5-6-2024

Optimizing Customs Efficiency: Detection for Illicit Shipments in Cargo

Ohood Alharbi
oha6342@rit.edu

Optimizing Customs Efficiency: Detection for Illicit Shipments in Cargo

by

Ohood Alharbi

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Professional Studies: Data Analytics

Department of Graduate Programs & Research

Rochester Institute of Technology

06/05/2024

RIT

Master of Science in Professional Studies:

Program Name

Graduate Thesis Approval

Student Name**: Ohood Alharbi**

Graduate Capstone Title**: Optimizing Customs Efficiency: Detection for Illicit Shipments in Cargo**

Graduate Thesis Committee:

Name: Dr. Sanjay Modak **Date:**

Chair of committee

Name: Dr. Hammou Messatfa **Date:**

Member of committee

Acknowledgments

First, there are no better words than the words of Allah, I would like to begin my research with a prayer from the Quran Surah Ta-Ha: “My Lord, increase me in knowledge.”

Second, I would like to express my gratitude to all the professors who have contributed positively to my learning journey, with special and warm-hearted thanks to Dr. Hammou. His mentoring and inspiration I will carry as a treasure in my heart.

Finally, thank you, my husband and partner, for all the support and motivation to follow my dreams. I dedicate this success to you.

Abstract

Aim- The study aimed for creating a machine learning algorithm which succeeded customs control procedures through estimating shipments being either illicit or non-illicit. The study aimed at boosting the level of accuracy in determining the critical shipments, which would in turn, increase customs productivity, minimize false positive and false negative, as well as augment security and ensure the smoothness of trade flow, amid a palpable surge in importation rates.

Methods- The study implemented the CRISP-DM methodology (Cross-Industry Standard Process for Data Mining), which has a well-structured approach of achieving standard data mining solutions. At the beginning, the efforts have been mainly made on defining both the project and operation targets, resolving the specific issues and developing the control and resources within the customs context. Then, data collection and understanding were the next tasks that the researcher needed to deal with, which held attributes like the number, types, quality, and association. Then data cleaning, transformation, and features construction were involved, which included data preparation for modeling. Modelling phase would be in strongly connected with the ETL process through construction, examination and assessment of models making sure that the model that is selected matches the requirements involving illicit and non-illicit shipments most of all. As the last step, the IBM SPSS Software has been utilized for statistical evaluations, considering ROC curve, precision, and confusion matrix as evaluation metrics for identifying the best suited model in terms of accuracy for predicting a complex dataset of customs operations.

Findings- The study revealed that customs authorities were confronted with insurmountable problems particularly the ability to differentiate between safe goods and contraband items, which is getting more challenging while keeping other objects flowing. In this context, the call to arms involved two tasks; the first one was aimed to enhance existing detection processes and the second one was to find out how to apply a neural network model that could specifically address the issues and improve the efficiency of air cargo shipments. The research results showed that the model was effective in detecting illicit shipments that are hidden within the huge range of international air shipments and therefore, it represents a distinctive solution to elevate the security level and defend the economic interests. The developed neural network model acted as a crucial tool in terms of countering the risks related to the national security threats and illicit activities which could have

consequentially impacted the trade flows and solidified economic instability. Through correct identification of shipments with a high likelihood of illicit cargo entering customs checkpoints, the model succeeded in reducing the chances of undiscovered contraband successfully passing through, which helped to prevent security breaches and losses. Such capability was especially significant in the cases of high-level customs visits where the repercussions of undetected clandestine shipments include breach of state security and the adverse economic implications. Key performance metrics of the neural network model was its high dependability in shipping contraband packages through its high recall performance measures. This metric had a center place in optimization of customs inspection procedures, which makes a possibility of accomplishment a more targeted and efficient control of entering shipments to a country and minimization of disruption of the legal trade. In addition, harnessing sophisticated technologies such as neural network models help customs agencies to increase their resilience as well as the ability to cope with evolving risks and new smuggling tactics. As a result, they are better placed to increase security and efficiency in the context of constant global trade dynamics advancement.

Keywords

Customs inspection, detection of illicit shipments, risk management, neural network, Decision tree, LSVM, XGBoost, missing values, feature transformation, feature importance analysis, recall, ROC curve, optimal cut-off point.

Table of Contents

ACKNOWLEDGMENTS ..... **II**

ABSTRACT ..... **III**

LIST OF FIGURES ..... **VII**

LIST OF TABLES ..... **VIII**

CHAPTER 1- INTRODUCTION **1**

1.1. BACKGROUND AND STATEMENT OF PROBLEM 1

1.1.1. Background Information *1*

1.1.2. Statement of the Problem *2*

1.2. PROJECT GOALS 2

1.3. AIMS AND OBJECTIVES 2

1.4. LIMITATIONS OF THE STUDY 2

1.5. STRUCTURE OF THE THESIS 3

CHAPTER 2 – LITERATURE REVIEW **4**

CHAPTER 3- RESEARCH METHODOLOGY **10**

CHAPTER 4- DATA ANALYSIS FINDINGS **11**

4.1. DESCRIPTION OF THE DATASET USED 11

4.2. EXPLORATORY DATA ANALYSIS (EDA) 11

4.2.1. Data Profiling and Summary Statistics *11*

4.2.2. Data Cleaning *13*

4.2.3. Approaches for Detecting Outliers *16*

4.2.4. Significance Analysis Between Second Inspection and Other Variables *17*

4.2.5. Features Engineering *21*

4.2.6. Summary of the Final Dataset *26*

4.3. MACHINE LEARNING MODEL DEVELOPMENT 26

4.3.1. Feature Importance Analysis *27*

4.3.2. Detailed explanation of the chosen machine learning algorithms *28*

4.3.3. Validation and Testing Procedures *29*

4.3.4. Results *34*

CHAPTER 5- DISCUSSION **50**

CHAPTER 6- CONCLUSION **51**

6.1. CONCLUSION 51
6.2. CONTRIBUTIONS TO KNOWLEDGE 51
6.3. PRACTICAL IMPLICATIONS 52
6.4. RECOMMENDATIONS AND FUTURE WORK 52
REFERENCES/BIBLIOGRAPHY 53

List of Figures

Figure 1: Summary of Used dataset 11
Figure 2: Customs Inspection Results 12
Figure 3: Data Cleaning 14
Figure 4: Import Status Values Proportion 15
Figure 5: Flight Status Values Proportion 15
Figure 6: Cleaned Data Record 16
Figure 7: Anomaly detection 17
Figure 8: “Expected Pieces” and “Expect Weight” with the target variable “Second Inspection” 19
Figure 9: ULD Type 20
Figure 10: Flight Type 20
Figure 11: Dest 21
Figure 12: Flight Day 22
Figure 13: Flight Month 22
Figure 14: Flight Night or Day 23
Figure 15: Dest Group 24
Figure 16: Level of Risk Origin 25
Figure 17: Carriere Risk Level 25
Figure 18: Data Partition 30
Figure 19: Data set Before Balancing 30
Figure 20: Dataset After Sampling 31
Figure 21: Business Features Input 34
Figure 22: Feature Automatic Selection Input 34
Figure 23: Predictor Importance 40
Figure 24: Predictor ULD Type 41
Figure 25: Predictor Flight Type 42
Figure 26: Predictor Level Risk Origin 42
Figure 27: Predictor Flight Day 43
Figure 28: Predictor Flight Month 43
Figure 29: Predictor Day Flight Night 44

Figure 30: Predictor Carrier Risk Analysis 44
Figure 31: Numeric Predictor Whitney U Test 45
Figure 32: Numeric Predictor Distribution 1 46
Figure 33: Numeric Predictor Distribution 2 46
Figure 34: Neural Network ROC 48

List of Tables

Table 1: Descriptive Statistics 1 12
Table 2:Descriptive Statistics 2 12
Table 3: Hypothesis Testing 1 18
Table 4: Hypothesis testing 2 18
Table 5: Feature selection Node 27
Table 6: The Random Tree 28
Table 7: Business Feature Selection 28
Table 8 Dataset Balancing Factor 31
Table 9: Hypothesis Testing 3 45
Table 10: Confusion Matrix 46
Table 12: Cherry pick data points 48

Chapter 1- Introduction

1.1. Background and Statement of Problem

1.1.1. Background Information

Customs are the main entity for maintaining the flow of goods while ensuring the security of nations. They play a crucial role in world trade and supplies by regulating clearance procedures and boosting efficient supply chain management. The main duties of customs personnel are to prevent the importation of illicit shipments and to impose customs trade laws while maintaining the flow of importation in a time-constrained manner. The method followed by customs personnel is a rule-based risk assessment to detect non-compliant imports, most rules rely on pre-established standards, including declared content, origin, route, or historical information. The growth rate of imports from August 2022 till 2023 has increased by 22% and 17% of the imports are by air [1]. Within the numerous import disclosures to isolate and target high-risk shipments, rule-based profiling has inherent limitations, despite its pragmatism. It could unintentionally divert attention to less urgent issues or lead to the incomplete inspection of dangerous imports. For customs controlling cargo imports, identifying high-risk shipments can be difficult for customs personnel due to the cargo's rapid mobility and limited resources. The risk assessment that is performed manually based on experience and personal judgment has the disadvantage of being dependent on the involvement and assessment of the human factor, thus, such risk assessment relies on real-time efforts and experiences to adapt to new forms of fraud or risks [2]. Customs limitations and lack of resources have a major effect on international trade, the large volume of international trade puts heavy pressure on customs personnel to control the high number of transactions efficiently [3]. Moreover, such a challenge could lead to delays and bottlenecks at border crossings, and impact the flow of goods and supply operations. This project emphasizes the importance of fully dropping this strategy of human intervention by utilizing more intelligence and automated risk analysis to increase customs productivity. This study aims to create a machine-learning model that will assist customs controls in taking better and more informed actions of importing high-risk goods or the ones that are illicit as a whole to compensate for resource deficiency and to guarantee security levels. The machine learning techniques integration into customs checkpoints decision making could make it possible for customs to move from a random or rule-based inspection towards

focused and flexible decision making. This could be done by using machine learning modeling based on historical data and making it possible for the calculation of potential risks.

1.1.2. Statement of the Problem

Customs officers now are facing a critical issue because the number of imported goods has increased by 22% it is difficult to stop illicit cases thus supporting legitimate trade. That aggravated inflow of imports has caused various failures, which are false disclosures, delays, insecurity, and others. To that end, this study is concerned with using machine learning that can help in the proactive detection of high-risk cargo shipments. Machine learning algorithms and historical data analysis have once again shown us that the digitization of customs procedures for controlling cargo can facilitate high accuracy in identifying contraband products in air cargo. The main goal is the considerable lowering of the cases of false identification and mentioning data-supported security plans as the way to advance customs security and to provide for both economic interests and the safety of the public.

1.2. Project Goals

- To Develop a machine learning model that predicts the likelihood of an air shipment being illicit or non-illicit.
- To improve the accuracy and efficiency of targeting high-risk shipments.
- To increase customs productivity and reduce false positives and false negatives in customs inspection procedures.
- To maintain effective security and trade flow.

1.3. Aims and Objectives

In light of 17% of imports being shipped by air, this project addresses the urging challenge faced by customs personnel in cargo. The surge in imports has strained customs resources and created inefficiencies, prompting the need for data-driven solutions to increase customs productivity and maintain effective security by improving the detection of illicit shipments.

1.4. Limitations of the Study

The limitations of this research are related to data availability and quality, fast changes in smuggling tactics that may not immediately reflect in the training dataset, and the major limitation

is the ethical considerations and privacy concerns regarding the usage. Moreover, deployment of the model in the real world to collect expert feedback for the result and evaluate business outcomes.

1.5. Structure of the Thesis

The thesis structure is as follows;

Chapter 2: This chapter contains a comprehensive review of different studies, which prove the success of machine learning solutions in the customs inspection domain. Moreover, studies that discuss data quality and enhancement techniques along with engineering new attributes are suggested to improve the result of the classification model. Finally, studies concerning the performance of the classification model such as performance metrics, data splitting, and balancing techniques.

Chapter 3: This chapter is the blueprint, which explains the methodology followed to achieve the main goals and objectives of this research.

Chapter 4: This chapter is about the findings of the analysis of the customs inspection dataset starting from understanding, preparing, and transforming the data using different techniques, along with data visualization, and building and evaluating the classification model.

Chapter 5: This chapter discusses the final result in alignment with the research goals and objectives.

Chapter 6: This chapter provides a conclusion of the research methodology and findings, highlighting its limits and offering recommendations for future research projects.

Chapter 2 – Literature Review

To enhance the effectiveness and accuracy of customs detection, machine learning has been applied in previous studies to support customs efficiency;

(Bassem Chermiti, 2019) According to the research published in the World Customs Journal, data mining techniques have predictive analytical capabilities, which can ultimately enhance risk management analytical capabilities. The main objective of the research is to identify which import declarations are most likely non-compliant using actual data so that customs personnel can utilize a data mining model that identifies import declarations that pose a high risk for further examination. The technique used in the paper is the CHAID decision tree. The model successfully determined the customs risk variables connected to import declarations entered into the system for customs clearance. The model also successfully predicted non-compliant customs declarations based on established guidelines**.**

(Xin Zhou, 2019) In other research published in the same above-mentioned journal, the paper mainly focuses on vulnerabilities in customs procedures and how they can be identified using data mining. It discusses cost-sensitive classification, which accounts for the expenses associated with incorrectly categorizing declarations posing a high risk. The dataset mentioned in the research, sourced from China Customs, contains 30,000 observations of customs-inspected declaration histories, including 82.73% classified as negative (indicating true declarations) and 17.27% classified as positive (indicating false declarations). The data mining technique used was a decision tree combined with a boosting technique, and it achieved higher accuracy with a testing set of 94.1% and a training set of 95.96%. The research concluded with the evaluation of decision trees and boosting by measuring AUC (area under the curve), and the results were 0.991 and 0.982, respectively, indicating how well the model separates the two classes. Significantly increases the precision of customs risk detection models which extend to customs operations and improve their risk detection capacities.

(Han et al., 2023) In addition to the previous techniques, the research introduced a new technique for establishing risk rules in the customs entry inspection routine. In the process, each of the features of the customs declaration data is assigned a different weight using a dynamic method, this is to guarantee that the most relevant elements are given the highest weight. The risk

parameters are also developed by the approach through the application of the enhanced dynamic-weight Can-Tree iterative extraction algorithm. The research aims are to explore unseen trends in customs risk data, design a risk rule base for entry inspection tests in security clearing, and increase the intelligence of customs risk screening.

(Camossi et al., 2012) It is worth noting that a different data mining technique was utilized instead. In the research, they used SVM (Support Vector Machines) for the one-class classification and unsupervised outlier detection to discover anomalies in the customs data. SVM is a sturdy machine-learning technique that is applicable for categorizing data in which an occurrence of customs data is either normal or anomalous within the customs-data used system. The dataset used in the research is historical data from three years of collection, including more than three hundred thousand itineraries for fifty thousand containers. The model effectively detected anomalous itineraries and achieved the desired classification accuracy to classify the anomalies as either high-risk or low-risk.

(González García & Mateos Caballero, 2021) In this paper, the authors discuss the significant threats of customs fraud, especially for the economy. They emphasize the need to optimize the accuracy of the inspection control process despite the challenges faced in customs, such as the requirement for in-the-moment decision-making, the qualitative nature of competing aims, and the abundance of item attributes. The study applied a new approach to detect customs fraud using MOBADO, which is short for Multi-Objective Bayesian with Dynamic Optimization. The approach is an integration of Bayesian decision theory and dynamic optimization with machine learning to achieve the main aim of enhancing customs inspection and better allocation of resources to increase the effectiveness of the customs inspection process. The result of this paper is promising, where the method not only increased and doubled the precision of the inspection but also optimized customs resources and automated 50% of human tasks.

(Singh et al., 2023) The paper proposed a new approach to tackle customs fraud detection by leveraging both supervised and unsupervised learning data. The aforementioned approach solves the critical problem of handling large amounts of trade transactions and the shortage of resources available for manual inspection. The authors collected data from three different countries, and the model used is a graph neural network (GNN). It is a semi-supervised technique that combines both labeled and unlabeled data to increase the effectiveness of customs fraud detection. The technique

designs and creates a transaction graph from tabular data, and the nodes in the graph are the transactions, and the model can connect between them based on joined features such as importer ID and HS-coded. The model can capture the correlations between different trade transactions and use the valuable information gathered from the vast amount of unlabeled data. The experimentation with the semi-supervised model shows tremendous improvement in customs fraud detection, where this model achieved an increase in recall, indicating the effectiveness of using different types of data and a semi-supervised model.

(Regmi & Timalsina, 2018) The study analyzed customs inspection data from Nepal in 2017, using 200,000 randomly selected datasets. They utilized a Deep Neural Network (DNN) model for optimizing customs inspections by identifying high-risk and low-risk shipments. The model was compared to a decision tree model and an SVM model, The DNN model significantly outperformed the other models, achieving an overall accuracy of 96.68% whereas decision tree and SVM models, achieved accuracies of 95.21% and 94.02%, respectively. The model neural network model was able to classify with high accuracy indicating a robust model capable of effectively reducing the volume of shipments that need a detailed inspection, thereby optimizing resource utilization in customs inspections.

To further streamline customs data mining solutions, it's essential to address performance metrics that complement the effectiveness of classifiers;

(M & M.N, 2015) The article mentioned that there are different metrics to measure the performance of the classification model, such as threshold-based discriminators. These metrics depend on the confusion matrix table of predicted and actual classes of the data points, which are accuracy, misclassification error, sensitivity, specificity, precision, and recall. The mentioned metrics measure the proportion, where accuracy measures the proportion of accurate predictions and misclassification error measures the inaccurate predictions. Sensitivity, on the other hand, measures the proportion of positive instances that are accurately classified, while specificity measures the proportion of negative instances that are accurately classified. In addition, precision measures the proportion of predicted positive instances that are positive and recall measures the proportion of actual positive instances that are accurately predicted. Meanwhile, a different metric used to measure the ranking performance of the classifier is the Area Under the ROC Curve (AUC), which shows how well the classifier can rank positive instances higher than negative

instances in the binary classification model. The values of AUC take the sub-set of 0-1. If the AUC is 1 it means, there is a perfect performance of the classifier on one side and when the value is 0.5 it means that the classifier is just randomly doing it. The description of the mentioned metrics assists in the selection of the top classifier in the course of the classification model's training.

Evaluation metrics are central quality improvement tools, but data quality should be also taken into consideration. If a classifier uses a dataset as a training sample and the dataset, contains missing values, then the classifier functions may be imperfect.

(Estabrooks et al., 2004) Implement informative research on the subject of dealing with the class imbalance problem in the dataset. The likelihood of the outcome is biased as the number of instances of one specific class in the data set is much more than the number of instances of other classes; to be precise, it creates challenges for the efficiency of the machine learning model. The paper discussed the resampling method to balance the dataset and focuses on two primary methods: through the case of over-sampling and under-sampling. The first sampling method was over-sampling which is just representing the minor class more in the data set to have a more balanced data. The second sampling method, under-sampling, is a decrease in the number of instances in the majority class to match the minority class in the data to gain balanced data. The authors were struck with finding out which up-sampling method was superior. Generally, it was noted that while oversampling and under sampling offered no universal superiority, they still proved to be effective tools for specific datasets and metrics. Subsequently, the measurement of the performance of the classification model in correctly categorizing distorted data. For that, the need to use the aforementioned methods proved to improve the classification model's performance when the data contains misbalanced data points.

(Singh & Upadhyaya, 2012) This paper presents clearly how outliers define and influence the model performance. Outliers in the paper are defined as the unusual data points that are abnormal to the other data points of the statistical data or behavior of the data. In the second part of the paper, the influence of outliers in the case of initial data on the analysis is shown; it will marginalize the results, distort the analysis, or lead to inaccurate results. Thus, getting the outlier data processed before deploying the model is a critical phase that might determine the appropriateness and perfection of the model.

(Smiti, 2020) The study is an insight into the various methods used in the detection of outliers. The first approach is a statistics tool used in evaluating the dataset's distribution values to know if the dataset contains outliers or not. However, this technique can be unproductive if data settlement is hazy and unknown. The second is the distance-based approach, which plots the distances between the data points. This method is known for its simplicity but may not work best for large and high-dimensional datasets. The third method is density-based; this technique sheds light on the local density around the data points. The points being in low-density regions compared to their neighbors are considered outliers. The problem with this method is that it requires high computational power. The fourth and last method mentioned in the study is cluster-based; any data points that do not fit into a cluster are considered outliers. This method works best for a dataset that contains clear groupings and needs domain knowledge to set the parameters of the clusters. The paper concluded that all methods have their advantages and disadvantages, and the proper choice depends on the type of dataset on hand and must take the best outlier detection method to improve the data analysis and machine learning model performance accuracy.

In line with enhancing the data, **(Nargesian et al., 2017)** definition led to consider feature engineering as a modified preexisting features addition to a dataset to have higher classification model performance. The goal of feature engineering is to create original features with the capacity to guide the learning and prediction process by discovering intricate patterns or relations in the data. One of the contributions of feature engineering is the ability to improve the predictive performance of the model. As a result, it is apt for classification tasks. Feature engineering gives the model an opportunity to explore the information deeper, in detail, so it can more accurately differentiate between the categories or classes.

(Bashir et al., 2020) explained model performance on training data to point to a particular issue where underfitting and overfitting are not desired. The model will be described as overfitting when it draws more information than needed, indeed, it perceives the noise. Yet, on the other hand, if it cannot detect the required data and does not get to know the complexity of the data well enough, the model would definitely be underfitting.

(Montesinos López et al., 2022) presents three pure models: a simple model that underfits, an intermediate model that is a good fit, and a complex model that overfits. The best predictive model, causing the lowest bias and the highest variance to capture dominant patterns which is not fitting

to noise, has the best performance in unknown data. The book talks about the problems of overfitting and underfitting and the role of the metric parameter of the model. In the case of the complex models of the neural network e.g. (NN) overfitting is more likely. Hereby, to solve the problem of overfitting deploying different measures such as adding restrictions, and dropout in the neural network model which is achieved by randomly reducing part of the weights to zero, and simplifying the model, should be applied. The modeling learning failures can be avoided using different approaches: for instance, increasing the sample size of the trained data, adjusting the hypothesis or parameters of the model, alternating the training data representation, or using a different machine learning algorithm.

In conclusion, the deployment of data mining techniques will enhance the productivity of customs and offer data-driven insights. Apart from data quality and performance metrics, the model efficacy of the system also is an important factor. As customs authorities are faced with growing expedition volumes, data mining will become more and more critical for inspecting illicit shipments.

Key takeaways

1. Machine learning models are applied to enhance customs efficiency and risk detection.
2. The machine learning models are used to identify customs risk variables and predict non-compliant customs declarations in customs clearance operations.
3. Feature selection is as important as feature engineering to build an efficient machine learning model.
4. The neural network model compared to the decision tree and SVM performance is better in classifying shipment as illicit and non-illicit with a high accuracy rate.
5. The metrics like accuracy, sensitivity, specificity, precision, recall, and AUC are necessary for the evaluation of classification model performance and are beneficial in selecting the model that suits the business decisions.

Chapter 3- Research Methodology

The chosen methodology is the Cross-Industry Standard Process for Data Mining (CRISP-DM), a standard approach for developing and launching data mining solutions. The choice behind the CRISP-DM method is because it guides the practitioner to follow a structured step-by-step process in order to enable the achievement of results. The first move should be to develop **business understanding** in terms of the major targets of the project, particular questions that are resolved by the data mining project as well as the constraints and resources available at hand. Once defining the business goals, the next step would be **gathering and comprehending the data**. This involves undergoing the data to get to know its characteristics which include but are not limited to quantity, type, quality, and correlation. The fourth step is **data preparation** which involves the preparation of the data for the modeling phases, this process includes cleaning transformation, and the creation of new features. The following phase is **modeling**, where the data mining models are constructed and **evaluated**. Additionally, the best technique that provides accurate predictions of illicit and non-illicit shipment will be chosen. The IBM SPSS Software platform will be utilized to perform quick statistical analysis of the data such as ROC curve, precision, confusion matrix, and additional measures for the evaluation of the model's performance. This action is important to measure the accuracy and determine the best-fitted model that answers the proposed business problem.

Chapter 4- Data Analysis Findings

4.1. Description of the dataset used

In this research, the source of the data is provided by a cargo company, and the data has been transformed for privacy reasons. The data contains historical information from the year 2022 about shipments imported from different countries by cargo. The data provided with previous customs inspection results (“first inspection” and “second inspection"). The initial records of the data are 135,813, and the data consists of 34 attributes of different types. However, the data have incomplete values and N/As.

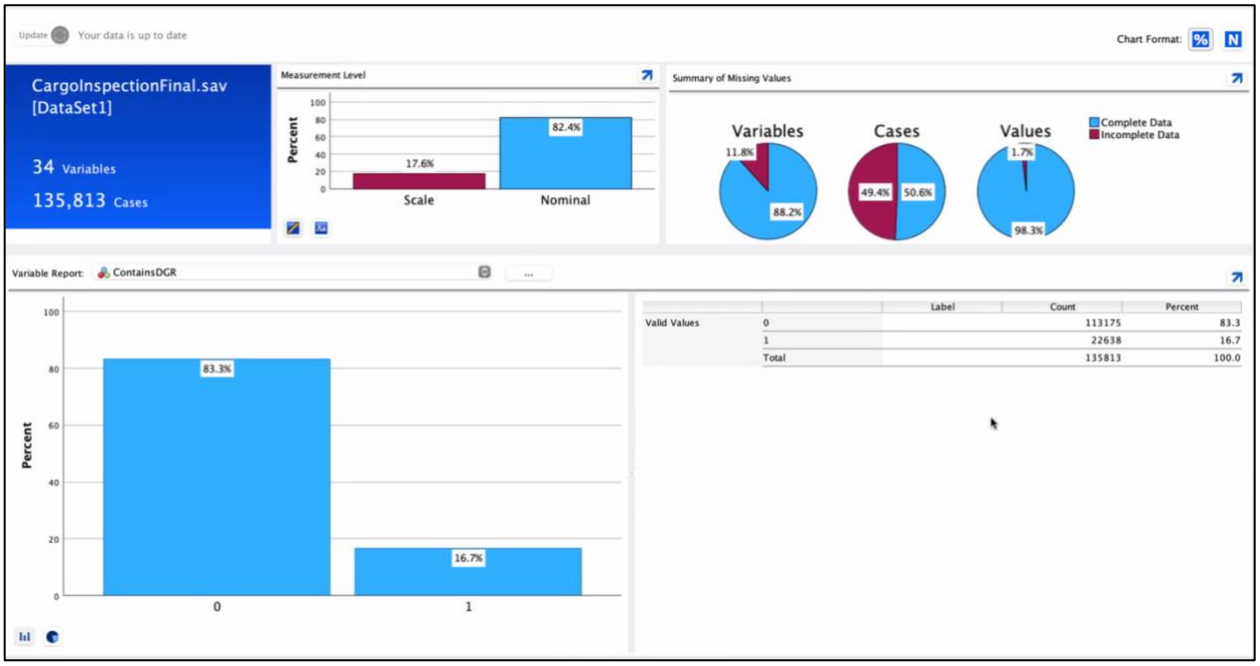


Figure 1: Summary of Used dataset

4.2. Exploratory Data Analysis (EDA)

4.2.1. Data Profiling and Summary Statistics

The dataset types are categorical, date, continuous, and nominal variables. Examples of nominal attributes of the customs inspection data are "first inspection" and "second inspection," where 0 indicates non-illicit shipments and 1 indicates that this shipment contains illicit shipments. The distribution of the aforementioned attributes indicates a major difference in the inspection results.

After customs at cargo performed a second inspection the number of illicit shipments detected increased noticeably revealing a significant issue in the shipment clearance decision making.

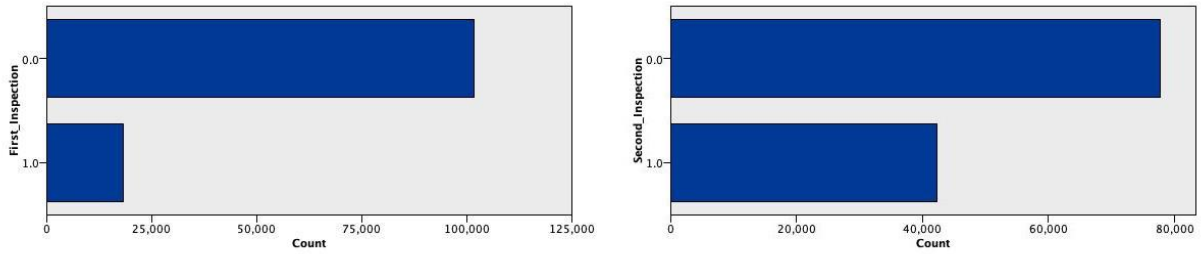


Figure 2: Customs Inspection Results

Understanding these data types is crucial for subsequent statistical analysis and data preparation. After performing a statistical summary of the data, it was observed that attributes such as "Dest" have 27 unique values, while "FlightType" exhibits 4 distinct values as shown in the below figure. The figure also shows that the attribute “Transaction Type” has only one value as “imported” and knowing this information from the unique values we can decide to drop this column to its insignificant to the model decision or prediction. Additionally, the summary provides insights into the quality of the data, revealing factors such as incomplete records, which lead to a thorough investigation to identify and address data quality.

Table 1: Descriptive Statistics 1

Numeric Attributes	Measurement	Min	Max	Mean	Standard deviation	Skewness
ExpectPieces	Continuous	0	1,200	21.791	56.875	6.673
Expect Weight	Continuous	0	20,2546	1,122.722	1,875.848	21.761

Table 2:Descriptive Statistics 2

Attribute	Measurement	Unique
Second_Inspection	Nominal	2
Dest	Nominal	27
LegOrigin	Nominal	117
LegDest	Nominal	27

Station	Nominal	27
Carrier	Nominal	83
AircraftType	Nominal	69
FlightStatus	Nominal	2
ULDType	Nominal	3
FlightType	Nominal	4
ImportStatus	Nominal	5
ActionStatus	Nominal	7
AWBType	Nominal	2
TransactionType	Nominal	1
First_Inspection	Nominal	2

4.2.2. Data Cleaning

The utilization of the data audit node in the IBM SPSS modeler helped to uncover the complete picture of the data quality, providing valuable insights about the number of outliers, extremes, and missing values. The below result of the data quality tab addresses that the data contained missing values in five attributes, which are (“Second inspection”, “Flight Number”, “Flight Status”, “Aircraft Type”, and “Import Status”). The records of the attribute “Second Inspection” are 96.795% complete, showcasing there are 3,844 null values. In the second attribute, “Flight Number," the percentage of complete data is 98.6% with 1,679 null values. The third attribute, “Flight Status”, shows that 98.77% of the records are complete and 1,467 are considered missing. Similarly, the attribute “Aircraft Type” has 99.551% complete data, and the attribute “Import Status” has 99.987%, where 538 and 16 are also considered missing. Moreover, the data quality tab also revealed outliers and extreme outliers in two attributes (“expected pieces” and “expected weight"). The attribute “Expect Pieces” exhibits 1,436 outliers and 587 extreme outliers, and the attribute “Expect Weight” also exhibits 1,430 outliers and 210 extreme outliers. The information above demonstrates a data quality of 77.27%, which requires further enhancement.

): 77.27% Complete records (%): 96.59%

	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value	Empty String	White Space	Blank Value
on	Nominal	--	--	--	Never	Fixed	96.795	116088	3844	0	0	0
	Continuous	5366	0	None	Never	Fixed	98.6	118253	1679	0	0	0
	Nominal	--	--	--	Never	Fixed	98.777	118465	0	1467	1467	0
	Nominal	--	--	--	Never	Fixed	99.551	119394	0	538	538	0
	Nominal	--	--	--	Never	Fixed	99.987	119916	0	16	16	0
	Continuous	39	54	None	Never	Fixed	100	119932	0	0	0	0
	Nominal	--	--	--	Never	Fixed	100	119932	0	0	0	0
	Continuous	0	0	None	Never	Fixed	100	119932	0	0	0	0
	Continuous	0	0	None	Never	Fixed	100	119932	0	0	0	0
	Nominal	--	--	--	Never	Fixed	100	119932	0	0	0	0
	Nominal	--	--	--	Never	Fixed	100	119932	0	0	0	0
	Nominal	--	--	--	Never	Fixed	100	119932	0	0	0	0
	Continuous	0	43	None	Never	Fixed	100	119932	0	0	0	0
	Nominal	--	--	--	Never	Fixed	100	119932	0	0	0	0
	Continuous	1436	587	None	Never	Fixed	100	119932	0	0	0	0
	Continuous	1430	210	None	Never	Fixed	100	119932	0	0	0	0
	Nominal	--	--	--	Never	Fixed	100	119932	0	0	0	0
	Nominal	--	--	--	Never	Fixed	100	119932	0	0	0	0
	Nominal	--	--	--	Never	Fixed	100	119932	0	0	0	0
	Nominal	--	--	--	Never	Fixed	100	119932	0	0	0	0
e	Nominal	--	--	--	Never	Fixed	100	119932	0	0	0	0
	Nominal	--	--	--	Never	Fixed	100	119932	0	0	0	0

Figure 3: Data Cleaning

4.2.2.1. Techniques for Handling Missing Values

There are different types of missing data, in this case, the data are missing are missing completely at random (MCAR). The data could be missing due to different reasons such as human error forgetting to record the values, loss of sample, or some technical errors while recording the values [16]. This indicates that there is no correlation between the values that are missing and the values within the dataset. The strategy to handle MCAR is to remove the missing values and utilize the available data within the dataset.

4.2.2.1.1. Discarding Missing Values

Records with empty values, such as in "Flight Number," a unique identifier for flights that cannot be replaced through mean imputation or estimation, were considered for removal. Similarly, for the attribute "Aircraft Type," which is the type of aircraft and is impossible to impute; therefore, the unavailable information was discarded. The missing values in the attributes "ULD Type" and "Second Inspection" were also discarded due to the values being missing completely at random.

4.2.2.1.2. Dropping Unnecessary Columns

The decision made for the attribute "Import Status" after viewing the distribution of the values showed that 98.62% of the values are "IMP" which is the short-term of the word imported.

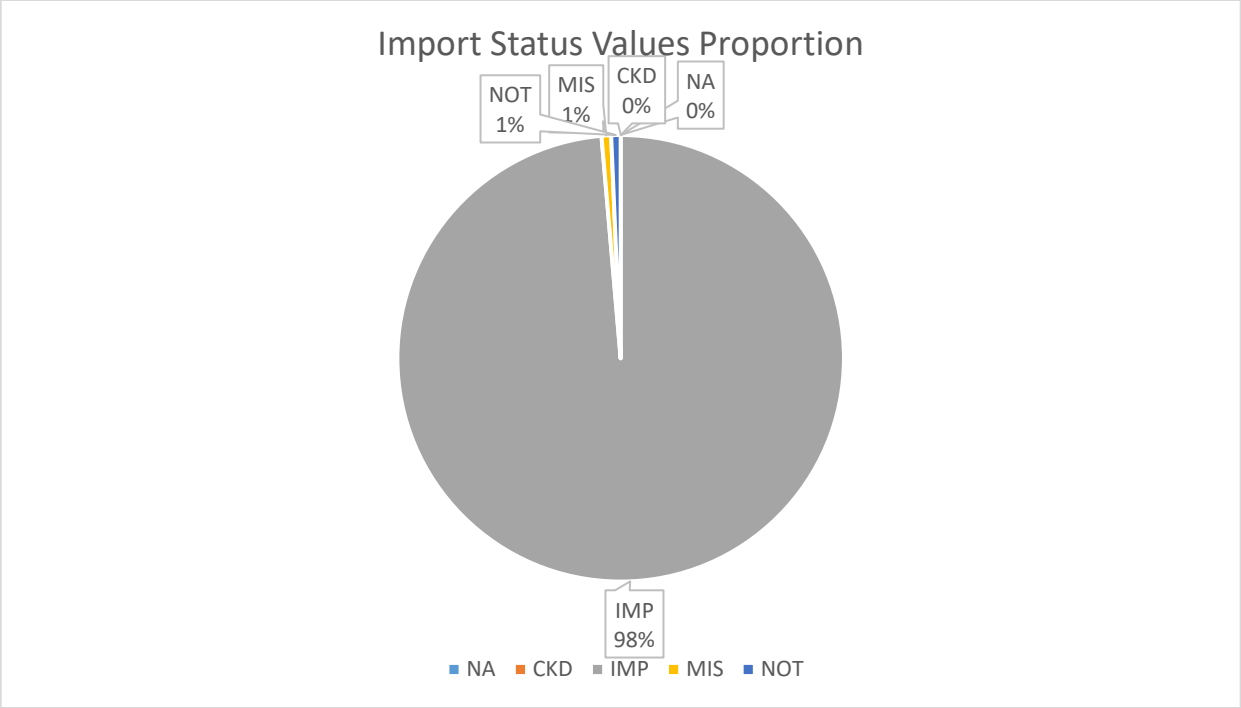


Figure 4: Import Status Values Proportion

The distribution of the attribute "Flight Status" was also conducted showing that the attribute contains only one value which is arrived.

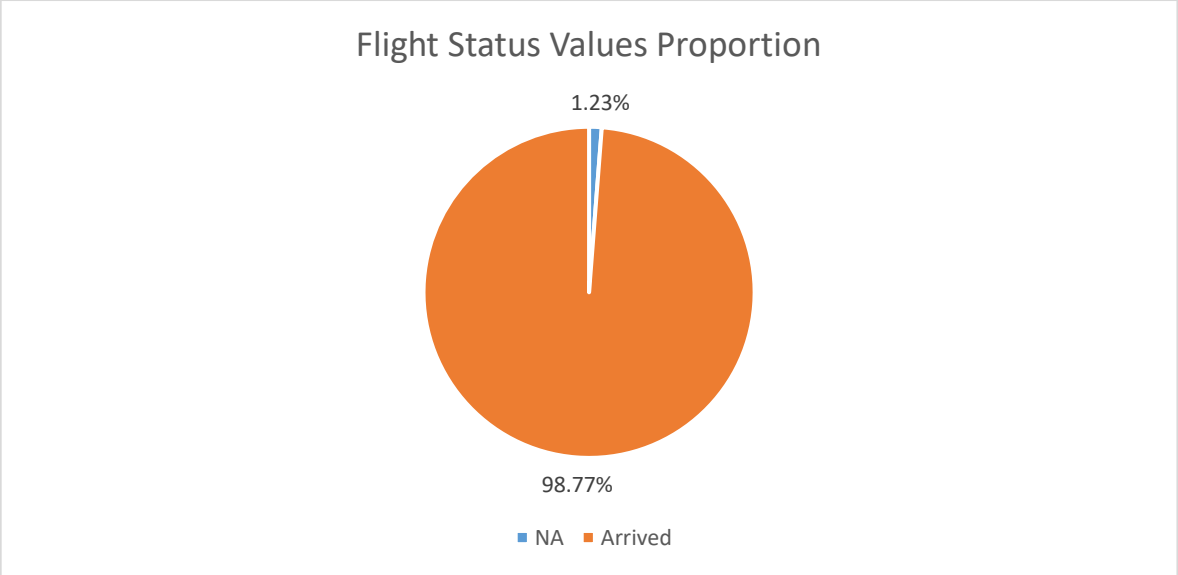
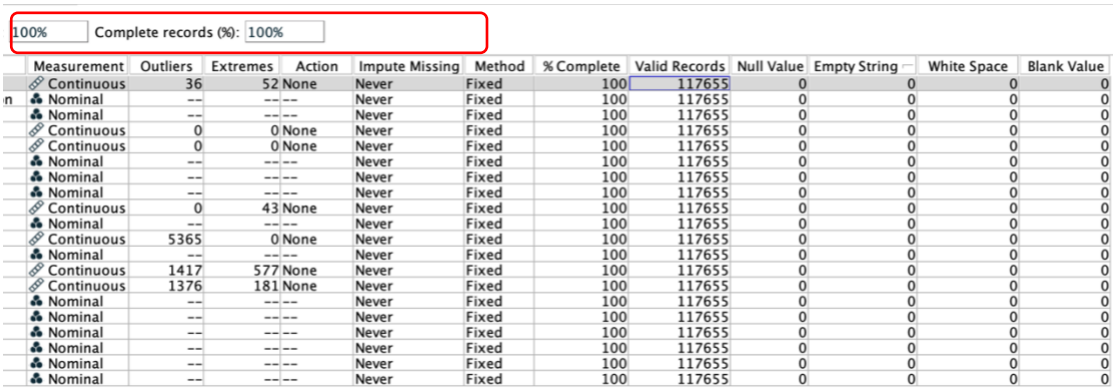


Figure 5: Flight Status Values Proportion

Considering the above, the mentioned attributes dropped using the filter function, this decision was justified by assessing the limited impact of discarding these columns on the overall research objectives.

4.2.2.1.3. Assessing Data Quality

As a result of removing missing values, and dropping insignificant columns the data quality was enhanced to reach 100%. The number of data records is 117,655 as shown in the below data quality tab;



Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value	Empty String	White Space	Blank Value
Continuous	36	52	None	Never	Fixed	100	117655	0	0	0	0
Nominal	--	--	--	Never	Fixed	100	117655	0	0	0	0
Nominal	--	--	--	Never	Fixed	100	117655	0	0	0	0
Continuous	0	0	None	Never	Fixed	100	117655	0	0	0	0
Continuous	0	0	None	Never	Fixed	100	117655	0	0	0	0
Nominal	--	--	--	Never	Fixed	100	117655	0	0	0	0
Nominal	--	--	--	Never	Fixed	100	117655	0	0	0	0
Nominal	--	--	--	Never	Fixed	100	117655	0	0	0	0
Continuous	0	43	None	Never	Fixed	100	117655	0	0	0	0
Nominal	--	--	--	Never	Fixed	100	117655	0	0	0	0
Continuous	5365	0	None	Never	Fixed	100	117655	0	0	0	0
Nominal	--	--	--	Never	Fixed	100	117655	0	0	0	0
Continuous	1417	577	None	Never	Fixed	100	117655	0	0	0	0
Continuous	1376	181	None	Never	Fixed	100	117655	0	0	0	0
Nominal	--	--	--	Never	Fixed	100	117655	0	0	0	0
Nominal	--	--	--	Never	Fixed	100	117655	0	0	0	0
Nominal	--	--	--	Never	Fixed	100	117655	0	0	0	0
Nominal	--	--	--	Never	Fixed	100	117655	0	0	0	0
Nominal	--	--	--	Never	Fixed	100	117655	0	0	0	0
Nominal	--	--	--	Never	Fixed	100	117655	0	0	0	0

Figure 6: Cleaned Data Record

4.2.3. Approaches for Detecting Outliers

Outliers a data point that differs from other data points to the extent that raises questions about how it is produced [17]. As shown in Figure (6), the dataset contains outliers in the attributes "Expect Pieces" and "Expect Weight". The method followed in this research is to detect those outliers and extremes by using statistical methods.

4.2.3.1. Removing Outliers and Extreme Datapoints

The derive node is used to create a new field from the attributes "Expected Pieces" and "Expect Weight" and compute the below condition and if the resulted value or datapoint varies from the computed formula then it will be considered as outlier or extreme data points, where *k* in the below formula is real number. If *k* =3, SPSS modeler considers the data point as an outlier, and if *k* =5, SPSS modeler considers the data point as an extreme outlier.

$$Value \geq mean(Expected\ Pieces) + k \times STD(expectedpicices)$$
$$or\ Value \leq \text{mean(expected pieces)} - k \times STD(expected\ pieces)$$
$$Value \geq \text{mean(Expect Weight)} + k \times STD(\text{Expect Weight})$$

$$\text{or Value} \leq \text{mean(Expect Weight)} - k \times STD\text{(Expect Weight)}$$

If the value for the above equation is within the considerable range of mean and standard deviation then it will be tagged in the new field as 0, if the value is outside that range, then 1 and considered as an outlier or extreme data point. Then, the node selected includes only the value 0, to remove the unusual pattern in the two mentioned attributes.

4.2.3.2. Anomaly Detection Model

The anomaly detection method is an unsupervised method used to detect outliers or unusual data points that deviate from normal behavior within the dataset. The model helped to generate an anomaly index filed within the customs inspection dataset and assigned each observation to the anomaly index to measure the deviation of the observation within its cluster. The model then identified each observation in a new field based on the calculated index as F or T where F refers to normal data points and T refers to outliers [18]. Finally, the outliers are removed from the dataset using the select function to include only "F" normal data points.

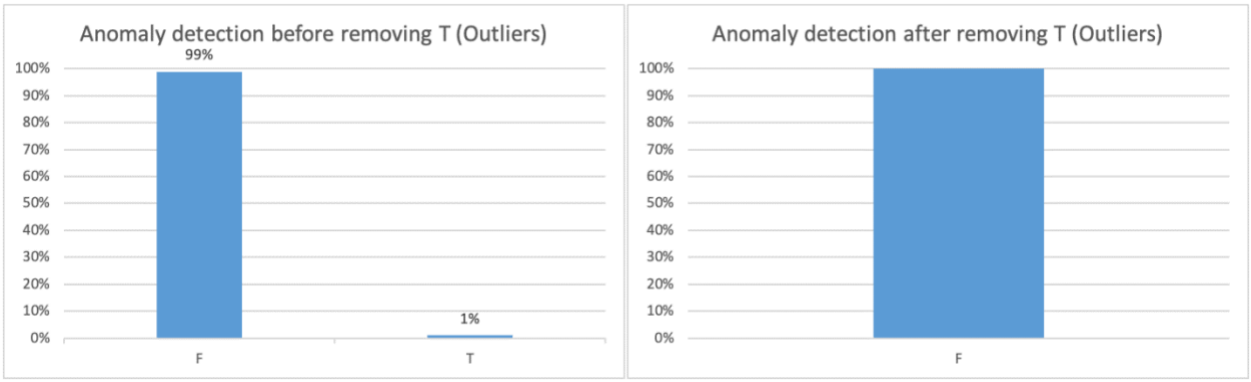


Figure 7: Anomaly detection

4.2.4. Significance Analysis Between Second Inspection and Other Variables

It's important to understand the significance of different attributes with the target variable "Second Inspection". This will help to capture the relationship between the attributes, guide the feature engineering process, and achieve the research objectives of building a prediction model with high accuracy in predicting imported shipments as illicit or non-illicit.

4.2.4.1. U-test comparison between numerical and nominal Attributes

The Mann-Whitney U test is utilized to test the null hypothesis for the significance of the two attributes "Expect Pieces" and "Expect Weight". The justification for using such a method is that

it is a non-parametric statistical test which is preferred when having a nominal dependent variable, and the distribution of that data is not normally distributed [19].

- **Null hypothesis:** There will be no differences in the central tendency between the independent variables and the dependent variable.

4.2.4.1.1. Numeric Attributes and Second Inspection

The SPSS Statistics modeler was utilized to test the null hypothesis using the Mann-Whitney U test. The results are shown in the below table for the comparison between the independent variable "Expected Pieces" and the target variable "Second Inspection" where the significant threshold is ≤ 0.050;

Table 3: Hypothesis Testing 1

Null hypothesis	Test Type	Significance
The distribution of "**Expect Pieces"** is the same across categories 0 and 1 of "Second Inspection"	Independent-Sample Mann-Whitney U test	< 0.001

Table 4: Hypothesis testing 2

Null hypothesis	Test Type	Significance
The distribution of "**Expect Weight"** is the same across categories 0 and 1 of "Second Inspection"	Independent-Sample Mann-Whitney U test	< 0.001

The above result summarizes that the test performed on the customs inspection dataset with 97918 records (Total N), and the result of the Mann-Whitney U test indicated a P-value less than 0.001 showcasing strong evidence against the null hypothesis. This comparison result shows a statistically significant difference between the independent variable "Expect Pieces" and "Expect Weight" with the target variable "Second Inspection" proving the relation between the attributes indicating to rejection of the null hypothesis.

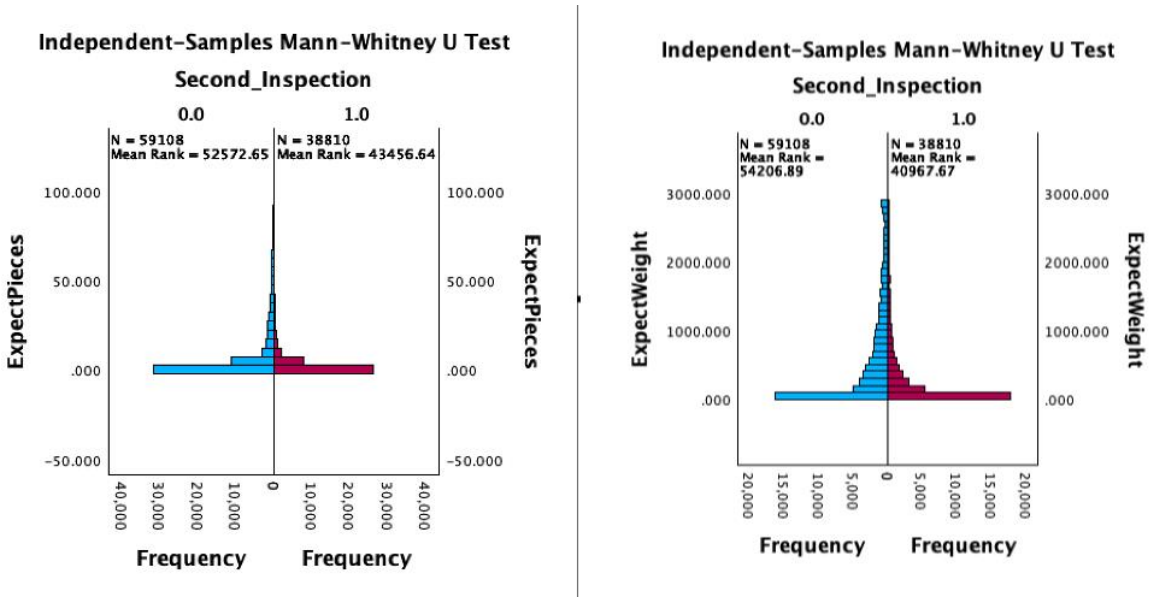


Figure 8: "Expected Pieces" and "Expect Weight" with the target variable "Second Inspection"

4.2.4.2. Significance between nominal and nominal attributes

Using a cross-tab table and histogram visualization to capture the relationship between two categorical attributes, such as "Flight Type" and "Second Inspection", or "ULD Type" and "Second Inspection" to understand the density of the target variable in a concerned attribute. The significance of the mentioned attributes will be tested using the Chi-square, which is utilized to determine the relationship between two nominal variables [20]. The Chi-square result can determine the null hypothesis that there is no significant relationship between the independent mentioned variables and the dependent variable "Second Inspection". If the Chi-square test value is high this will indicate that there is no evidence to support the null hypothesis therefore will be rejected.

- "ULD Type": The below visualization indicated that ULD contains more illicit shipments than Bulk. The Chi-Square is 5,216.788, the degree of freedom is 1, and the calculated p-value is <.00001 indicating a significant relation between the mentioned attribute with the target "Second Inspection" rejecting the null hypothesis.

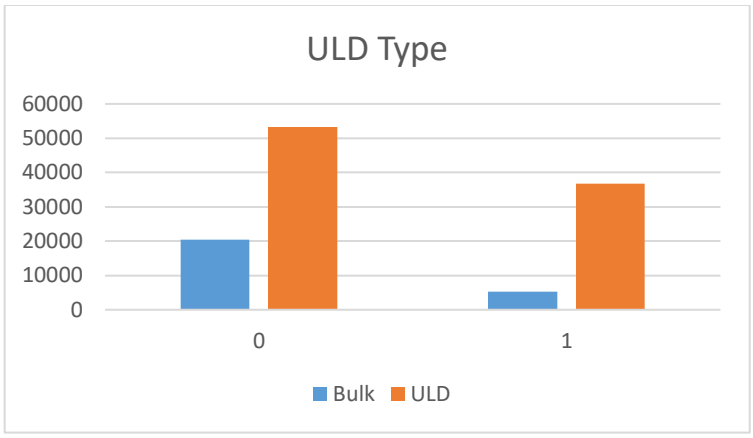
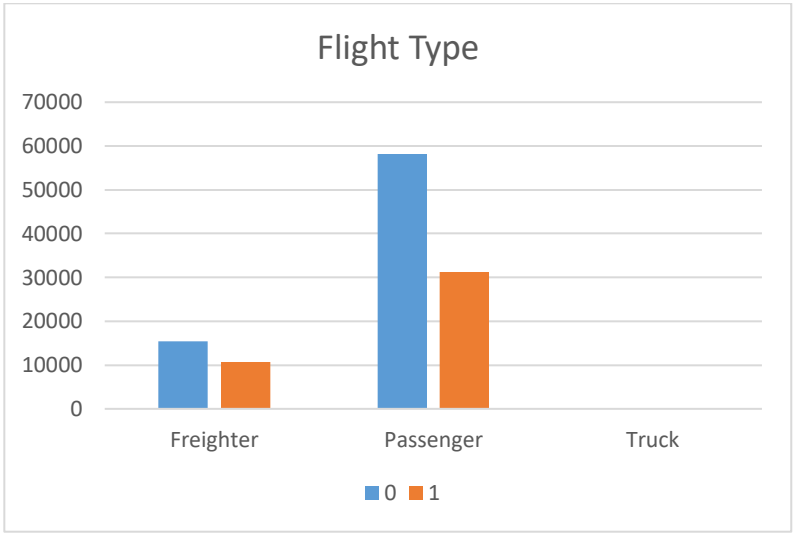


Figure 9: ULD Type

- "Flight Type": The below visualization indicated that passenger flights contain illicit shipments more than freighter and are almost non-existent if the type is truck. The Chi-Square is 513.549, the degree of freedom is 2, and the calculated p-value is <.00001 indicating a significant relation between the mentioned attribute with the target "Second Inspection" rejecting the null hypothesis.



-

Figure 10: Flight Type

- "Dest": The cross-tab values show that there are 3 main busy destinations that contain a high count of illicit shipments detected by customs, which are R, J, and D, and the remaining destinations have view counts. The Chi-Square is 1,101.072, the degree of freedom is 19, and the calculated p-value is <.00001 indicating a significant relation

between the mentioned attribute with the target “Second Inspection” rejecting the null hypothesis.

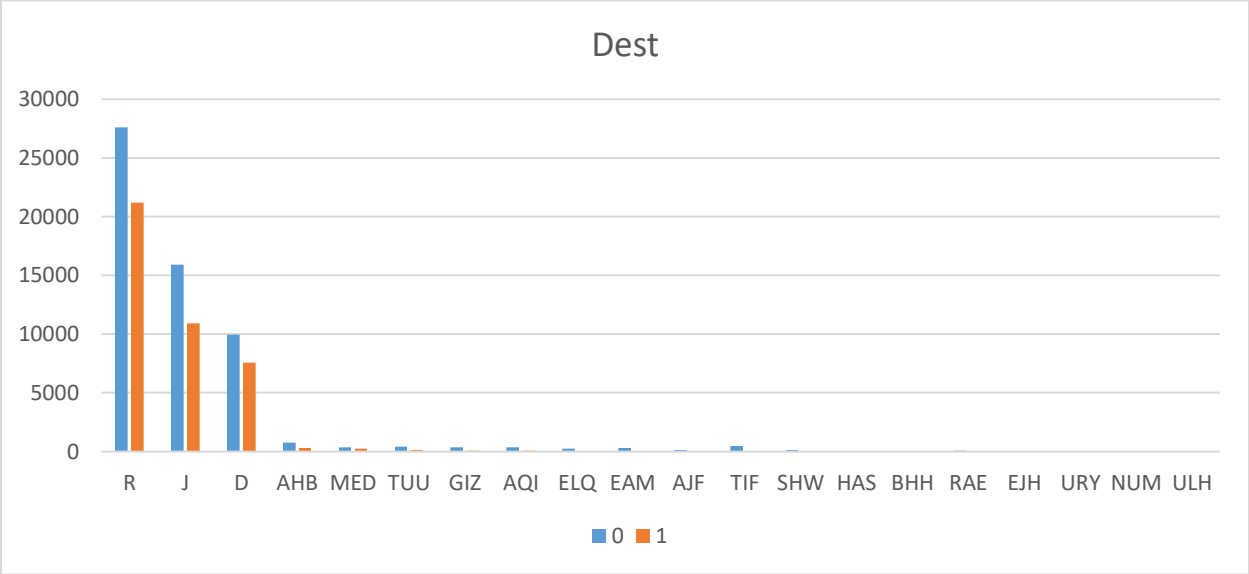


Figure 11: Dest

4.2.5. Features Engineering

The book explained the importance of features engineering to have the required result out of the model as they treat the features of the dataset as input to learn from [21]. In this section, new features will be extracted and constructed from preexisting ones to be more convenient for the machine learning model.

4.2.5.1. Feature Extraction (Date-Time Extraction)

The attribute “Flight date” was split into three attributes “Flight day”, “Flight month”, and “Day Flight or Night”. The flight day represents the day of the week to understand the density of the importation concerning the weekdays or weekends. The flight month is the month of the flight as numeric. Finally, day flight or night is a nominal variable indicating if the flight is scheduled PM or AM time.

- The flight day indicated that Friday, Saturday, and Sunday had more illicit shipments detected than the remaining days. The Chi-Square is 1,064.064, the degree of freedom is 6, and the calculated p-value is <.00001 indicating a significant relation between the mentioned attribute with the target “Second Inspection”.

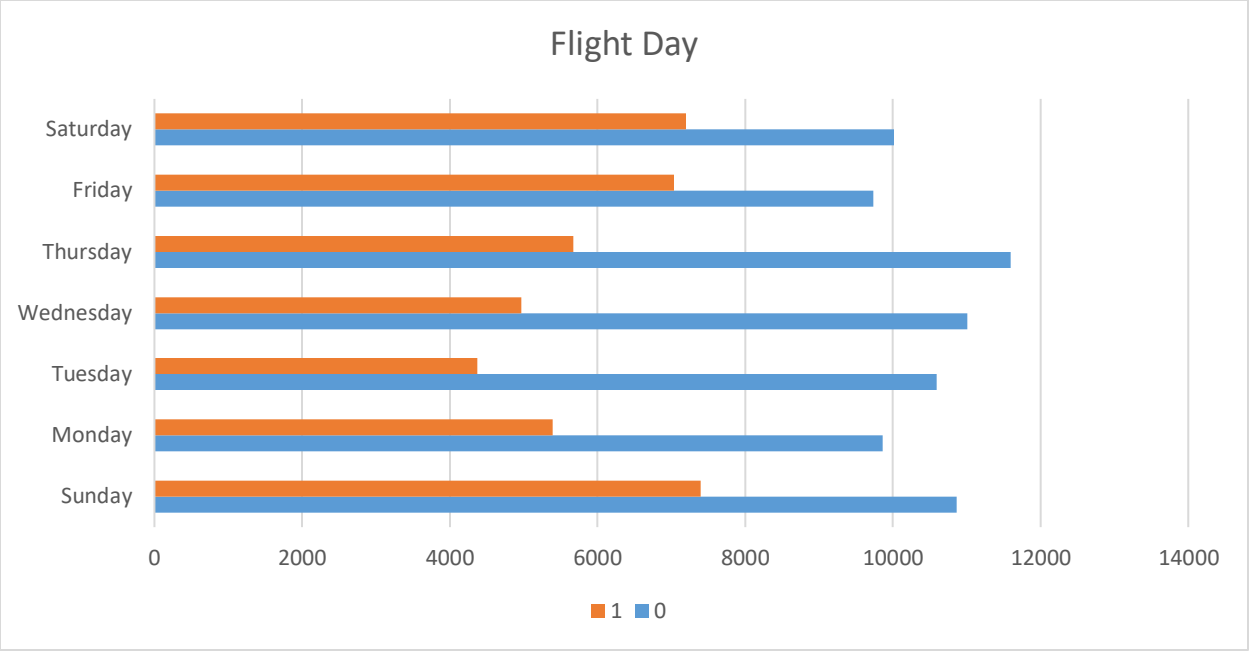


Figure 12: Flight Day

- Flight month indicated that months 2,3,4,11, and 12 had a higher count of illicit shipments detected than the remaining months. The Chi-Square is 219.268, the degree of freedom is 11, and the calculated p-value is <.00001 indicating a significant relation between the mentioned attribute with the target "Second Inspection".

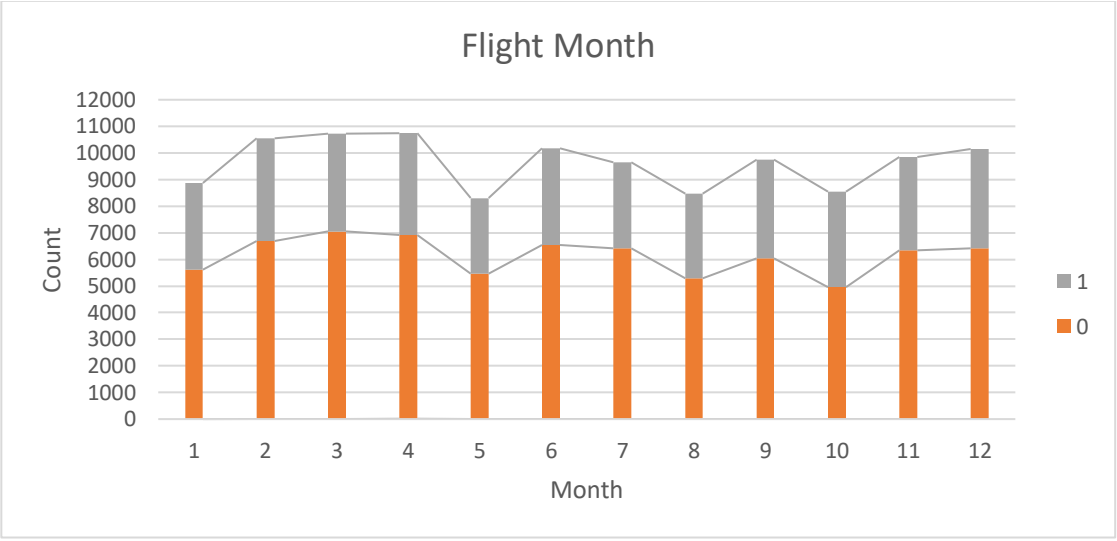


Figure 13: Flight Month

- Day flight or night indicated that the shipments imported at night (PM) had a higher count than in the morning time. The Chi-Square is 1,893.538, the degree of freedom is 1, and the calculated p-value is <.00001 indicating a significant relation between the mentioned attribute with the target “Second Inspection”.

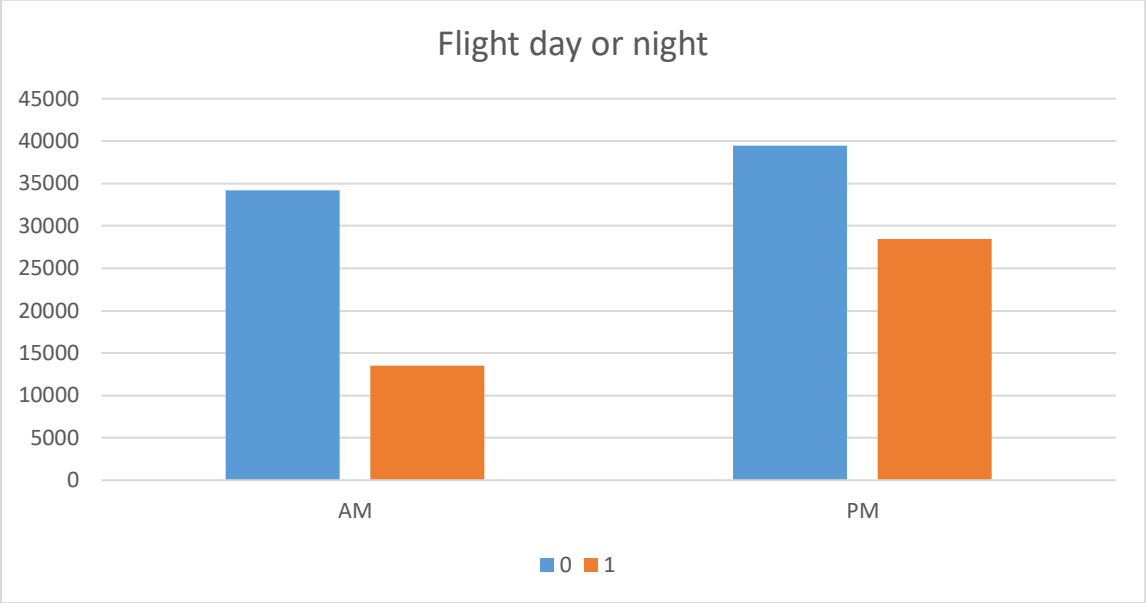


Figure 14: Flight Night or Day

4.2.5.2. Category Aggregation (Binning categorical features)

The main objective of category aggregation is to reduce the noise ratio in features that have many categories. [33] addresses this problem by using feature engineering to create a new feature that aggregates multiple categories into a single category to reduce the noise in the dataset and enable the model to learn from meaningful signals in the data. This method is utilized on the attribute "Dest,” which has 27 different categories shown in Figure (11), with only 3 unique categories that contain the major counts, while the remaining categories have negligible counts. Therefore, a new field was engineered from the initial attribute and named "DestGroup,” where the categories R, J, and D are the same and all the remaining categories are grouped in a single group as "Others.”

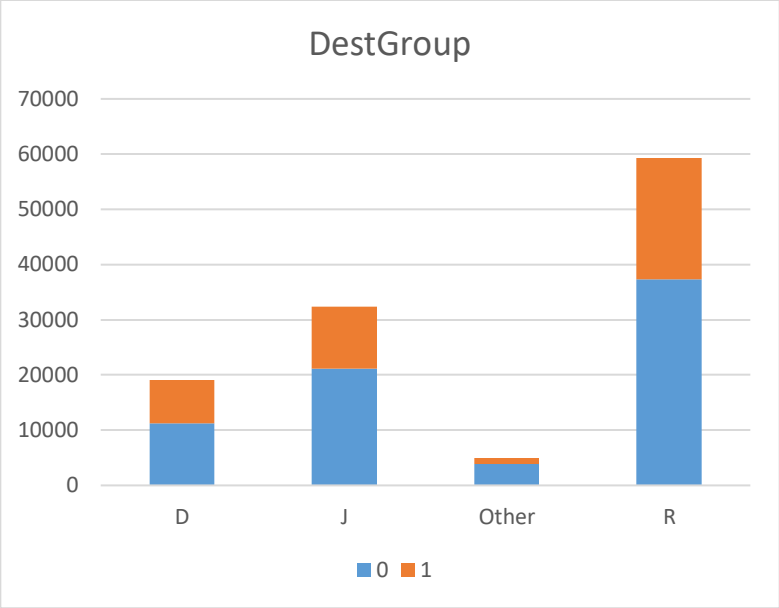


Figure 15: Dest Group

4.2.5.3. Features Transformation

New features were constructed using domain knowledge from the cargo company where they demonstrated that the attribute "Shclist" is important because it provides information about the content of each airwaybill as one AWB may contain different subfamily groups. The explanation of each subfamily is provided by the company, for example, if the AWB contains the code "DGR" inside the "Shclist" value it means that the content of the imported shipment is categorized or subjected as dangerous goods. Therefore, new fields were engineered from the preexisting attribute:

- **"NB_PRODUCTS":** This attribute quantifies how many products are in each airway bill or container.
- **"NB_DGR_PROUCTS":** This attribute quantified the proportion of dangerous goods among the total number of products.
- **"DGR_INDEX":** Divided the attribute "NB_DGR_PROUCTS" on "NB_PRODUCTS" to calculate the index of the dangerous goods inside the airway bill or container.
- **"OriginLevelOfRisk":** Numeric cross tab between the origin and risk indicators 0 if the container has no dangerous products and 1 if the container contains a risky product, the risk carrier is computed as the value of the number of risky containers divided by the total number of containers.

$$Risk\ origin = \frac{Number\ of\ containers\ with\ dangerous\ goods}{Total\ number\ of\ containers}$$

- **"Level of risk Origin":** Cross-tab between origin and second inspection and indicators of frequency to categorize the origin as no risk_origin, low risk_origin, medium risk_origin, high risk_origin, and very high risk_origin.

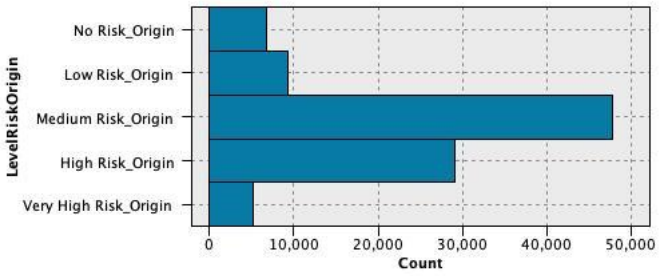


Figure 16: Level of Risk Origin

- **"Risk Carriere":** Numeric cross tab between the carrier and risk indicators 0 if the container has no dangerous products and 1 if the container contains a risky product, the risk carrier is computed as the value of the number of risky containers divided by the total number of containers.

$$Risk\ carriere = \frac{Number\ of\ containers\ with\ dangerous\ goods}{Total\ number\ of\ containers}$$

- **"Carriere Risk level":** Cross-tab between Carriere and second inspection and indicators of frequency to categorize the Carriere as Car_Risk0=no risk, Car_Risk1=low risk, Car_Risk2= medium risk, and Car_Risk3=high risk.

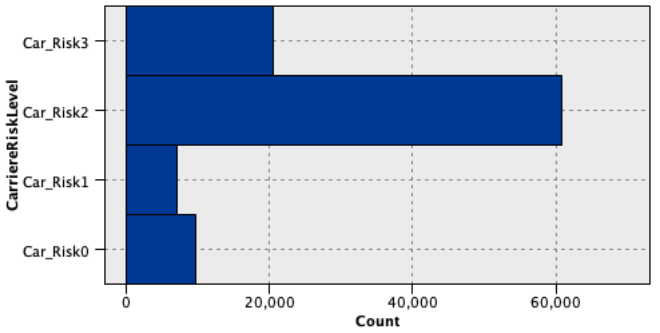


Figure 17: Carriere Risk Level

4.2.6. Summary of the Final Dataset

The final dataset is the enhanced version of the customs inspection dataset and this has been achieved by discarding the missing values from the attributes "Flight Number," “Aircraft Type,” “ULD Type” and “Second Inspection” due to the values being missing completely at random. The decision was also made to drop unnecessary columns "Import Status” and “Flight Status” due to their insignificant, resulting in the enhancement of the data quality from 77.27% to 100%. The robustness of the data improved by removing extreme data points using statistical computation and anomaly detection nodes. Furthermore, exploratory data analysis techniques were applied to gain insights into the customs inspection data structure, distributions, and relationships among variables with the support of statistical testing to prove the hypothesis. Finally, feature engineering techniques were implemented to enhance the dataset’s interpretability and predictive capacity before the deployment of the classification model.

4.3. Machine learning model development

This research aims to delve into the urging challenges faced by customs personnel in cargo and develop a prediction model that predicts the likelihood of imported shipments being illicit or non-illicit using classification models. The project begins with exploring the customs inspection data and understanding its attributes and records. The exploration continues with further examining the data quality and performing summary statistics to grasp the data patterns and behaviors. Following that, data processing for missing values and outliers, as emphasized in Chapter 3 of the literature review, handling missing values and discarding outliers using statistical techniques are necessary before deploying the model. Furthermore, the transformation of the customs inspection data using feature engineering to modify preexisting attributes, aggregate, split, and join to improve the model performance as proven in different studies. Moreover, capturing the relationship between different attributes using statistics and data visualization. Utilizing algorithms to determine the important features such as random trees and feature selection. Finally, the deployment of classification models to two different scenarios, and measure the accuracy of each model for both scenarios, and calculate precision, recall, F1-measure, ROC curve, and AUC to select the best classification model.

4.3.1. Feature Importance Analysis

Feature selection is the process of selecting only relevant features in the dataset to build the machine learning model. There are many benefits to selecting the relevant features, like reducing the training time, avoiding repetition and high dimensionality, and helping to improve the machine learning model's performance [34]. There are different techniques to determine the relevant features; one example of these techniques is feature importance. The important features in the customs inspection dataset have been determined in two different scenarios; the first scenario applied was automatic feature selection using a feature selection algorithm in the IBM SPSS Modeler and by the random tree algorithm. The second scenario applied was to select features advised by the subject matter expert in Cargo to gain a business perspective and determine important features.

4.3.1.1. Detailed explanation of the chosen Input

4.3.1.1.1. Automatic Feature Selection

The dataset may contain hundreds of features, and choosing which features are important and used as model input can be a time-consuming and complex task. In this scenario, two feature selection techniques are proposed by the IBM SPSS modeler. The first technique is the feature selection algorithm, which is used to identify the most important features in the customs inspection dataset in three steps. The first step is screening which discarding unnecessary columns that are not important or have missing values or unbalanced variations. The second step is ranking, where the remaining useful input will be sorted and ranked as per its importance. Finally, the feature selection algorithm will select only the important features to be used as input for the model [23]. The features are shown in the below table;

Table 5: Feature selection Node

The feature selection node	
Target	Input
Second_Inspection	AircraftType, ExpectPieces, ExpectWeight, ULDType, FlightType, LevelRiskOrigin, Flight_Day, Flight_Month, DayFlightorNight, RiskCarriere, CarriereRiskLevel

The random tree algorithm is used to select the most important features where it provides sample data through the use of bootstrap sampling with replacement. Using a random selection of predictors, the best predictor is utilized to divide a tree node. The sample data from customs inspection data is used to grow a tree model and it randomly selects part of the predictors and uses the best one to split a tree node. This process is repeated when splitting each tree node [24]. The features are shown in the below table;

Table 6: The Random Tree

The random tree	
Target	Input
Second_Inspection	ExpectPieces, ExpectWeight, ULDType, LevelOfRisk, LevelRiskOrigin, Flight_Day, Flight_Month, DayFlightorNight, RiskCarriere, CarriereRiskLevel, DestGroup

4.3.1.1.2. Business Feature Selection

In this scenario the input selected by the subject matter expert in customs, and the features selected are as below;

Table 7: Business Feature Selection

The random tree	
Target	Input
Second_Inspection	ExpectPieces, ExpectWeight, ULDType, FlightType, NB_PRODUCTS, LevelRiskOrigin, Flight_Day, Flight_Month, DayFlightorNight, CarriereRiskLevel,

4.3.2. Detailed explanation of the chosen machine learning algorithms

The selection of the model is essential, in this section, to determine the most effective machine-learning model to optimize customs efficiency in detecting illicit shipments, five different models will be applied and their performance will be compared using performance evaluation metrics. The models are as follows;

- ◊ **Neural Network:** The model was selected due to its major ability to learn from the data and capture complex and non-linear relationships. Also, it can produce accurate predictions on unseen or new data by identifying different data patterns and learning from them, then using these patterns to perceive similar patterns in the new data. The model can also handle data volatility, allowing it to uncover hidden patterns in the data without configuring assumptions about the behavior of the data, which is an important advantage in the customs world [25].
- ◊ **Logistic regression:** The model was selected because it is simple and easy to comprehend for the binary classification problems. The model is also able to provide an assumption of the probability of a shipment being illicit or non-illicit based on the input of the variables with no requirements of high computation and tuning [26].
- ◊ **Decision Tree:** The model was selected due to its ability to interpret the human decision-making process. The model is simple, easy to explain, and can capture the relationship between continuous and categorical variables. It predicts the outcome based on the relations of the other variable rather than the input, thus, the model is beneficial in determining the main and important feature in the customs inspection dataset [27].
- ◊ **XGBoost:** The model was selected due to its high ability in classification problems where the model learns from decision trees and applies regularization techniques to increase and improve its performance this process is known as ensemble learning. Other benefits of the XGBoost model such as the ability to handle large datasets with efficient computation, feature importance analysis, and accurate prediction results.
- ◊ **Linear Support Vector Machine (LSVM):** The model was selected due to its effectiveness in separating two different classes and capturing linear patterns in the data especially when the margin of separation is clear between the classes. The model can handle the high dimensionality of the data and perform well because of its generalization performance.

4.3.3. Validation and Testing Procedures

This section extends the data preparation process by splitting the customs inspection dataset into a training set and a testing set. After that, to achieve the full learning capacity of the model in the training set, a balancing technique is needed to reach the project's objective of building a prediction model to detect shipments as illicit or non-illicit with great accuracy. Finally, different classification models will be deployed to both scenarios such as neural network, decision tree,

logistic regression, XGBoost, and LSVM. The models will be evaluated and the main aim is to identify the most effective model for detecting imported shipments.

4.3.3.1. Data partition

The dataset was split into a training set and a testing set, where 70% of the data is used to train the machine learning model, and 30% of the data is utilized to test and validate the model performance on unseen data. The percentage of the data partitioned is shown in the below figure;

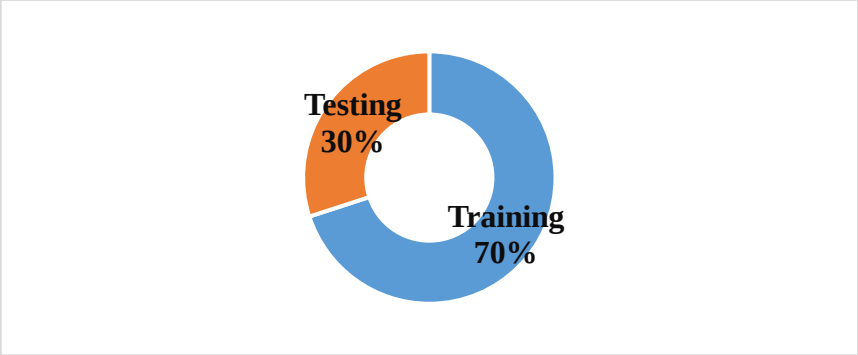


Figure 18: Data Partition

4.3.3.2. Balancing

The method followed in this research to balance the dataset is an under-sampling method. The over-represented class in the target variable "Second Inspection" is 0 and instances were randomly removed to develop more balanced data for the modeling.

Figure 19: Data set Before Balancing

The balance directives for the under-sampling in the portioned training dataset are as follows;

Table 8 Dataset Balancing Factor

Factor	Condition
0.644	Second Inspection=0
1	Second Inspection=1

The distribution of the second inspection feature in the training dataset is balanced as shown below;

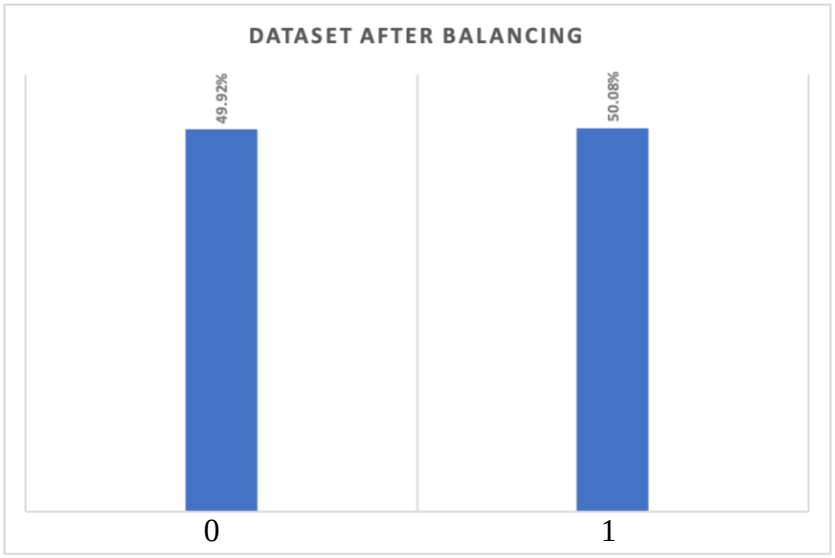


Figure 20: Dataset After Sampling

Balancing the training set is very important before the classification model because the model learns from the training dataset. If the data is imbalanced, the model learns very well about the majority class while the minority class is underrepresented. Therefore, the customs inspection dataset is balanced before the model stage to increase accuracy and reduce biases towards one class [35].

4.3.3.3. Evaluation metrics used to assess model performance

There are different performance metrics to measure the classification model and used to evaluate the model performance on unseen data, such as confusion matrix, accuracy, precision, recall, F-measure, and AUC [28]. The details of each metric are as follows;

Confusion matrix: It is a table of predicted values and actual values used to measure the performance of the classification model, the table is demonstrated below;

Actual Values

Predicted Values	Positive (1)	Negative (0)
Positive (1)	TP	FP
Negative (0)	FN	TN

Accuracy: This metric measures how often the classification model accurately predicts, where the correctly predicted TP (True positive) and TN (True negative) will be added and divided by the total number of predictions TP, TN, FP (False positive), and FN (False negative).

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

Precision: This metric is important if the cost FP is high, it measures how many accurate cases turned out to be positive by dividing the TP by the sum of TP and FP.

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

Recall: This metric is important if the cost of FN is high which concerns this case of customs control, it measures how well the classification model accurately predicts shipment being illicit. If the model has a high recall value this means that the model will capture illicit shipments reducing the cost of shipment passing undetected. Recall measures the TP and divide it by the sum of TP and FN.

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

F-measure: This metric is measured by taking the harmonic mean of the recall and precision scores and the result values will be between the range of 0-1 if 0 means poor performance and higher means better performance. F-measures help to determine if the classification model has

balanced performance and can correctly identify true positives and the proportion of actual positives it can grasp.

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Area Under the Curve (AUC): This graph helps to measure how well the classification model can separate between two classes in this case illicit and non-illicit. The best value for AUC is 1 or almost 1.

ROC (Receiver Operating Characteristic): The graph is used to measure the performance of the classification model and to find out how it takes decisions among different values of thresholds or certainty. It is originated by figuring out the True Positive Rate (TPR) versus the False Positive Rate (FPR) on the x and y-axis of the graph.

- Sensitivity highlights the items those are correctly classified

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

- FN rate highlights shipments those are incorrectly classified by the model

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

- Specificity demonstrates non-illicit shipments those are correctly classified

$$Specificity = \frac{TN}{TN + FP}$$

- FP rate measures illicit shipment incorrectly classified by the model

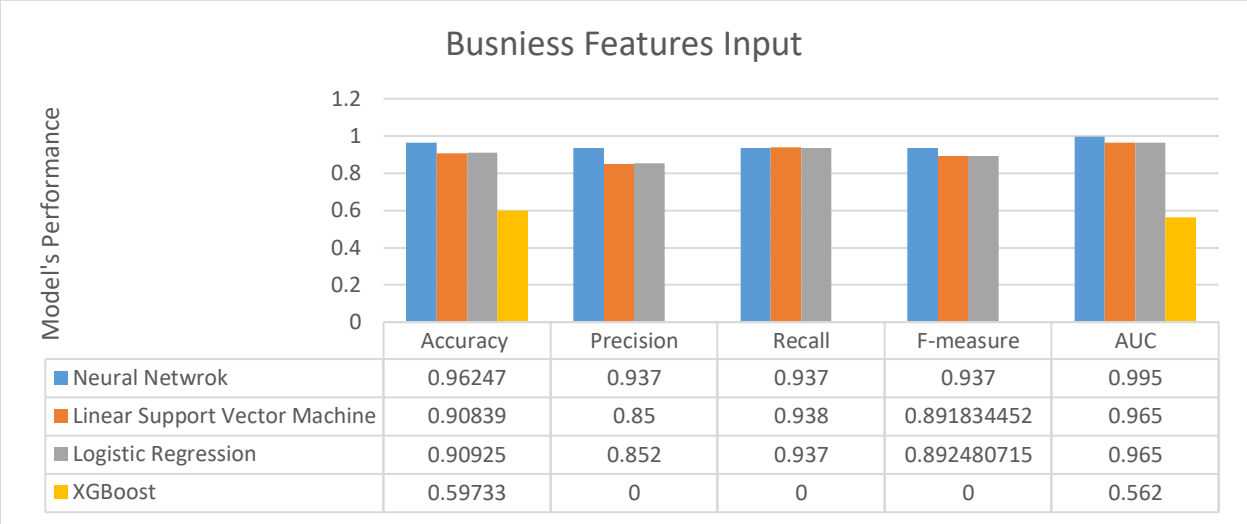
$$FP\ rate = \frac{FP}{TN + FP} = 1 - specificity$$

Youden's J statistic: This method looks for the point of cut-off of X (TPR) and Y (FPR) at the corresponding threshold values. First of all, it not only considers the classifier sensitivity and specificity but also gives the overall accuracy of the model in a single score. If the value of Youden's J is high, it indicates a better discriminatory capacity of the test, meaning it is able to classify both illicit shipments and non-illicit shipments appropriately [36].

4.3.4. Results

4.3.4.1. Presentation of the Experimental Results

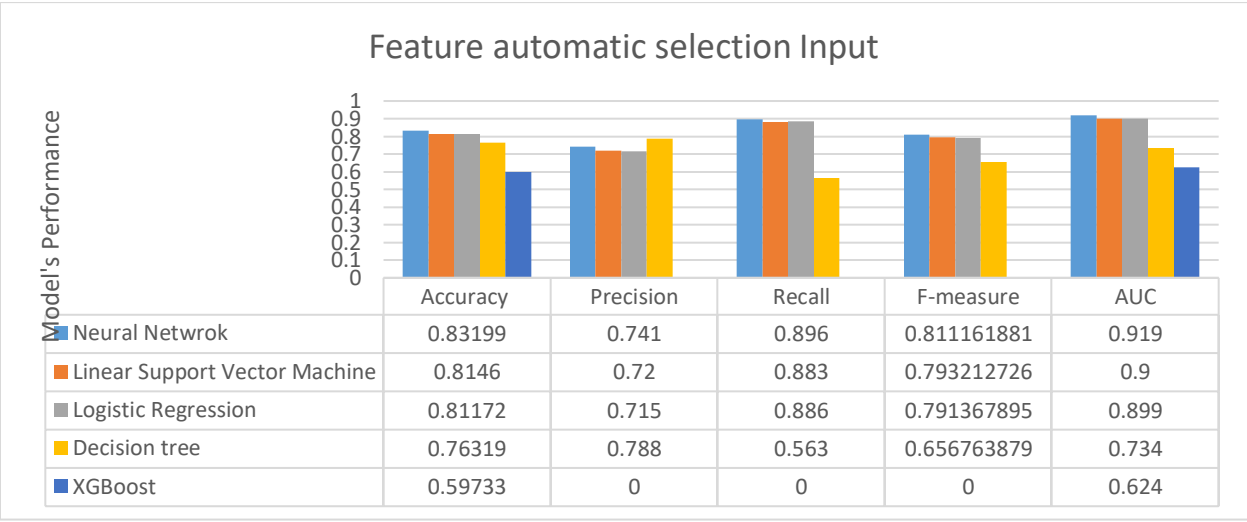
Business Features selection models performance as shown below figures; the classification decision for the illicit container was made with a probability of the model higher than 0.5.



	Accuracy	Precision	Recall	F-measure	AUC
Neural Netwrok	0.96247	0.937	0.937	0.937	0.995
Linear Support Vector Machine	0.90839	0.85	0.938	0.891834452	0.965
Logistic Regression	0.90925	0.852	0.937	0.892480715	0.965
XGBoost	0.59733	0	0	0	0.562

Figure 21: Business Features Input

Feature automatic selection models performance as shown below figures; the classification decision for the illicit container was made with a probability of the model higher than 0.5.



	Accuracy	Precision	Recall	F-measure	AUC
Neural Netwrok	0.83199	0.741	0.896	0.811161881	0.919
Linear Support Vector Machine	0.8146	0.72	0.883	0.793212726	0.9
Logistic Regression	0.81172	0.715	0.886	0.791367895	0.899
Decision tree	0.76319	0.788	0.563	0.656763879	0.734
XGBoost	0.59733	0	0	0	0.624

Figure 22: Feature Automatic Selection Input

Key Insights:

1. **Logistic Regression Model:**
    - Initialized 11 predictors for automatic features selection and 10 predictors as business features selection variables for shipment prediction.
    - Accuracy reached 81.172% and 90.925% in automatic feature selection and business feature selection, respectively.
    - Established an Area Under Curve (AUC) of 0.899 in automatic features selection and 0.965 in business features selection.
    - Precision, recall, and F1-measure are as an automatic feature selection: 0.715, 0.886, and 0.791. For business feature selection 0.852, 0.937, and 0.89 respectively.
2. **Decision Tree Model:**
    - The model adopted a minimalist structure with only 3 predictor variables.
    - Obtained a lower overall accuracy score (76.319%) in the process of automatic feature selection than the logistic regression model.
    - Achieved an area under curve (AUC) of 0.734
    - These results demonstrated a precision of 0.788, recall of 0.563, and F1-measure of 0.656.
3. **XGBoost Model:**
    - The automatic feature selection used 11 predictor variables, while the business used 10 feature variables for shipment prediction.
    - Reached the lowest overall accuracy of 59.733% for both automatic features selection and business features selection.
    - Shows an area under the curve (AUC) of 0.624 and 0.562 in automatic features and business features selection, respectively.

4. **Linear Support Vector Machine Model:**

    - The 11 predictors were considered in automatic feature selection and the 10 predictors in business feature selection for the prediction of shipments.
    - Achieved an accuracy of 81.46% in automated features selection while 90.839% in business features selection.
    - AUC (Area Under Curve) is calculated to be 0.9 for automatic features selection and 0.965 for business features selection.
    - Precision, recall, and F1-measure were for feature selection were found to be 0.72, 0.883, and 0.793. Business feature selection was 0.85, 0.938, and 0.891 respectively.

5. **Neural Net Model:**

    - Applied 11 predictors for automatic feature selection and utilized 10 predictors for business feature selection to predict shipments.
    - Attained the greatest overall accuracy of 83.199% in automated feature selection and 96.247% in business feature selection.
    - Performs the best among the models and fetches the automated features selection AUC of 0.919 and the business features selection AUC of 0.995.
    - Precision, recall, and F1-measure are 0.741, 0.896, and 0.811 for automatic feature selection. In terms of business feature selection of 0.937, 0.937, and 0.937, the balanced trade-off is distinctly the case in this model.

On the whole, the results showcased in the graphs reveal that features chosen by subject matter experts in customs performed better than the automated feature selection technique. The neural network model proved to outperform logistic regression, XGBoost, LSVM, and Decision tree at 96.247% with the highest accuracy. The evaluation of the precision, recall, and f-measures showed the same result, precision, and recall of 0.937, which testifies that the model is well-balanced. Another important point is that the NN model's AUC was 0.995, which is very high for the effective separation of two classes. Additionally, with regard to the trade-offs between model

complexity, performance metrics, and operational practicality that need to be given attention, the implementation of the most favorable customs shipment prediction model will be an easy task.

Comparative Analysis:

1. **Overall Accuracy:**
    - With an overall accuracy of 96.247%, the neural network model outperformed the LSVM (90.839%) and the logistic regression model (90.925%).
    - The accuracy demonstrates the model's capacity to correctly categorize shipments that are both positive and negative.
2. **Area Under Curve (AUC):**
    - The neural network model has the greatest AUC value (0.995), demonstrating how well it can differentiate between shipments that are illegal and those that are not.
    - With AUC values of 0.965, respectively, come logistic regression and LSVM.
3. **Precision and Recall:**
    - With a True Positive Rate of 0.938, the LSVM demonstrated the highest recall, which emphasizes its potency in detecting actual illicit shipments.
    - With recall levels of 0.937, the neural network and logistic regression models were almost equal.
    - The neural network leads with a precision of 0.937. Precision quantifies the percentage of real positives among all shipments forecasted as positive.
4. **F1-Measure:**
    - The neural network model has the highest F1-Measure (0.937), a harmonic mean of precision and recall that shows a balance between the two.
    - F1-Measure values of 0.89 are attained by both LSVM and Logistic Regression, respectively.

In brief, logistic regression, LSVM, and neural networks have been shown to be good across the metrics; however, the neural network seems to be the best at the recall sub-metric due to its ability to correctly identify illicit shipments. In spite of a favorable model demonstrating balanced precision and recall, it gives the best discriminating power with the highest AUC. On that note, the ideal choice for optimizing customs efficiency and predicting contraband shipments is to use a neural network model. The decision depends on the model's recall performance metric and its ability to accurately classify shipments carrying high risks of being illicit that must be detected and, on the other hand, to reduce the costs of normally passing undetected shipments that are very important in the customs environment.

4.3.4.2. Predictor Importance of the Neural Network Model

The important predictors can be decided based on the reduction variance of the target attribute "Second Inspection" from each predictor by using (sensitivity analysis) which finds, under a given set of assumptions, the effects of varying values of a single independent variable on the specific dependent variable [29]. The predictors given in the graph above are by the formula below;

$$Y = f(x_1, x_2, \dots, x_k)$$

Where Y=Second Inspection (Target)

X_j predictor, where j=1..k

K=The number of predictors

F()=Function factor

In accordance with the sensitivity measure described below, the predictors are ranked;

$$S_i = \frac{V_i}{V(Y)} = \frac{V(E(Y|X_i))}{V(Y)}$$

where the unconditional output variance is denoted by V(Y), the expectation operator E in the numerator requires an integral over X_{-i} ; that is, all factors except X_i, (Y), the expectation operator E in the numerator requires an integral over X_(-i); that is, all factors except X_i. The variance operator V then implies an additional integral over X_i. Thereafter, the normalized sensitivity is used to calculate predictor importance.;

$$VI_i = \frac{S_i}{\sum_{j=1}^{k} S_j}$$

The concept of sensitivity in depth and ranked the predictor as a way of measuring sensitivity, which is the crucial component of any combination of interaction and non-orthogonality among predictors [29]. The description below gives the normalized significance of the major characteristics of the neural network model;

$$VI_{LevelOfRiskOrigin} = 0.29$$

$$VI_{NB\_PRODUCTS} = 0.24$$

$$VI_{ExpectWeight} = 0.13$$

$$VI_{CarriereRiskLevel} = 0.09$$

$$VI_{ExpectPieces} = 0.08$$

$$VI_{ULD\ Type} = 0.06$$

$$VI_{Flight\ Type} = 0.04$$

$$VI_{Flight\_month} = 0.03$$

$$VI_{Flight\_day} = 0.03$$

$$VI_{DayFlightorNight} = 0.01$$

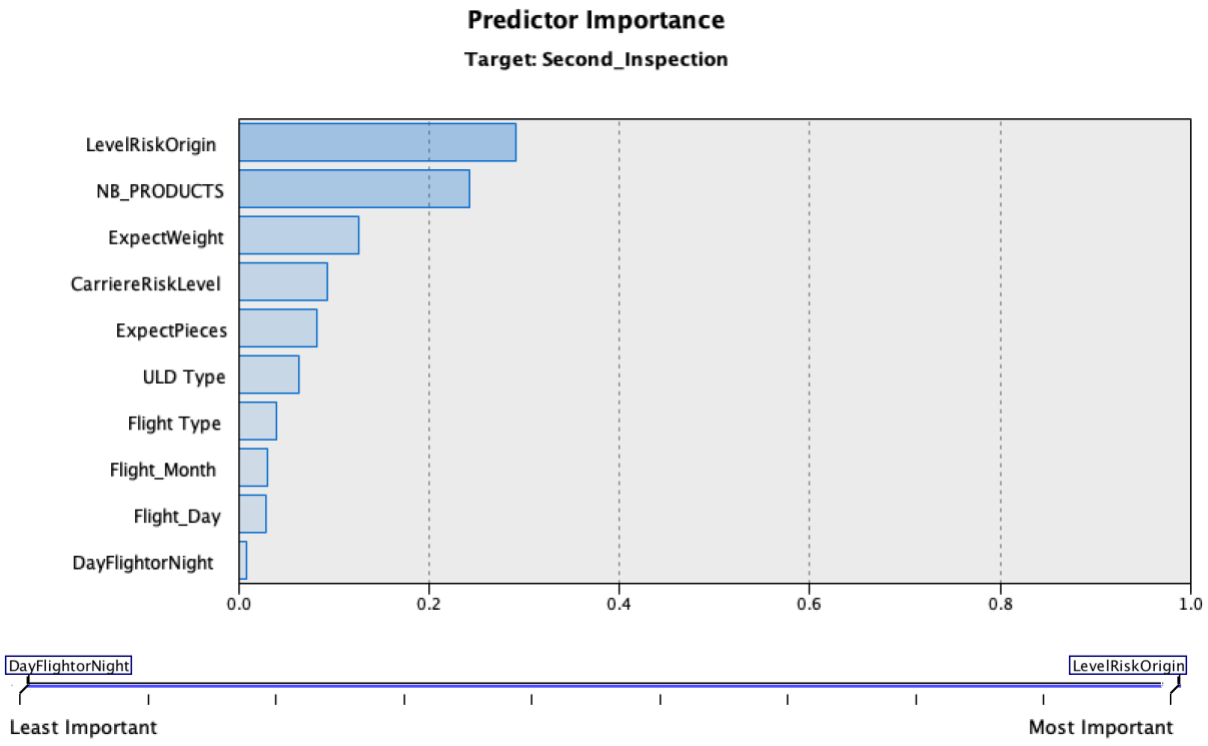


Figure 23: Predictor Importance

These factors are taken into account from the most significant sensitivity factor to the least significant one.

LevelOfRiskOrigin: The origin risk level imparts information that can be utilized to predict the shipments for control purposes. Analysis of the degree of sensitivity related to risk level origin can allow customs to carry out the most impactful inspections first.

NB_PRODUCTS: This evaluates the number of goods each airway bill or container carries and when changes happen in trade volume, regulations, and operational factors it may affect container control outcomes. For example, the enormous amount of devices increases the risk of security holes. Sensitivity analysis of what is the number of products inside the container revealed that customs offices would have an opportunity to allot resources thoughtfully and rearrange inspection priorities so as to lessen risks.

ExpectWeight: In the container weighing context, load dimension is often a hint of likely payload contents, action, or non-compliance. The authority will feel curious about its particular mass and bear examination further. The side effect of container weight sensitivity analysis is that it helps

tricky customs to detect whether the weight of the container is anomaly or not and channel the inspections based on the detection, in this way smuggling activities will be reduced.

In conclusion, application in customs inspection: it is necessary to understand the influence on the level of risk origin, NB_PRODUCTS, and volume weight which in turn enhance customs control. Sensitivity analysis will help customs to discover those factors that predict inspections the most, and this way they will be able to create predictive models that are based on the most influential predictors. By applying this strategy, time and funds be are saved, the delay for the clearance of freight is reduced and security measures are improved.

4.3.4.4. Visualization of the Most Important Predictors

4.3.4.4.1. Nominal features

The importance of each of the predictors has been measured with the help of a cross-tab for the calculation of the chi-square for proving the statistical significance of the below predictors-

- Chi-square value that is "ULD Type" is 4797.807, 1 is df value, and the p-value is less than .001 that reflects the statistical significance.

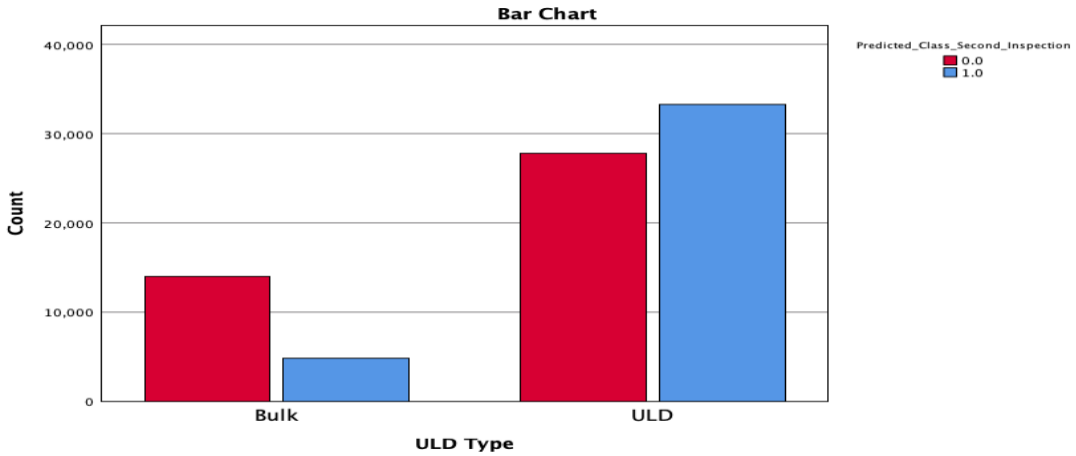


Figure 24: Predictor ULD Type

- Chi-square value of "Flight Type" is 334.423, 1 is the value of df, and the p-value is less than .001 which shows the statistical significance of the variable.

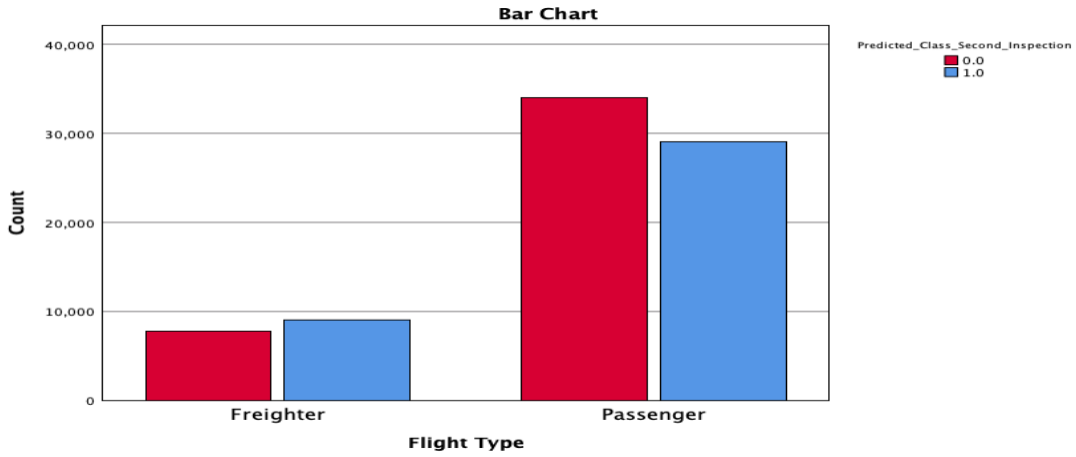


Figure 25: Predictor Flight Type

- "LevelRiskOrigin" chi-square value is 24785.745, the df value is 4, and the p-value is less than .001 which proves the statistical importance of this attribute.

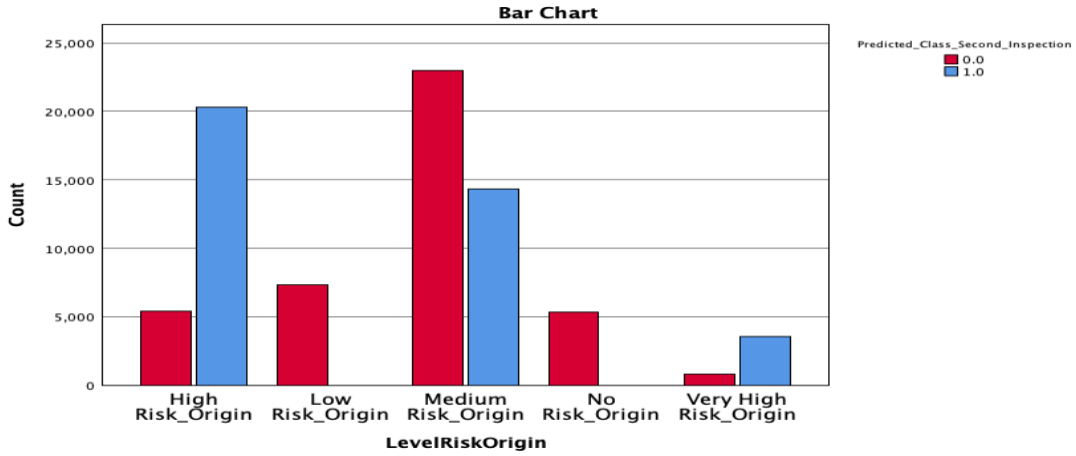


Figure 26: Predictor Level Risk Origin

- "Flight_Day" chi-square value is 847.893, the df value is 6, and the p-value is less than .001 which proves the statistical importance of this attribute.

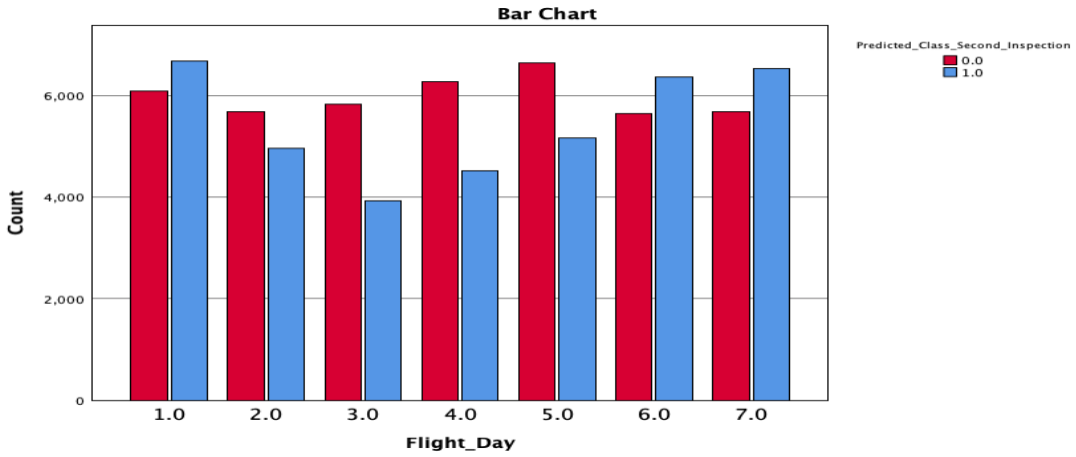


Figure 27: Predictor Flight Day

- "Flight_month" chi-square value is 174.102, the df value is 11, and the p-value is less than .001 which proves the statistical importance of this attribute.

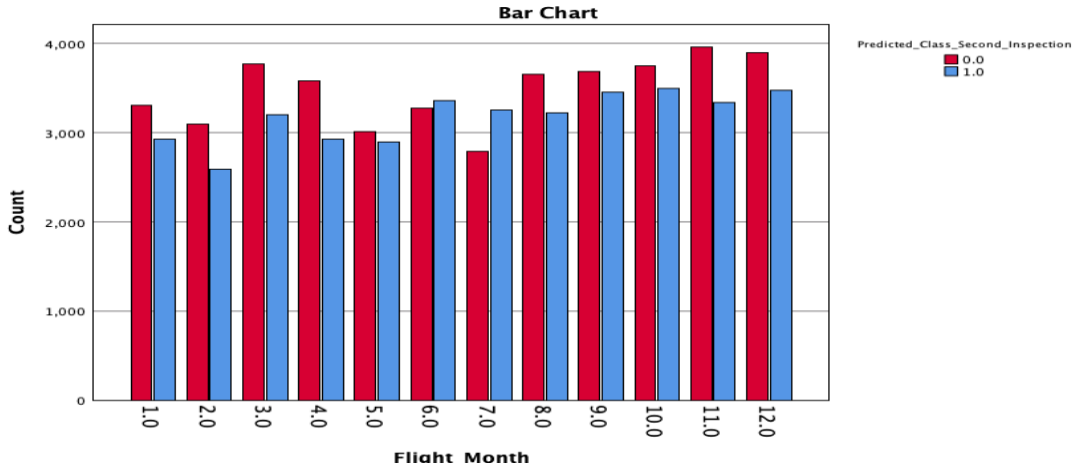


Figure 28: Predictor Flight Month

- "DayFlightorNight" chi-square value is 1534.484, the df value is 1, and the p-value is less than .001 which proves the statistical importance of this attribute.

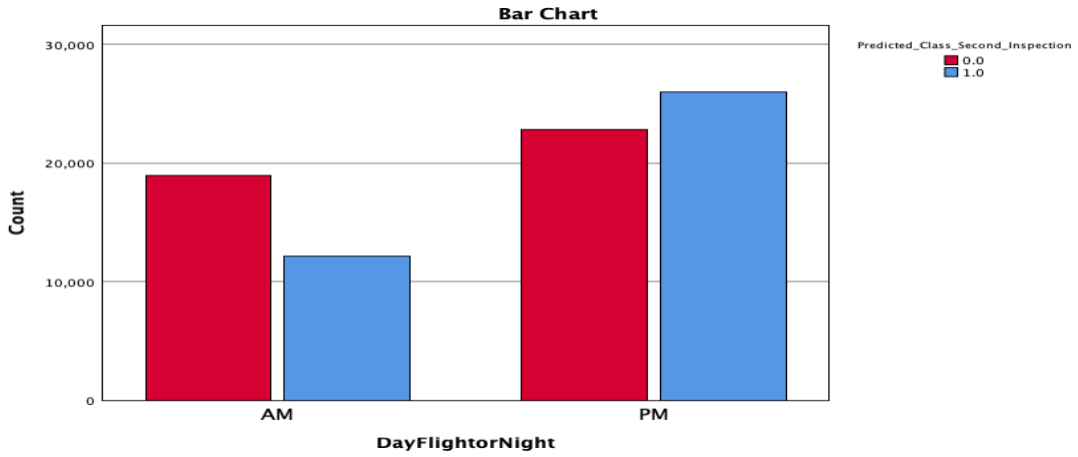


Figure 29: Predictor Day Flight Night

- "CarriereRiskLevel" chi-square value is 11726.295, the df value is 3, and the p-value is less than .001 which proves the statistical importance of this attribute.

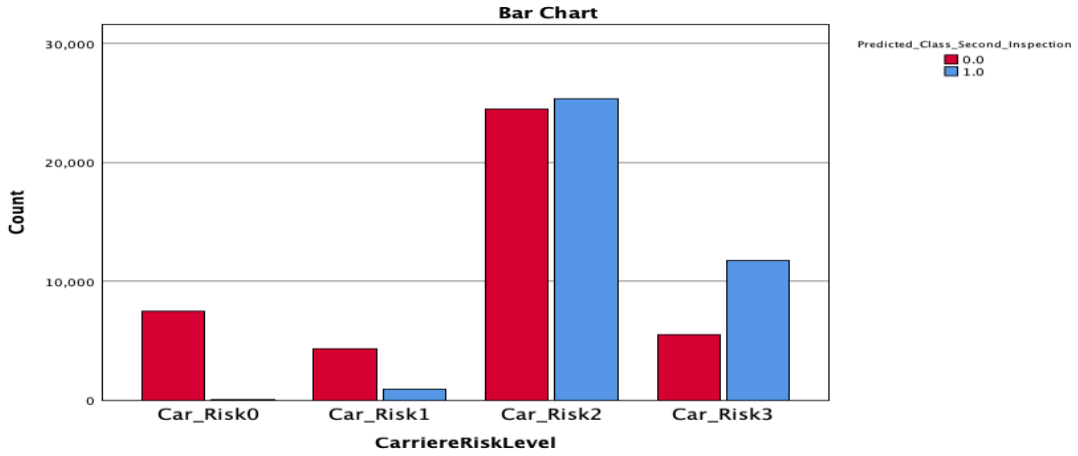


Figure 30: Predictor Carrier Risk Analysis

4.3.4.4.2. Numeric features

The Mann-Whitney U test is utilized to test the null hypothesis. The results are shown in the below table for the comparison between the predictors "Expect Weight", "Expected Pieces", and "NB_PRODUCTS" and the predicted class "Second Inspection" where the significant threshold is ≤ 0.050;

Table 9: Hypothesis Testing 3

Null hypothesis	Test Type	Significance
The distribution of **"ExpectPieces"** is the same across categories of Predicted_Class_Second_Inspection.	Independent-Sample Mann-Whitney U test	< 0.001
The distribution of **"ExpectWeight "** is the same across categories of Predicted_Class_Second_Inspection.	Independent-Sample Mann-Whitney U test	< 0.001
The distribution of **"NB_PRODUCTS"** is the same across categories of Predicted_Class_Second_Inspection	Independent-Sample Mann-Whitney U test	< 0.001

The above result illustrates that the p-value is less than 0.001 which showcases strong evidence against the null hypothesis. This result shows a statistically significant difference between the mentioned numeric predictors and the predicted class "Second Inspection" proving the relation therefore rejecting the null hypothesis.

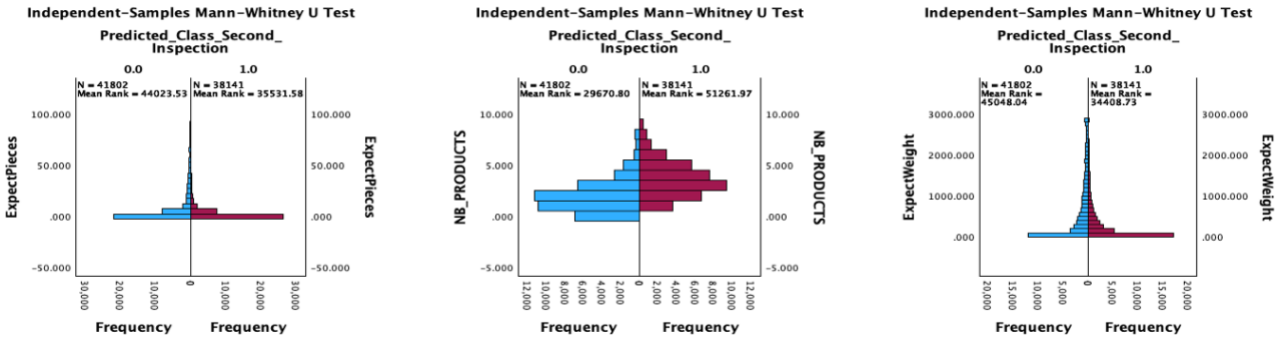


Figure 31: Numeric Predictor Whitney U Test

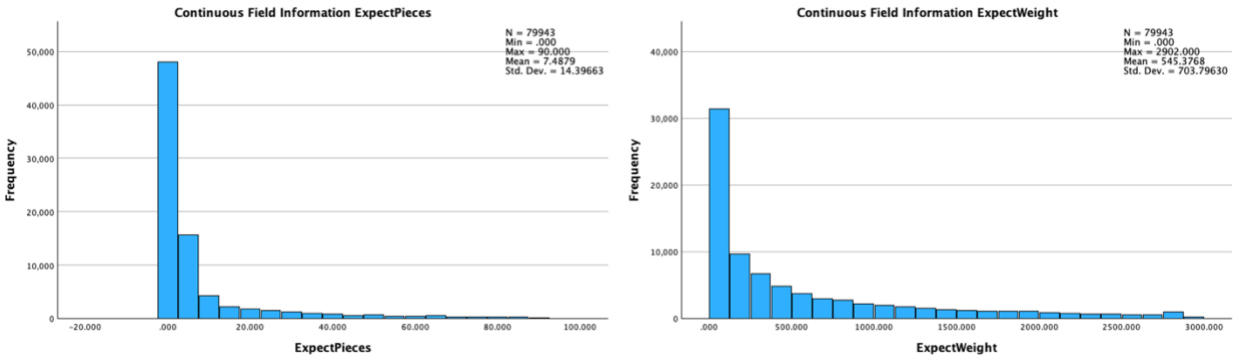


Figure 32: Numeric Predictor Distribution 1

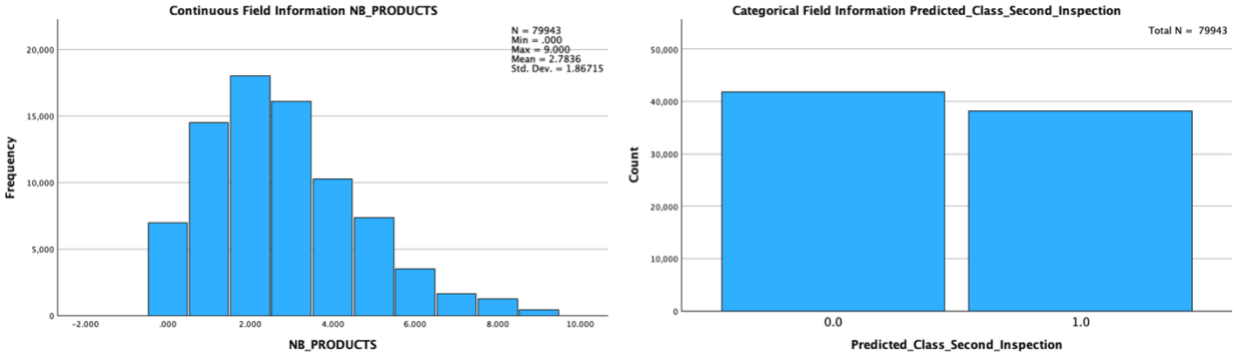


Figure 33: Numeric Predictor Distribution 2

4.3.4.6. Analysis of Confusion Matrix

The neural network model's confusion matrix is displayed in the table below, which aids in clarifying how well the algorithm detects shipments;

Table 10: Confusion Matrix

Actual	Predicted	
	1	**0**
1	96.10%	3.90%
0	2.70%	97.30%

True Positive (TP): 96.10% of the actual illicit shipments were accurately predicted by the model to be illicit.

False Negative (FN): 3.90% of the actual illicit shipments were mispredicted by the model as non-illicit.

True Negative (TN): 97.30% of the actual non-illegal shipments were accurately predicted by the model to be non-illegal.

False Positive (FP): The model miscalculated the proportion of actual non-illegal shipments to be unlawful, estimating 2.70%.

The sensitivity is calculated as follows in order to determine the shipments that are accurately categorized as unlawful.;

$$\text{sensitivity} = \frac{96.10\%}{96.10\% + 3.90\%} = 96.10\%$$

Additionally, the specificity is assessed in order to determine the accurately categorized non-illicit shipments;

$$\text{specificity } = \frac{97.30\%}{97.30\% + 2.70\%} = 97.30\%$$

4.3.4.7. Analysis of ROC Curves and AUC Values

The ROC curve (Receiver Operating Characteristic) is a graph that helps to determine how well the classification model is performing and how it takes decisions at different thresholds or certainty [30]. The ROC curve contains two axis X and Y, the Y axis shows the classification model's ability to classify illicit shipments (True positive rate) correctly in other words, it measures the sensitivity of the model, and the higher the value the better the model detects illicit shipment for further inspection by customs. The X-axis shows the (1-specificity) of the classification model where it measures if the model classifies non-illicit shipments as illicit and requires customs to further inspect the shipment and wasted time and resources (False positive rate), the lower the value of FP in the X axis the better which lower value indicate that the model correctly classifies illicit shipments. The AUC (Area Under the Curve) is an evaluation metric as well where it measures the classification model performance on differentiating between two classes and it is used as a summary of the ROC curve. If the value of AUC is 1 or near 1 this means the model can differentiate between positive and negative classes in this case, illicit or non-illicit, and if the value

of AUC is 0.5 this means the model classification at random and not accurate, and finally if 0 means the model fail to differentiate between positive and negative classes.

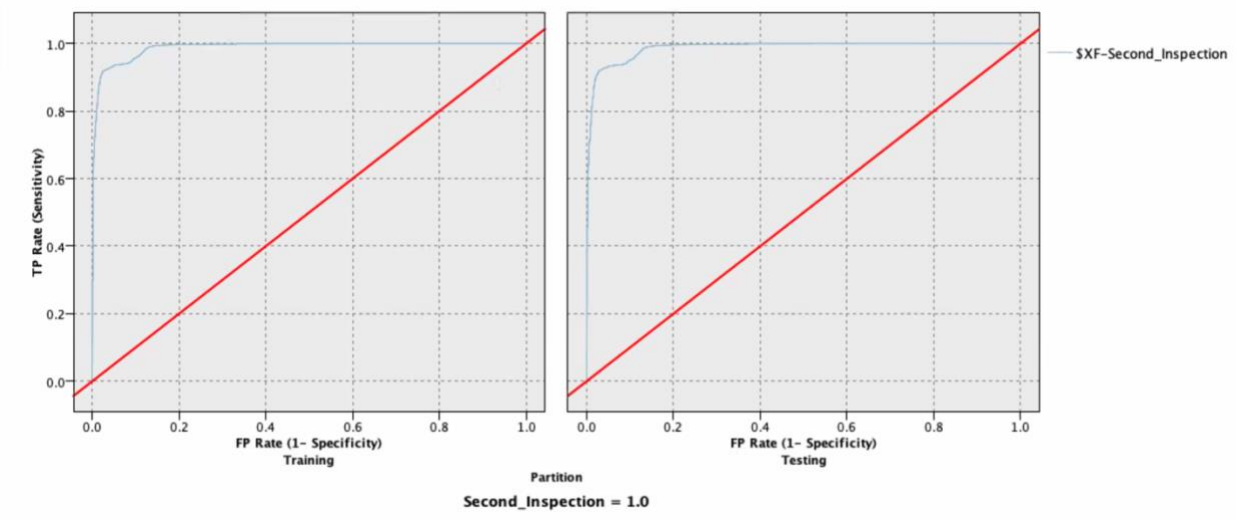


Figure 34: Neural Network ROC

The neural network model's ROC curve, seen above, provides a graphical depiction of performance as the discrimination threshold changes. The True Positive Rate (TPR, or Sensitivity) and False Positive Rate (FPR, or 1 - Specificity) are compared in this research experimental set at a threshold of 0.5. This curve illustrates the fine balance between correctly categorized negatives and reliably detecting genuine positives, with each data point representing a different threshold level. Through Youden's J statistic, the ideal cut-off point can be determined. The ROC curve's maximum vertical distance from the point (X, Y) on the diagonal (random line) is maximized by this statistic, which also improves the net accurate classification through maximization of the difference between sensitivity and (1 - specificity).

The present analysis employs a selective selection of data points from the below table to demonstrate the sensitivity, 1-specificity, and related Youden value:

Table 11: Cherry pick data points

Sensitivity	1-Specififty	Youden
0.1993	0.0007	0.1986
0.6053	0.0029	0.6024
0.8267	0.013	0.8137

0.92295	0.0468	0.87615
0.9556	0.107	0.8486
0.993	0.1551	0.8379
0.6121	0.003	0.6091
0.9105	0.0255	0.885
0.9963	0.2002	0.7961

Based on the aforementioned results, the cutting point is identified as point A = (X;Y) = (0.92295;0.0468), which corresponds to the Youden value of 0.87615. The True Positive Rate (TPR) or Sensitivity at the selected threshold is represented by the X-coordinate (0.92295), which is the percentage of true positive instances among all actual positive instances that the model properly recognized. The Y-coordinate (0.0468), on the other hand, represents the False Positive Rate (FPR) or the complement of Specificity at the selected threshold. This represents the percentage of false positive cases that the model mistakenly categorized as positive out of all real negative cases. The results mentioned above demonstrate the neural network model's dependability and efficiency to recognize both illegal and lawful cargo, which could enhance customs productivity throughout inspection and better utilize available resources.

Chapter 5- Discussion

This research aims to achieve main four objectives; the first objective was to build a machine-learning model that can predict or classify imported shipments as illicit or non-illicit. This objective was successfully reached by gaining a comprehensive understanding of the customs inspection data, enhancing its quality, exploring important factors using EDA analysis, and data preparing the data for the model. The classification models applied were neural network, logistic regression, decision tree, LSVM, and XGBoost. However, the neural network achieved the best result in terms of performance in detecting shipments with a great accuracy of 96.247% exceeding all models applied. These results have also been confirmed by **(Singh et al., 2023)** which they used a neural network model and achieved an accuracy of 86.9% for customs detection, and this proves the model reliability in this research achieving higher model performance. The second objective of this research is to improve the accuracy and efficiency of targeting high-risk shipments, and this objective was fulfilled by applying different techniques such as features engineering and performing feature importance analysis. Engineering new features helped the model to capture the pattern and learn from the data along with important analysis where an automatic feature selection algorithm is used, a different technique was utilized to improve the model's accuracy which is the advice from customs expert matter for feature selection. This helped to enhance the accuracy and efficiency of targeting high-risk shipments which was seen in the automatic feature selection algorithm where the neural network model accuracy is 83.199% and improved after following the domain knowledge of the customs matter expert for feature selection improved the neural network model accuracy to be 96.247%. The mentioned results restate the paper of **(Nargesian et al., 2017)** that one of the contributions of feature engineering is the ability to improve the predictive performance of the model. The feature selection by customs experts helped in selecting the relevant features, reducing the training time, avoiding repetition and high dimensionality, and helping to improve the machine learning model's performance as mentioned in the article published by **(Wu, n.d.)**. According to **(Xin Zhou, 2019)** paper, discusses the importance of cost-sensitive classification, which accounts for the expenses associated with incorrectly categorizing declarations posing a high risk this leads to the third objective of this research, which is to increase customs productivity and reduce the false positives in customs inspection which can be a waste of resources and time that can be utilized on illicit shipments. The

neural network model built has a high precision rate of 0.937 which indicates that the model accurately predicts illicit shipments which increases the customs productivity and lowers the false positive case percentage which is less than 6%. The final objective is to maintain effective security and trade flow, this is achieved by the neural network model where the recall result was .937 and this result was confirmed by **(Singh et al., 2023)** where the model recall in their research was .92. This shows that the model's ability to capture illicit shipments reduces the cost of shipments passing undetected as non-illicit and this ensures the security of the nation and maintain effective trade flow.

Chapter 6- Conclusion

6.1. Conclusion

This research aims to address the urging challenges faced by customs personnel in cargo. The surge in imports has strained customs resources and created inefficiencies, prompting the need for data-driven solutions to increase customs productivity and maintain effective security by improving the detection of illicit shipments. Through the application of different machine learning models to predict imported shipments and extensive evaluation and analysis, it was found that the neural network model developed as the most efficient algorithm, accomplishing a superb accuracy of 96.247%. This result was obtained through major efforts in data cleansing, feature transformation, preparation, and feature importance analysis and selection to gain the highest reliability and robustness of the classification model.

6.2. Contributions to Knowledge

By incorporating the newest scientific knowledge, this research helps out the urgent situation of heightening customs performance for the purpose of detecting smuggling. The proof of the superiority of the neural network model among other classification models, as well as the accent of data quality and transformation techniques, in the present research, presents valuable remarks in the fight to upgrade and automatize the detection of contraband imports in the customs arena.

6.3. Practical Implications

The findings have the application, among customs functions, for the inspection. The machine learning model acts as a helpful weapon for detecting illegal imports with high precision for trade security as well as collective social peace. With the help of classification models, customs can automate and improve the declaring procedures through staff allocation.

6.4. Recommendations and Future Work

The study though provides deep insight, the need for more research and improvement is still there. The future study can explore using more data sources and/or predictor variables like shipment tracings and historical compliance data to further enhance the model's prediction precision. First of all, there is a necessity to involve customs experts in order to check the created model and to prove its relevance for the purposes of customs declaration practices. Similarly, the validation of the model utilizing actual data and tracking the real-time performance of the model are effective approaches to test its effectiveness and adaptability in actionable settings. Furthermore, the ethical questions raised about using AI in customs should be considered for future studies. Responsible adoption of AI necessarily comprises talks related to the privacy, openness, as well as the fairness of applying classification algorithms implicitly to given contexts. In the mentioned context, researchers ought to act proactively when it comes to people's trust and adoption of AI-based solutions by dealing with these ethical issues.

References/Bibliography

1. Abu Dhabi, C. (2023). Indicators of Foreign Trade Through the Ports of Abu Dhabi Emirate. https://www.adcustoms.gov.ae/Foreign-Trade-Statistics

2. Constanta, I., & Stefan, Z. (2012). RISK MANAGEMENT - A NEW PRIORITY SYSTEM CUSTOMS AND ITS CONSEQUENCES. https://mpra.ub.uni-muenchen.de/39352/

3. Chen, L., & Ma, Y. (2015). A Study of the Role of Customs in Global Supply Chain Management and Trade Security Based on the Authorized Economic Operator System. 5(2).

4. Han, D., Zhang, J., Wan, Z., & Liao, M. (2023). Dynamic Weights Based Risk Rule Generation Algorithm for Incremental Data of Customs Declarations. Information, 14(3), 141. https://doi.org/10.3390/info14030141

5. Camossi, E., Dimitrova, T., & Tsois, A. (2012). Detecting Anomalous Maritime Container Itineraries for Anti-fraud and Supply Chain Security. 2012 European Intelligence and Security Informatics Conference, 76–83. https://doi.org/10.1109/EISIC.2012.39

6. González García, I., & Mateos Caballero, A. (2021). A Multi-Objective Bayesian Approach with Dynamic Optimization (MOBADO). A Hybrid of Decision Theory and Machine Learning Applied to Customs Fraud Control in Spain. Mathematics, 9(13), 1529. https://doi.org/10.3390/math9131529

7. Singh, K., Tsai, Y.-C., Li, C.-T., Cha, M., & Lin, S.-D. (2023). GraphFC: Customs Fraud Detection with Label Scarcity (arXiv:2305.11377). arXiv. http://arxiv.org/abs/2305.11377

8. Regmi, R. H., & Timalsina, A. K. (2018). Risk Management in customs using Deep Neural Network. 2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS), 133–137. https://doi.org/10.1109/CCCS.2018.8586834

9. M, H., & M.N, S. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. International Journal of Data Mining & Knowledge Management Process, 5(2), 01–11. https://doi.org/10.5121/ijdkp.2015.5201

10. Estabrooks, A., Jo, T., & Japkowicz, N. (2004). A Multiple Resampling Method for Learning from Imbalanced Data Sets. Computational Intelligence, 20(1), 18–36. https://doi.org/10.1111/j.0824-7935.2004.t01-1-00228.x

11. Singh, K., & Upadhyaya, D. S. (2012). Outlier Detection: Applications And Techniques. 9(1).

12. Smiti, A. (2020). A critical overview of outlier detection methods. Computer Science Review, 38, 100306. https://doi.org/10.1016/j.cosrev.2020.100306

13. Nargesian, F., Samulowitz, H., Khurana, U., Khalil, E. B., & Turaga, D. (2017). Learning Feature Engineering for Classification. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, 2529–2535. https://doi.org/10.24963/ijcai.2017/352

14. Bashir et al. - 2020—An Information-Theoretic Perspective on Overfittin.pdf. (n.d.).

15. Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). Multivariate Statistical Machine Learning Methods for Genomic Prediction. Springer International Publishing. https://doi.org/10.1007/978-3-030-89010-0

16. Tamboli, N. (2023). Effective Strategies for Handling Missing Values in Data Analysis (Updated 2023). https://www.analyticsvidhya.com/blog/2021/10/handling-missing-value/

17. Aggarwal, C. C. (2013). Outlier Analysis. Springer New York. https://doi.org/10.1007/978-1-4614-6396-2

18. Anomaly node. (2024). IBM Corporation. https://www.ibm.com/docs/en/watsonx-as-a-service?topic=modeling-anomaly-node

19. McClenaghan, E. (2022). Mann-Whitney U Test: Assumptions and Example. https://www.technologynetworks.com/informatics/articles/mann-whitney-u-test-assumptions-and-example-363425

20. Moore, D. S. (2017). The basic practice of statistics (3rd ed.). New York : W.H. Freeman c2004.

21. Zheng, A., & Casari, A. (2018). Feature Engineering for Machine Learning Principles and Techniques for Data Scientists (1st ed.). O'Reilly Media, Incorporated.

22. Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. Computer Physics Communications, 145(2), 280–297. https://doi.org/10.1016/S0010-4655(02)00280-1

23.IBM. (2024a). Feature Selection node. https://www.ibm.com/docs/en/cloud-paks/cp-data/4.8.x?topic=modeling-feature-selection-node

24. IBM. (2024b). Random Trees node. https://dataplatform.cloud.ibm.com/docs/content/wsd/nodes/randomtrees.html?context=cpdaas

25. Jahnavi, M. (2017, July 10). Introduction to Neural Networks, Advantages and Applications. Towards Data Science. https://towardsdatascience.com/introduction-to-neural-networks-advantages-and-applications-96851bd1a207

26. Buhl, N. (2023, November 27). Logistic Regression: Definition, Use Cases, Implementation. Encord. https://encord.com/blog/what-is-logistic-regression/

27. Duggal, N. (2023, February 20). Advantages of Decision Trees. Simplilearn. https://www.simplilearn.com/advantages-of-decision-tree-article

28. Agrawal Sumeet. (2024). Metrics to Evaluate your Classification Model to take the right decisions. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/07/metrics-to-evaluate-your-classification-model-to-take-the-right-decisions/

29. Saltelli, A. (Ed.). (2007). Sensitivity analysis in practice: A guide to assessing scientific models (Reprinted). Wiley.

30. Aniruddha Bhandari. (2024, April 4). Guide to AUC ROC Curve in Machine Learning: What Is Specificity? Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/06/auc-roc-curve-machine-learning/

31. Acuña, E., & Rodriguez, C. (2004). The Treatment of Missing Values and its Effect on Classifier Accuracy. In D. Banks, F. R. McMorris, P. Arabie, & W. Gaul (Eds.), Classification, Clustering, and Data Mining Applications (pp. 639–647). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-17103-1_60

32. Zhou, X. (2019). Data mining in customs risk detection with cost‑sensitive classification. 13(2).

33. Steele, M. (2021, March 31). Feature Engineering Examples: Binning Categorical Features. Towards Data Science. https://towardsdatascience.com/feature-engineering-examples-binning-categorical-features-9f8d582455da

34. Wu, F. (n.d.). Feature Selection: The “why”, the “what” and the “how.” TurinTech CSO. https://www.turintech.ai/the-why-the-what-and-the-how/

35. Encord & Encord. (2023, August 9). Introduction to Balanced and Imbalanced Datasets in Machine Learning. https://medium.com/cord-tech/introduction-to-balanced-and-imbalanced-datasets-in-machine-learning-be60c6eeb8be#:~:text=Balancing%20a%20dataset%20makes%20training,because%20it%20contains%20more%20data.

36. Mojtaba Hassanzad & Karimollah Hajian-Tilaki. (2024). Methods of determining optimal cut-point of diagnostic biomarkers with application of clinical data in ROC analysis: An update review. https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-024-02198-2