts on a country's economy. This fraudulent practice not only results in revenue loss for the government but also distorts trade statistics and affects fair competition in the market. When importers pay lower customs duties and taxes due to false declarations, the country's treasury misses out on crucial revenue streams that could fund essential public services and infrastructure development. Furthermore, it can create an uneven playing field for honest businesses that adhere to regulations, potentially discouraging foreign investments and undermining the overall economic stability of the nation. Detecting and preventing undervaluation fraud is, therefore, of paramount importance for safeguarding a country's financial interests and ensuring the fairness and integrity of international trade.

To tackle these challenges, our research highlights the substantial body of work in the field of customs fraud detection, with a particular emphasis on the utilization of Artificial Intelligence (AI), specifically Machine Learning (ML) and Deep Learning (DL). AI presents a promising avenue for improving fraud detection within customs administrations. There are several approaches to deal with fraud detection. We highlight the use of neural networks [1, 2], bayesian networks [3], expert systems [4], rule based systems [5] and the detection of statistical outliers [6, 7, 8].

These approaches can be subdivided in two groups: supervised and unsupervised. In the supervised approaches there is a training set of operations that are labeled either as fraudulent or normal. These operations are used as input to some systems, such as neural network systems, that need labeled inputs to construct the model that will be used to detect frauds.

This paper delves into the potential applications of AI in this context, showcasing its advantages over traditional methods and its ability to process vast amounts of data while identifying irregularities efficiently. Through the deployment of diverse AI techniques and

algorithms, such as Natural Language Processing, ML, and DL, the goal is to enhance fraud detection and alleviate the workload of customs officials by swiftly identifying suspicious transactions.

Furthermore, this paper explores the future directions for AI in fraud detection, emphasizing the continuous evolution of AI technologies, seamless integration with existing customs procedures, and the importance of training and capacity-building initiatives. By embracing AI-driven solutions, customs administrations can bolster their fraud detection efforts, ultimately leading to more effective and efficient customs operations.

In addition to this, our study investigates the potential application of the "BACUDA" project, which brings together customs authorities and data scientists to develop cutting-edge data analytics methodologies and algorithms. Leveraging open-source languages and tools, this initiative empowers customs administrations to implement customized solutions tailored to their specific requirements, effectively combating fraud and enhancing their operational efficiency.

This paper seeks to underscore the significance of employing AI as a viable solution for addressing the challenges associated with customs fraud detection, while also shedding light on the intricacies involved in managing customs data. The remainder of the paper is organized as follows in section1 we will present the process of fraud detection the types of fraudulent activities and discuss the rule engine-based system for fraudulent activities and the problems that comes with it. In section2 we will demonstrate the use of AI technologies in the domain of fraud detection in customs. And finally discuss our approach for this project.

II. Landscape of Customs Fraudulent Activities in:

In this section we present in details the process of fraud customs inspection and explain each type of fraudulent activities in addition to this we illustrate what are the rule engine based system and there downsides.

II.1 Process of fraud customs inspection

The process of fraud customs inspection involves a series of steps and measures taken by customs administrations to identify and prevent fraudulent activities in international trade. While the specific process may vary among countries, the following are common steps involved in fraud customs inspection.

1. **Risk assessment:** Customs administrations conduct risk assessments to identify high-risk shipments or entities based on various factors such as transaction history, importer/exporter profiles, intelligence reports, and known fraud patterns. This helps prioritize inspections and allocate resources effectively.
2. **Target selection**: Based on the risk assessment, customs authorities select specific shipments or transactions for inspection. This can be done using profiling techniques, data analysis, or random selection methods. The objective is to focus on high-risk cases that have a higher likelihood of fraud.
3. **Document verification:** Customs officers review and verify the documentation accompanying the shipments, including invoices, bills of lading, packing lists, and certificates of origin. They compare the information provided in the documents with the actual goods being imported or exported to ensure accuracy and compliance with regulations.
4. **Physical inspection**: In cases where a physical inspection is deemed

necessary, customs officers physically examine the goods, containers, or packages. They may open and inspect individual items to verify their nature, quantity, quality, and value. This step helps detect discrepancies between the declared and actual goods, as well as identify potential smuggling or misclassification.

5. **Technology-assisted inspections:** Customs administrations may employ various technologies to enhance fraud detection. This can include X-ray scanning systems, automated cargo inspection systems, and other advanced tools that help identify hidden or illegal goods. These technologies aid in non-intrusive inspections and can expedite the inspection process.
6. **Post-clearance audits**: After the customs clearance process, customs administrations may conduct post-clearance audits to verify the accuracy and compliance of the declared information. This includes cross-checking with financial records, conducting site visits, and ensuring that importers/exporters are adhering to tax and customs regulations.
7. **Training and capacity building:** Continuous training programs are provided to customs officers to enhance their skills in fraud detection and keep them updated with the latest techniques and trends in fraudulent activities. Capacity building initiatives aim to improve the overall efficiency and effectiveness of the fraud detection process.

By following these steps and leveraging technological advancements, Moroccan customs administration aims to detect and prevent fraudulent activities, protect national revenue, and ensure fair and compliant international trade.

II.2 Fraudulent activities

The current landscape of fraudulent activities in Moroccan customs presents significant challenges for trade facilitation and revenue protection. As globalization continues to drive increased imports and exports, illicit practices such as *misclassification of products*, *manipulated origin countries*, and *undervaluation of imported goods* have become more prevalent.

Fraudulent practices in customs, including *misclassification of products* and *manipulation of origin countries*, have detrimental effects on government revenue, trade fairness, and national security in Morocco. *Misclassification* leads to reduced tariff payments due to incorrect product categorization, while *origin manipulation* exploits trade agreements to gain competitive advantages. *Undervaluation* of goods results in lower customs duties and taxes. These practices not only erode government income but also undermine fair competition and pose security risks. To address these issues, Moroccan customs administration must enhance its fraud detection through AI and data analytics, automating risk assessment and pattern recognition to curb fraudulent activities and safeguard the integrity of trade transactions.

By addressing the current landscape of fraudulent activities through improved detection and prevention measures, Moroccan customs administration can ensure fair trade practices, protect national revenue, and contribute to a secure and transparent international trade environment.

II.3 Rule engine based system for fraudulent activities

Customs fraud detection rules play a crucial role in identifying suspicious activities and potential fraud within the import and export processes. These rules are typically implemented within a rule-based engine, utilizing “if/then” programming logic to automate the detection and classification of fraudulent behavior. For the privacy concerns of real fraud detection rules implementation, we provide only

some examples of generic fraud detection rules.

II.3.1 Rules implementation

The actual implementation of fraud detection rules within the customs administration may involve a broader and more comprehensive set of rules tailored to their specific needs and risk factors. These rules are developed based on extensive data analysis, domain expertise, and ongoing monitoring of fraudulent trends. Here are some examples of customs fraud detection rules:

Rule 1: If the declared value of a high-value item exceeds a predetermined threshold, then flag the transaction for further inspection.

> **Purpose**: Identifies potential cases of undervaluation or misrepresentation of goods to evade customs duties or taxes.

Rule 2: If the country of origin specified in the declaration does not match the expected country for a particular product, then mark the transaction as suspicious.

> **Purpose**: Identifies cases where the origin country is manipulated to take advantage of preferential trade agreements or avoid import restrictions.

Rule 3: If the declared quantity of goods significantly deviates from the expected quantity for a specific product category, then investigate the transaction further.

> **Purpose**: Detects instances of quantity misreporting or smuggling attempts by underreporting or overreporting the quantity of goods.

Rule 4: If the declared weight of a shipment exceeds the weight limit for a particular transportation mode or container type, then trigger an alert.

> **Purpose**: Helps uncover cases of potential illegal shipment practices, such as overloading or misclassification of goods to avoid higher freight costs.

Rule 5: If the declared classification code for a product does not match the corresponding tariff code, then flag the transaction for closer scrutiny.

> **Purpose**: Identifies cases of intentional misclassification to benefit from lower import duties or to bypass specific regulations or restrictions.

Rule 6: If the declared description of goods is inconsistent with the accompanying documentation or previous records, then investigate the transaction further.

> **Purpose**: Uncovers potential cases of document forgery or false declarations to conceal illicit activities or prohibited items.

Rule 7: If multiple shipments with similar characteristics originate from different companies but share common intermediaries or beneficiaries, then initiate a comprehensive investigation.

> **Purpose**: Detects potential schemes involving front companies or shell entities to facilitate fraud, money laundering, or smuggling.

Those aren't real rules they were generated using AI due to confidential purposes.

II.3.2 Problem of rule-based system: Continuous refinement

Continuous refinement and adaptation of fraud detection rules are critical for the effectiveness of a rule-based system in customs agencies. Collaboration with experts and data analysts ensures optimal accuracy and minimizes false positives. The system should allow for rule customization and adaptation to address emerging fraud tactics. Comprehensive audit reports help monitor and improve fraud detection. Fraud reporting mechanisms strengthen

stakeholder collaboration, while mutual assistance agreements and information sharing are vital for collective fraud combat. Sector-specific studies aid in identifying susceptible industries. Privacy and data protection regulations are strictly followed. The Moroccan customs administration aims to enhance fraud detection to protect national revenue, trade security, and fair competition through a rule-based engine.

III. Applying AI technologies in customs fraud detection:

For an efficient fraud detection system, AI technologies play a crucial role in various domains, including credit card fraud detection. Several AI technologies have been developed, such as:

Machine Learning (ML) for Predictive Analysis:ML algorithms are used for predictive modeling to identify patterns of fraudulent behavior based on historical data, transactional information, and user behavior.

Deep Learning (DL) for Anomaly Detection: DL techniques, such as deep neural networks and autoencoders, excel in detecting subtle anomalies in large datasets, making them valuable for uncovering sophisticated fraud schemes.

Natural Language Processing (NLP) for Text Analysis: NLP is utilized to analyze text-based data, including emails, chat logs, and customer feedback, to identify linguistic cues or anomalies that may indicate fraudulent activities.

Generative AI for Synthetic Data Generation: Generative AI models can create synthetic datasets that mimic real-world data, which can be used to train fraud detection models and improve their accuracy and robustness.

These AI technologies, when integrated into a fraud detection system, enhance its ability to detect and prevent fraudulent activities effectively and efficiently.

III.1 Related works

Detecting fraud through conventional audit procedures can be both expensive and labor-intensive, particularly in customs where there may be a shortage of officers with the necessary expertise. This challenge has led to the exploration of computational solutions for the automatic identification of suspicious operations. These solutions leverage data mining and statistical techniques to uncover fraudulent activities.

Identifying suspicious activities poses a challenge across various sectors, including credit card fraud, telecommunications fraud, terrorism prevention, financial crime detection, and computer intrusion detection. Detecting fraud is crucial because preventive measures can sometimes prove inadequate [9], and an effective detection system must be capable of adapting to recognize emerging fraudulent behaviors on its own.

Brazilian customs initially implemented supervised learning for the purpose of choosing goods for human verification, a method first detailed in [10]. While alternative strategies were explored in [11] without yielding significant advantages, a novel approach outlined in [12] delivered noteworthy enhancements in certain performance metrics. In contrast, unsupervised methods do not require labeled inputs; they rely on a set of rules to categorize an operation as potentially fraudulent or assess it against previous operations to flag potential anomalies (outliers).

Rule-based systems, which fall under unsupervised approaches, utilize a set of predefined rules to categorize operations as either fraudulent or normal. They may also assign a probability score to each operation, indicating the likelihood of it being fraudulent. These rules are usually crafted based on expert guidance, harnessing the expertise of specialists to formulate the evaluation criteria for each operation. While these systems offer the advantage of being unsupervised and benefiting from expert

knowledge, a notable drawback is their need for frequent rule updates to address emerging fraudulent behaviors. Without regular updates, these rules can become outdated and less effective over time.

Using outlier detection (e.g., [7, 8]) to spot potential fraud is an unsupervised technique that involves comparing each operation to prior ones. This approach has several advantages, including its ability to adapt and identify new behaviors as fresh data is added to the system. Furthermore, it provides a transparent statistical interpretation for each suspicious operation. For instance, the system can compute that a particular operation deviates by four standard deviations from its anticipated value and that such an occurrence happens only once in every thousand operations. In such cases, the operation is considered an outlier and warrants further investigation as it's deemed suspicious.

A crucial requirement for outlier detection systems in fraud detection is that most of the stored operations should be classified as normal (non-fraudulent). Additionally, it's vital to underscore that being identified as an outlier does not automatically imply fraud. In addition to this assumption, it's also essential to guarantee the accurate registration and classification of importers, exporters, and products.

Other research in the field of customs fraud detection has proposed solutions along two primary avenues. The first set of approaches centers on the exploration of new cases, ranging from basic yet intuitive random inspections [13] to more advanced methods like active learning, which leverages uncertainty [14] and diversity [15], [16]. The second group of approaches focuses on utilizing knowledge acquired previously, including the adaptation of heuristic methods and the use of widely available machine learning techniques such as XGBoost [17], SVM [18], and sophisticated deep-learning-based approaches like DATE [19]. While inspecting random items can enhance fraud detection performance, it may involve sacrificing the identification of some known instances of fraud [20].

III.2 Discussion

Despite the existence of previous research in the field of customs fraud detection, our project faces unique challenges that require us to start from scratch. The complexity of our dataset, characterized by rare events and a lack of resources due to dealing with millions of rows and a large number of complex features, makes finding patterns challenging.

Our fraud detection efforts primarily focus on three core datasets: the Operators dataset, the Customs Declaration dataset, and the Litigation dataset. These datasets are sourced from the Moroccan customs e-clearance system and serve as critical sources of information for identifying potential fraudulent activities.

The Operators dataset contains information about the individuals and entities involved in import and export transactions, providing valuable insights into the actors within the customs ecosystem.

The Customs Declaration dataset, on the other hand, plays a central role in fraud detection by offering data on declared goods, payment details, origins, and various attributes. This dataset also benefits from a rule-based engine, which assigns color flags to products based on their risk level, aiding in prioritizing inspection efforts.

Lastly, the Litigation dataset is instrumental in understanding legal aspects related to customs operations and provides in-depth information on detected fraud cases. It classifies fraud based on a solid legal foundation and customs inspectors' expertise.

Analyzing the litigation dataset allows us to gain insights into various types of fraud and adapt our detection strategies

accordingly. This dataset evolves as new cases are investigated, ensuring that we stay vigilant against emerging fraud schemes.

While our current focus is on these datasets, future research will explore additional data sources, including unstructured web data related to operators and goods in the context of e-commerce. This integration of diverse data types will further enhance our customs fraud detection capabilities.

In the next section, we present an overview of how we have applied AI for undervaluation

IV. AI based approach for customs fraud detection

Our methodology for implementing a system to address the issue of undervaluation involves two distinct approaches:

1. Exploring the Potential of the Bacuda Project:

- We will begin by examining the Bacuda project to identify any existing solutions or insights that can be leveraged to tackle undervaluation effectively.

2. Building a Machine Learning Pipeline from Scratch:

- In parallel, we will develop a machine learning pipeline solution from the ground up. This approach will involve designing and implementing a custom solution tailored to our specific needs.

IV.1 Bacuda Project

In the collaborative BACUDA project, an advanced artificial neural network called DATE has been developed for detecting under-declared imports while maximizing customs revenue. DATE was successfully applied to Nigeria Customs, achieving impressive accuracy and recall rates with minimal inspections. This innovative approach showcases the potential of artificial intelligence to enhance customs operations and revenue collection worldwide.

IV.1.1 DATE architecture

DATE's robustness stems from combining a tree-based model for interpretability with transaction-level embedding and dual-attention mechanisms. It accurately identifies illicit transactions and predicts tax revenues by simultaneously learning from both illegality and overcharging in each transaction. To fully understand its functioning, the DATE model utilizes a diverse ecosystem of machine learning and deep learning programs and tools, dissecting it into different parts that ultimately contribute to the complete model.

IV.1.2 Implementation of DATE on Moroccan SGDs

The obstacles we found when trying to apply the BACUDA approach to our data is the differences in data structure. Therefore, we cannot apply the same feature engineering techniques they used because our datasets have distinct features. Additionally, their feature engineering was conducted by experts, making it costly and challenging for us to replicate. As a consequence, not adhering to their feature engineering methods results in lower accuracies and catastrophic recalls.

IV.2 ML based system for undervaluation:

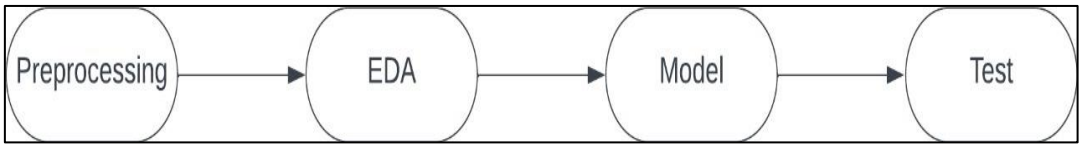

Figure 1 : proposed approach workflow

IV.2.1 Preprocessing:

In this project, the initial step was preprocessing the data to ensure its suitability for machine learning analysis. The dataset 'SGD' was the cornerstone, comprising a vast amount of data. Two preprocessing approaches were explored: classic preprocessing and the PyCaret methodology. However, early results were compromised due to an inaccurate target column. A pivotal breakthrough emerged when a more comprehensive dataset was introduced. The merging of datasets based on a common identifier, 'IDDUM'(the id for each SGD) was a crucial preprocessing step that laid the foundation for subsequent machine learning endeavors. In addition to the challenges mentioned, our project grappled with several data-related obstacles that we encountered in the process of the classical preprocessing. One significant hurdle was handling rare events within the dataset, where fraudulent cases were considerably outnumbered by legitimate transactions. This rarity posed a challenge in effectively training the models to identify these infrequent instances accurately.

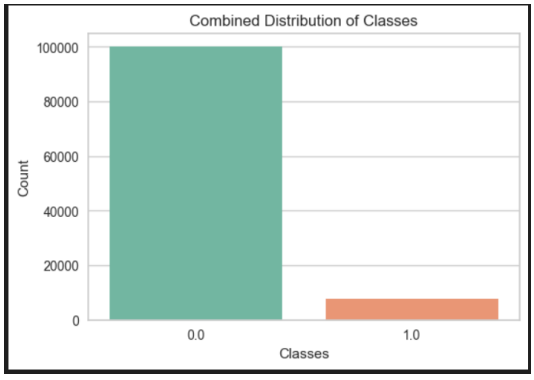

Figure 2 : Class distribution

Dealing with missing values was another vital aspect of data preprocessing. The dataset often had gaps, requiring careful strategies to impute or handle these missing entries effectively. Moreover, addressing duplicate records was crucial to ensure the integrity and accuracy of the dataset, necessitating thorough deduplication processes.

Lastly, variations in data types across different features added to the complexity. Harmonizing these diverse data types for coherent analysis and modeling required meticulous data transformation techniques.

IV.2.2 Exploratory Data Analysis (EDA):

The project necessitated a thorough Exploratory Data Analysis (EDA) phase to comprehend the intricacies of the dataset and glean insights crucial for machine learning model development. This involved delving into the merged dataset, understanding the composition of fraudulent and non-fraudulent cases, and identifying patterns and trends. EDA also revealed the challenges posed by the rare occurrence of fraud events, prompting the adoption of specialized techniques like autoencoders. Understanding the dataset's structure and nuances was instrumental in informing feature selection and model development.

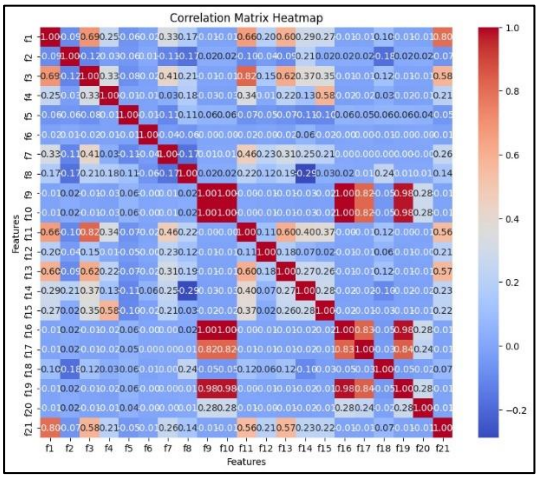

Figure 3 : Correlation Matrix of all features

IV.2.3 Model Development and model testing and evaluation:

With a well-preprocessed dataset and insights from exploratory data analysis (EDA), the project moved into model development. Since we were building everything from scratch, we decided to try various machine learning and deep learning algorithms in order to select the one with the

highest accuracy and recall rate using the conventionally preprocessed dataset.

We utilized neural networks, One-Class SVM, which is known for its use in unsupervised machine learning models, particularly for rare events, and the XGBoost model.

For the svm model the initial results were not promising, especially in terms of recall for fraudulent events. After some adjustments to the threshold, we managed to achieve a recall of 82% for fraudulent events. Unfortunately, this improvement in recall for fraudulent events came at the cost of a decreased recall rate of 52% for non-fraudulent events.

This trade-off between recall for fraudulent and non-fraudulent events presents a common challenge in classification tasks and might require further exploration and fine-tuning to strike a better balance between the two classes.

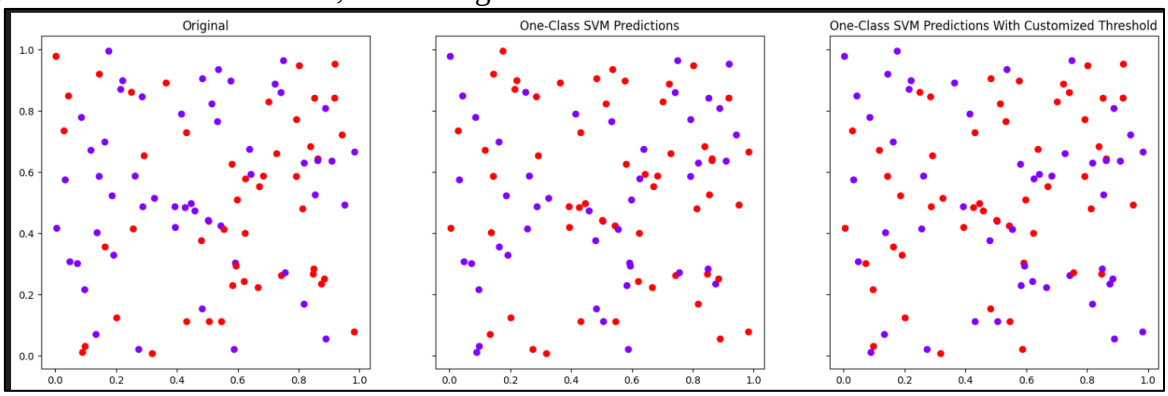

Figure 4 : One class SVM

Regarding the neural network (NN), the recall achieved was only 36%, which is relatively low. After hyperparameter tuning, the XGBoost model showed improved performance, with recall rates ranging between 62% and 65%, making it the best-performing model among the proposed models.

Then, we explored the PyCaret approach, which is a Python library designed to streamline the machine learning workflow from data preprocessing to experimenting with various machine learning models. When we applied PyCaret to the dataset, the best-performing models were Quadratic Discriminant Analysis and Naïve Bayes.

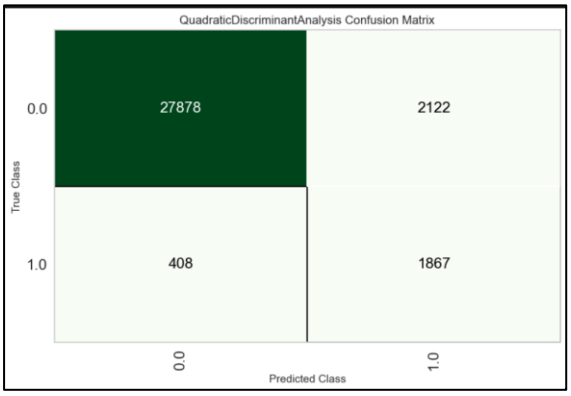

Figure 5 : Confusion Matrix of Quadratic Discriminant Analysis

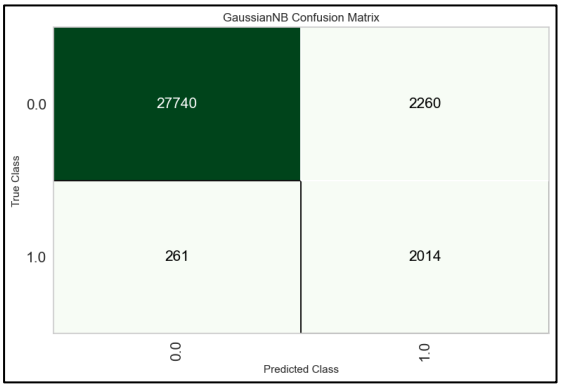

Figure 6: Confusion Matrix of Naive Bayes

After observing that we could achieve better recall with the preprocessing done by PyCaret, we decided to take a different approach to preprocess our data, which involved using target mean encoding. Subsequently, we applied the newly preprocessed dataset to the optimized XGBoost model, and as a result, we achieved a recall of 82% and an accuracy of 97%.

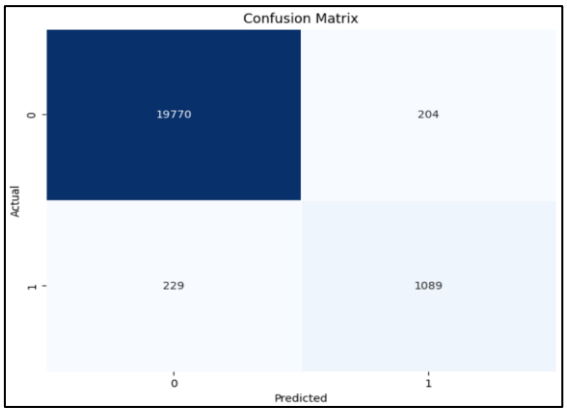

Figure 7 : Confusion Matrix of XGBoost

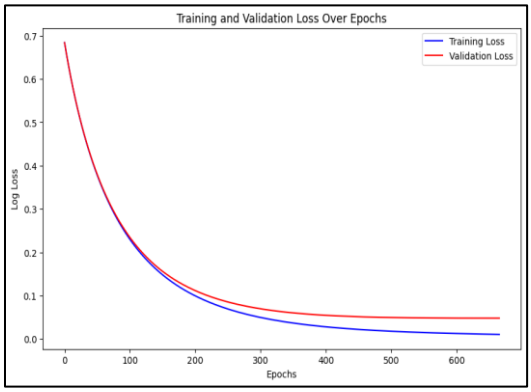

Figure 8 : Training and Validation loss over epochs (XGBoost)

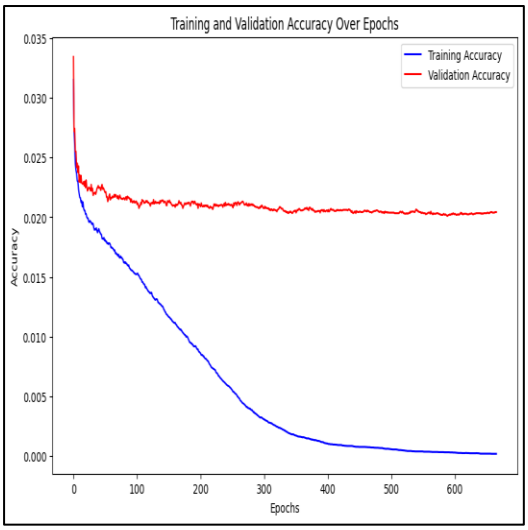

Figure 9 : Training and Validation Accuracy over epochs (XGBoost)

Model	Preprocessing technique	Recall	Accuracy
XGBoost	Classical Preprocessing (Z-score Normalization, Label Encoding categorical variables)	65 %	96%
One Class SVM		82%	54%
Neural Network		36%	94%
Naïve Bayes	PyCaret Preprocessing	87%	92%
QDA		92%	92%
XGBoost	**Target Mean Encoding**	**82%**	**97%**

V. Conclusion

The ever-evolving landscape of international trade, propelled by globalization, necessitates a robust response to combat the rising tide of fraudulent activities. The traditional manual inspection methods, marred by inefficiencies and potential errors, highlight the urgency for advanced technological solutions. This research delves into the potential of Artificial Intelligence (AI),machine learning, and deep learning, to significantly

enhance fraud detection within the Moroccan customs administration,

The study sheds light on the shortcomings of conventional customs inspection methods, where time-consuming physical inspections and paper-based documentation reviews are prevalent. It emphasizes the need for a paradigm shift toward AI-driven systems that can efficiently process vast amounts of data, identify irregularities, and expedite the identification of potential fraudulent activities.

Our research presents a tailored machine learning pipeline specifically designed to tackle undervaluation fraud. It navigates through the challenges posed by vast datasets, feature selection, and the delicate balance between data richness and model performance. The significance of preprocessing, exploratory data analysis, and the selection of appropriate models, notably XGBoost, has been underscored.

In conclusion, the potential of AI to enhance fraud detection in Moroccan customs administration is immense. As AI technologies continue to evolve and seamlessly integrate into customs operations, there is a promising future ahead. Embracing AI-driven solutions will not only bolster fraud detection efforts but also fortify customs operations, ultimately contributing to fair trade practices, revenue protection, and a more secure international trade environment. The journey towards an AI-powered future for customs fraud detection has just begun, and its potential to reshape the dynamics of international trade is vast and exciting.

VI. Acknowledgment

We gratefully acknowledge support and the copyediting performed by the MOROCCAN customs administration. I would like to express my sincere appreciation to Pr. Hind LAMHARHAR the project manager. For her valuable contributions and support throughout this research.

VII. Reference:

[1] K. Fanning and K. Cogger. Neural network detection of management fraud using published financial data. International Journal of Intelligent Systems in Accounting, Finance & Management, 17(1):21–24, 1998.

[2] B. P. Green and J. H. Choi. Assessing the risk of management fraud through neural network technology. Auditing: A Journal of Practice and Theory, 16(1):14–28, 1997.

[3] S. Maes, K. Tuyls, B. Vanschoenwinkel, and B. Manderick. Credit card fraud detection using Bayesian and neural networks. In Proceedings of the 1st International NAISO Congress on Neuro Fuzzy Technologies, 2002.

[4] M. M. Eining, D. R. Jones, and J. K. Loebbecke. Reliance on decision aids: an examination of auditors assessment of management fraud. Auditing: A Journal of Practice and Theory, 16(2):1–19, 1997.

[5] A. Deshmukh and T. Talluru. A rule based fuzzy reasoning system for assessing the risk of management fraud. Journal of Intelligent Systems in Accounting, Finance & Management, 4:669–673, 1997.

[6] V. Hodge and J. Austin. A survey of outlier detection methodologies. Artificial Intelligence Review, 22(2):85–126, 2004.

[7] N. T. Roman, E. R. Constantino, H. Ribeiro, J. J. Filho, A. Lanna, S. K. Goldenstein, and J. Wainer. Carancho – a decision support system for customs. In Proceedings of ECML PKDD Workshop on Practical Data Mining: Applications, Experiences and Challenges, pages 100–103, September 2006.

[8] K. Yamanishi, J. Takeuchi, G. Williams, and P. Milne. On-line unsupervised outlier detection using finite

mixtures with discounting learning algorithms. Data Mining and Knowledge Discovery, 8(3):275–300, 2004

[9] D. Yue, X. Wu, Y. Wang, Y. Li, and C.-H. Chu. A review of data mining-based financial fraud detection research. In International Conference on Wireless Communications, Networking and Mobile Computing (WiCom), pages 5514–5517, September 2007.

[10] M. A. C. Ferreira. Uso de redes de cren¸ca para sele¸c˜ao de declara¸c˜oes de importa¸c˜ao. Master's thesis, Instituto Tecnol´ogico de Aeron´autica, 2003.

[11] J. Jambeiro Filho and J. Wainer. Analyzing Bayesian networks with local structure and cardinality reduction over a practical case. In Proceedings of the Workshop on Computational Intelligence (WCI), 2006.

[12] J. Jambeiro Filho and J. Wainer. Using a hierarchical Bayesian model to handle high cardinality attributes with relevant interactions in a classification problem. In Proceedings of the International Joint Conference

[13] C. Han and R. Ireland, "Performance measurement of the KCS customs selectivity system," Risk Management, vol. 16, no. 8, pp. 25–43, 2014

[14] N. Houlsby, F. Huszar, Z. Ghahramani, and M. Lengyel, "Bayesian ´ active learning for classification and preference learning," 2011.

[15] N. Houlsby, F. Huszar, Z. Ghahramani, and M. Lengyel, "Bayesian ´ active learning for classification and preference learning," 2011.

[16] O. Sener and S. Savarese, "Active learning for convolutional neural networks: A core-set approach," in ICLR, 2018.

[17] J. T. Ash, C. Zhang, A. Krishnamurthy, J. Langford, and A. Agarwal, "Deep batch active learning by diverse, uncertain gradient lower bounds," in ICLR, 2020.

[18] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in KDD, 2016, pp. 785–794.

[19] J. Vanhoeyveld, D. Martens, and B. Peeters, "Customs fraud detection: Assessing the value of behavioural and high-cardinality data under the imbalanced learning issue," Pattern Analysis and Applications, vol. 23, 2020.

[20] S. Kim, Y.-C. Tsai, K. Sigh, Y. Choi, E. Ibok, C.-T. Li, and M. Cha, "DATE: Dual attentive tree-aware embedding for customs fraud detection," in KDD, 2020, pp. 2880–2890.

[21] S. Kim, T.-D. Mai, T. N. D. Khanh, S. Han, S. Park, K. Singh, and M. Cha, "Take a chance: Managing the exploitation-exploration dilemma in customs fraud detection via online active learning," arXiv preprint arXiv:2010.14282, 2020.

[22] https://bacuda.wcoomd.org/