Open Peer Review on Qeios

Supply Chain Fraud Prediction with Machine Learning and Artificial intelligence

Mark Lokanan[1], Vikas Maddhesia[1]

1 Royal Roads University

Funding: The author(s) received no specific funding for this work.

Potential competing interests: The author(s) declared that no potential competing interests exist.

Abstract

The increasing complexity of supply chains is putting pressure on businesses to find new ways to optimize efficiency and cut costs. One area that has seen a lot of recent development is machine learning (ML) and artificial intelligence (AI) to help manage supply chains. This paper employs machine learning (ML) and artificial intelligence (AI) algorithms to predict fraud in the supply chain. Supply chain data for this project was retrieved from real-world business transactions. The findings show that ML and AI classifiers did an excellent job predicting supply chain fraud. In particular, the AI model was the highest predictor across all performance measures. These results suggest that computational intelligence can be a powerful tool for detecting and preventing supply chain fraud. ML and AI classifiers can analyze vast amounts of data and identify patterns that may evade manual detection. The findings presented in this paper can be used to optimize supply chain management (SCM) and make predictions of fraudulent transactions before they occur. While ML and AI classifiers are still in the early stages of development, they have the potential to revolutionize SCM. Future research should explore how these techniques can be refined and applied to other domains.

Mark Lokanan[1], **Vikas Maddhesia**[1]

[1] *Royal Roads University 2005 Sooke Road, Victoria, BC Canada V9B 5Y2 | royalroads.ca*

Keywords: Supply chain fraud; Machine learning; Artificial intelligence; Predictive analytics.

Introduction

Supply chain fraud is becoming more common with the digitization of business operations and e-commerce transactions. Fraud in the supply chain can take many forms, which pose a serious threat to businesses. In some cases, it can result in the loss of merchandise or the theft of confidential information. In other cases, it can lead to the disruption of supply chains and the waste of resources. With the network of intermediaries across the globe, organizations are more vulnerable to supply chain fraud in countries with lax or less stringent rules governing misconduct. The increase in cross-border transactions presents a higher risk of fraudsters using phony invoices or other false documents to commit fraud. As

top management shifts its focus to handling urgent operational concerns, prospective fraudsters see this as an opportunity to infiltrate the supply chain.

The vastness and complexity of the global supply chain make it challenging to trace the origin of items. The rise of digitization and e-commerce transactions presents opportunities for fraudsters, who can introduce counterfeit goods or tamper with products in the supply chain. As a result, companies can lose millions of dollars annually to supply chain fraud. Many businesses use machine learning (ML) and artificial intelligence (AI) in their supply chains to solve this issue. By analyzing large data sets, ML intelligence in the form of computational technology can help to identify patterns and provide early warning signs of fraud in the supply chain. ML and AI algorithms can detect patterns in data that humans would not be able to identify. In cases where the ML algorithm detects a suspicious transaction, it can be flagged for manual review. Companies can use ML intelligence to automate fraud detection, making it more efficient and effective.

In recent years, there has been an increase in ML and artificial neural networks (ANN) to identify and prevent fraud in the supply chain. These two computational technology approaches are known as “ML" and "deep learning." This paper aims to use ML and ANN to predict fraud in the supply chain of a manufacturing company. We will train our models on data from past fraud cases and use these models to identify the key features associated with fraud in the supply chain. Our aim is to provide a robust technique that manufacturing companies can use to detect and prevent fraud in their supply chains. This work presents several important advances to the existing body of research about using computational techniques to identify fraudulent activities to supply chains.

By identifying irregularities in the supply chain, these technologies can help companies take steps to prevent fraud before it occurs. In addition, ML and AI can also monitor the supply chain in real time, allowing companies to quickly identify and respond to potential threats. As the world becomes increasingly digitized, ML and AI will play an important role in protecting businesses from supply chain fraud. In addition, ML and AI algorithms can be used to automatically flag suspicious activities, making it easier for investigators to detect fraud. As supply chains become more complex in the digital space, ML and AI will become essential tools for detecting and preventing fraud.

The remaining section of this paper is organized according to the manner described below. Section one reviews the extant literature on ML technology and fraud detection. Emphasis will be placed on using ML and AI techniques to detect fraud in the supply chain. Section three describes the research methodology and experimental setting. Section four provides an analysis of the findings. Section five concludes with a discussion and outlines areas for future research.

The Literature on Smart Technology and Supply Chain Fraud

ML intelligence has emerged as a new approach for detecting fraud in supply chains. In big data, supply chain management (SCM) increasingly employs AI to identify and prevent fraudulent activities[1][2][3]. Significant losses may result from an interruption in the supply chain, which can have ripple effects across the whole system. By evaluating transactional data, computational intelligence in the form of ML and AI may uncover trends indicating fraud[4]. For instance, a supplier's abrupt spike in order volume might be reported as suspicious, or a sudden delay in shipment may suggest that the items have been redirected to a different location. ML can be used to monitor real-time data and discover transaction irregularities in the order and shipping process. Overall, ML intelligence is a powerful tool to detect red flags of fraud, and its use in SCM

will likely grow in the near future[2].

The examination of data is critical in supply chain management (SCM). While ML is important in many areas of SCM, including but not limited to procurement, inventory management, warehousing, and logistics, there is one area where computational intelligence has a significant impact: predicting smart supply chain fraud[2]. ML in the form of predictive analytics is increasingly being used to diminish the risk of fraud in the supply chain[5][6][7]. When it comes to fraud prevention in the supply chain, research has shown that ML can be used to flag potential irregularities in large data sets[2][5]. While the research in this area is in its infancy, early commercial application findings have indicated the potential for more widespread use of computational technology in supply chain fraud management[1][8][9]. Predictive analytics to detect risk in the supply chain is an emerging area of research that has been successfully used in several fields, including risk prediction.

ML can be a powerful tool for identifying risk in supply chains (Melançon et al. 2021). Recent research has used ML techniques to predict the risk of illegitimate transactions in supply chain data[2][7][10]. Others have used ML techniques to model disruptive events and detect risk in the supply chain of financial market transactions[11]. Researchers have used ML intelligence in container shipping to develop risk analysis tools to discover irregular container shipments in the supply chain[12]. This stream of research is important because it helps protect businesses and consumers from fraudsters increasingly using sophisticated methods to infiltrate the supply chain. Research has shown that by analyzing data from various sources, ML algorithms can identify patterns that indicate a potential risk to the company's supply chain[12][13]. For example, an algorithm might examine historical data to identify suppliers that have frequently been late with deliveries. By flagging these suppliers as high-risk, businesses can make sure to monitor their performance and take steps to avoid disruptions closely. Companies can avoid costly interruptions and keep operations running smoothly by identifying risks early and taking proactive measures to manage them[2].

Predictive Analytics and Supply Chain Fraud

Predictive analytics is a powerful tool that is becoming increasingly important in the fight against fraud in the food supply chain. By analyzing large amounts of data, predictive analytics can help identify patterns and trends and enable transparency and visibility in the food distribution supply chain[3][14]. As more sophisticated predictive analytics tools become available, they will likely play a more significant role in food supply chain fraud detection. Predictive analytics can help to identify anomalies and red flags that may indicate fraud and, by doing so, can help to protect consumers from adulterated or counterfeit food products[15][16]. Predictive analytics has also been used in other areas to improve food tracing and recalls and optimize food safety compliance programs[3]. As food SCM moves to blockchain distribution, predictive analytics is likely to play an even more critical role in protecting the food supply chain from fraud[15][17].

Another area that received particular attention is predictive analytics in healthcare[18][19]. Recent research has used ML to safeguard payments and identify fraud in medical insurance supply chains[20]. Others[19][21] note that due to the complexity and volume of data and the sheer number of intermediaries in the health care supply, it is difficult to detect fraud through manual detection. ML and AI are valuable tools to detect fraudulent transactions in vast health care and insurance databases[21][22]. [22] state that supervised ML models can be used to identify fraudsters with high accuracy. Furthermore, ML and AI models can be fine-tuned to focus on specific types of fraud, such as claims or provider fraud[18]. AI systems can also be used to monitor claims for suspicious activity, such as unusually high claims volumes or out-of-network providers[21].

In other areas, Abbas and company employ ML intelligence to detect counterfeit drugs in the pharmaceutical supply chain[5]. As health care and insurance data continue to grow in volume and complexity, ML and AI will play an increasingly important role in detecting fraud and protecting patients and providers from financial losses[18][22].

Despite its application, many potential applications of predictive analytics to detect fraud in SCM have not yet been explored. For example, predictive analytics can be used to detect fraud from the medium of payment (i.e., electronic or cash) or through the order status of the buyer (on-site or online). The first avenue, detecting fraud from the medium of payment, is currently being used by a few companies with success. This method relies on understanding patterns in past payments (i.e., historical data) to flag suspicious transactions for further investigation. The second avenue, detecting fraud from the order status of the buyer, has shown great potential but is not yet being used due to data limitations. This paper will use ML and AI algorithms to address an area of research lacking in the literature to detect fraud through order status. This methodological approach uses ML algorithms to make predictions about whether or not an order is likely to be fraudulent based on historical fraud patterns.

Experimental Setting and Research Design

The data was analyzed using the Python programming language. Python is an open-source programming language that is free to use for ML tasks. Scikit-learn is the name of the ML software package that was used to analyze the data. The Scikit-learn library is a collection of Python tools that may be used for statistical modelling and ML. Scikit-learn is the go-to library for classification, regression, or clustering tasks.

Dataset

The data for this paper came from a large manufacturing company. The data comes from real-world business transactions and can be used to identify fraud detection opportunities in the company's supply chain. There are around 180000 observations taken from supply chain transactions that have been mined over three years. The dataset is housed in the Mendeley data repository and is licensed under Creative Commons 4.0[2]. The goal is to find patterns in the data to help the company stop fraud in the supply chain.

Variables and Measurements

Table 1 displays the independent variables along with their descriptions and measurements. The variables are transactional indicators collected by the company (supply chain data provider) to monitor and analyze the performance of various supply chain operations (transactions). Generally, the data comprises production, orders, sales, and distribution features. The features assess how well a company utilizes resources to manage its supply chain connections. In this project, these features were used to forecast fraudulent transactions.

Table 1. Descriptions of Features and their Attributes

Features	Description	Measure
Type	Type of Payment	Categorical
Days for shipping (real)	Actual time taken for shipping of the product	Continuous
Days for shipment	Estimated shipping time	Continuous
Delivery Status	Update on delivery status	Categorical
Late_delivery_risk	Risk indicator of delivery (i.e., late or not)	Binary
Category Name	Category name of the item being shipped	Categorical
Customer City	Customer's city name	Categorical
Customer Country	Customer's country name	Categorical
Customer Segment	Type of customer	Categorical
Customer State	Customer's state name	Categorical
Department Name	Department name of the product being sold	Categorical
Market	Region where country belongs	Categorical
Order City	City where product was ordered	Categorical
Order Country	Country where product was ordered	Categorical
Order Item Discount Rate	Discount rate on product being ordered	Continuous
Order Profit Per Order	Profit on the order	Continuous
Order Region	Region from where product being ordered	Categorical
Order State	Country from where order is being placed	Categorical
Product Name	Name of the product	Categorical
Product Price	Price of the product	Continuous
Order_Year	Year of Order Placed	Categorical
Order_Week_day	Weekday of Order Placed	Categorical
Order_Month	Month of Order Placed	Categorical
Order_Hour	Hour of Order Placed	Categorical
Total Price	Order Item Quantity*Order Item Total	Continuous
Shipping Mode	Shipping Mode of the product	Categorical

Dependent variable

The dependent variable is fraud. Fraudulent transactions do not meet the business's expectations or standards. These transactions may be accidental or intentional, but they all involve some form of fraudulent behaviour. This paper defines fraud as a binary two-class problem where the classes are coded as fraudulent and non-fraudulent transactions. We have chosen to label the classes this way to simplify the analysis and give a clear definition of each set. We coded fraud as (0 = no) when the transaction was not fraudulent and (1 = yes) when the transaction was flagged as suspicious or fraudulent. The formula to represent financial sanctions is shown in equation 1:

$$y = \{1, Fraud, 0\text{Non-Fraud}\} \quad \text{eq. 1}$$

Data Cleaning and Preprocessing

There were many redundant variables in the dataset. The following categorical variables were thrown out because they were related to customer demographics and would not have helped create the classification models:

- 'Customer Email', 'Product Status', 'Customer Password', 'Customer Street', 'Customer Fname', 'Customer Lname', 'Product Description', 'Product Image', 'Order Zipcode'

The following variables were dropped because there are replications of the same features in the dataset and contain the same values:

- 'Benefit per order', 'Order Customer Id', 'Product Card Id'

The following numerical variables were dropped from the dataset because they represent a unique numerical ID corresponding to any department or product. However, the categorical names were kept, which will come in handy during model creation after transposing them to numerical features using the Get_Dummies method. These variables include:

- 'Category Id', 'Department Id'

The variables 'order date (DateOrders)' and 'shipping date (DateOrders)' were dropped because we had already extracted the year, month, and day to be used in the model.

Addressing the Multicollinearity

Multicollinearity is when two or more independent variables demonstrate a high correlation in the data. It affects the interpretability of ML models as it compromises the contribution of independent variables to the target outcome. In the case of a linear regression model, multicollinearity among independent variables leads to inflated variances. Inflated variance, in turn, affects the stability of the model and makes it difficult to interpret the coefficients. In other words, the model becomes less reliable. A correlation test is used to identify if multicollinearity existed in the independent variables within the data.

As can be seen in Figure 1, a correlation coefficient was calculated for each pair of independent variables. The higher the correlation coefficient between a pair of independent variables, the stronger their relationship and the more multicollinear they are. The classification model's quality can be enhanced by eliminating or removing the impacts of multicollinearity variables in the data. A correlation of .70 was used as the cut-off. To avoid multicollinearity, the variables that showed a high correlation (over 0.70) were dropped from the dataset. We followed successive variable reduction methods to eliminate highly correlated variables. Figure 1 shows the relationship between features before addressing multicollinearity issues. Note

that the features highlighted in dark blue are highly correlated.

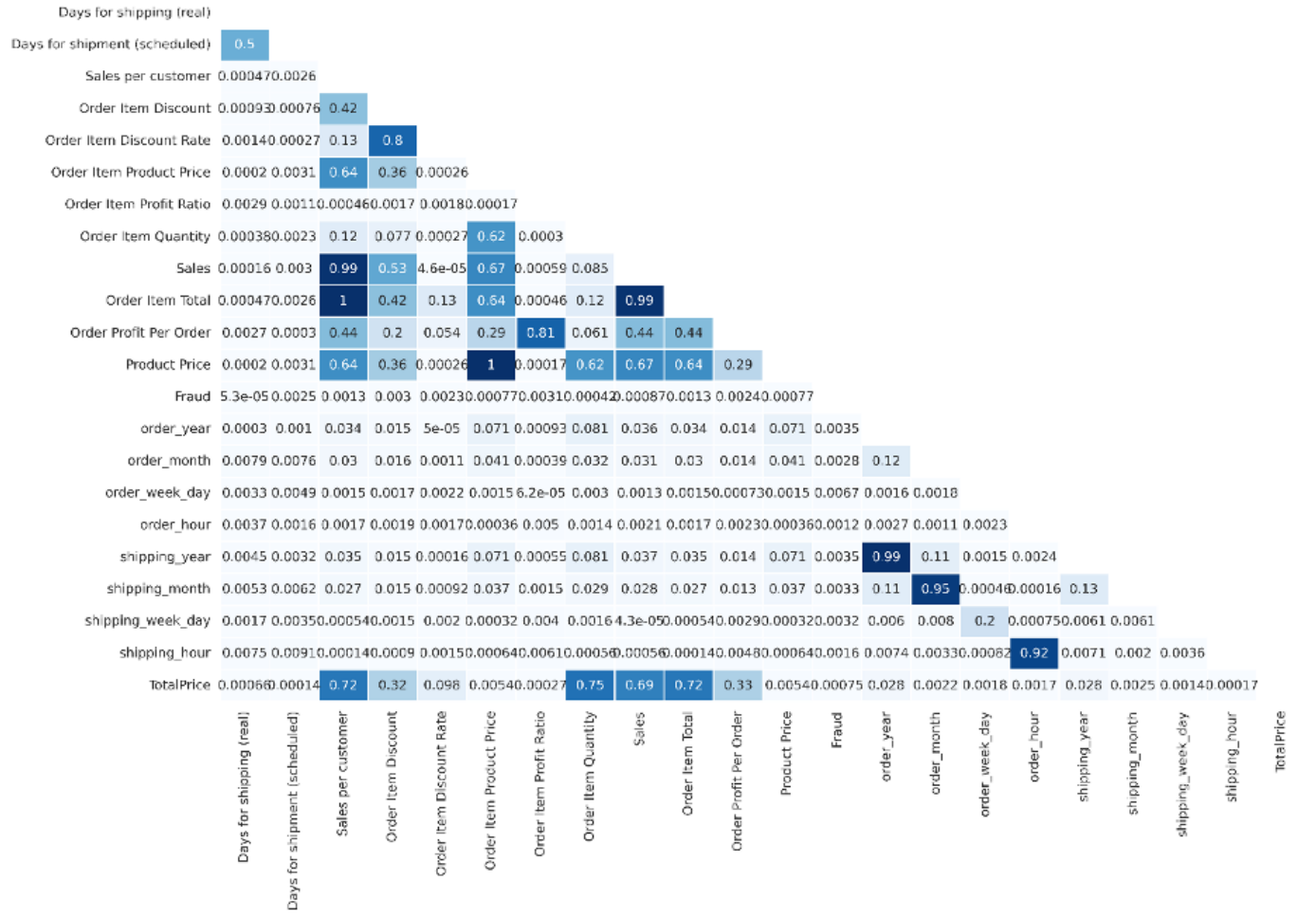

Figure 1. *Before removing Multicollinearity from the data*

We used successive variable reduction (SVR) to eliminate highly correlated features from the dataset[23]. SVR is a technique used to detect multicollinearity in data. It is based on the idea that if two or more variables are highly correlated, then one of the variables can be removed without losing any information. Once the correlated features are identified, the variable with the highest correlation with the response variable is removed from the data set. Next, we recalculated the correlations of all the features and repeated this process until all the highly correlated features were removed from the dataset. This technique is particularly useful when there are many variables, and one of the objectives is to reduce the number of features to eliminate the presence of redundancy in the data[16][24]. Using SVR, the following variables have been eliminated from the dataset: 'Sales per customer,' 'Order Item Product Price,' 'Sales,' 'shipping_year,' 'shipping_month,' 'shipping_week_day,' 'shipping_hour,' 'Order Item Discount,' 'Order Item Profit Ratio,' 'Order Item Total,' 'Order Item Quantity.' As shown in Figure 2, once SVR was conducted, there were no issues with multicollinearity in the dataset.

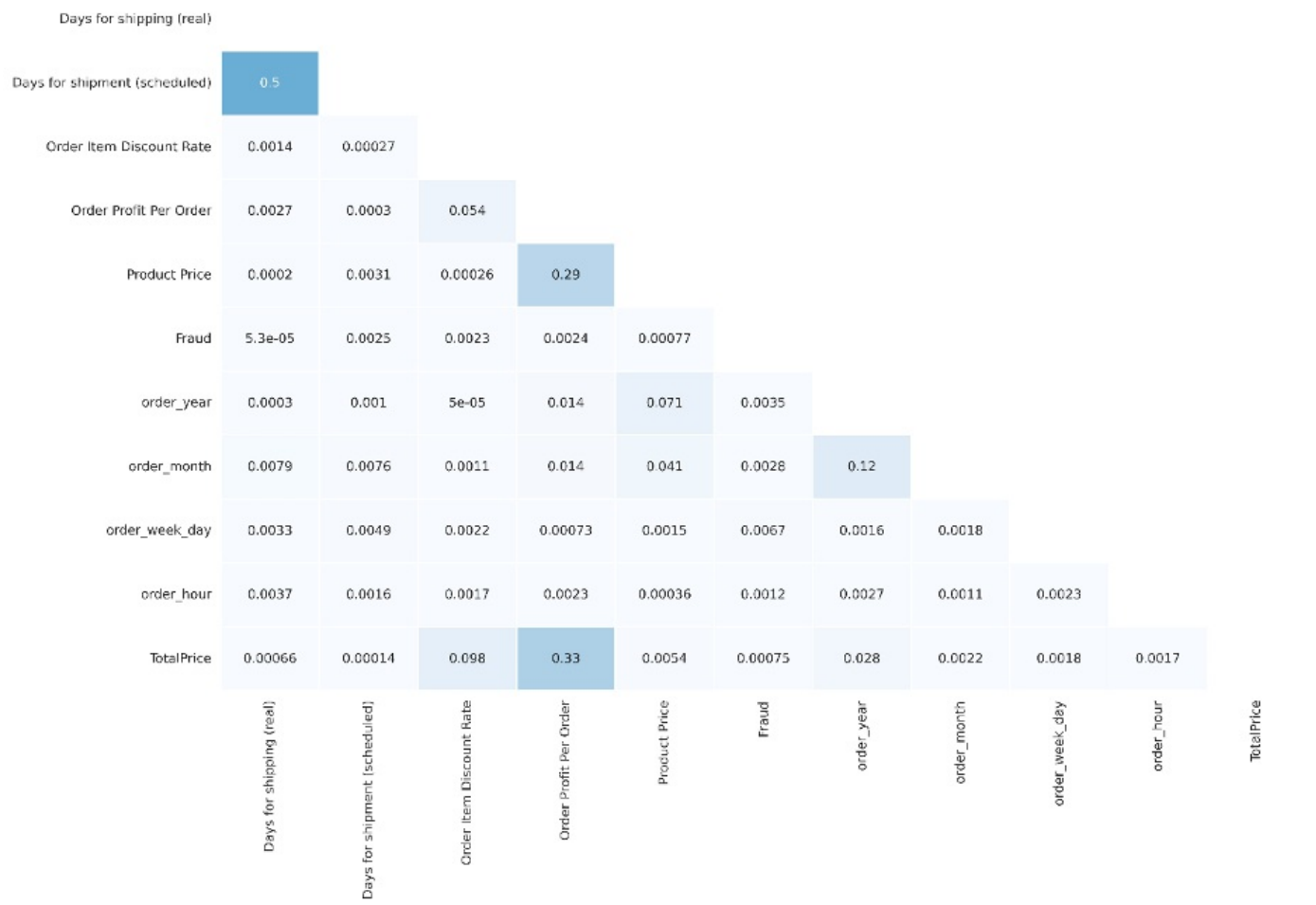

Figure 2. *After removing Multicollinearity from the Data*

Dealing with Class Imbalance Problems

This project involves working with an unbalanced dataset. As seen in equation 2, class imbalance occurs when one label of the target variable is much higher than the other. The RandomUnderSampler (RUS) with Synthetic Minority Over-Sampling Technique (SMOTe)

was used to address the class imbalance situation. RUS is a simple and uncomplicated method for balancing data that involves randomly picking a part of the data for the targeted classes and then undersampling the majority class via a random sample of the observations. The randomness ensures that every observation has an equal chance of being selected to be included in the analysis dataset. By randomly selecting a subset of observations, we can ensure that the data represents the studied population. As seen in equation 2, the fraudulent transactions account for only 2.25% of the data, while transactions that were not fraudulent accounted for 97.75%. A model's learning from highly unbalanced training data is incorrect because classification algorithms usually misclassify minority classes as majority classes. Classification imbalance is a problem for fraud detection because fraudulent activities are misclassified, making it challenging for the algorithms to predict based on the unseen data[25].

$Fraud_{Yes}$ = (Transaction with Fraud indicator 1 / Total Transactions) * 100 $Fraud_{yes}$ = (4062/180519) * 100 = 2.25% eq. 2

Figure 3 and Figure 4 represent the data before and after under-sampling was applied. Note from Figure 3 that there were only 2.25% of fraudulent transactions. As shown in Figure 4, the majority class (non-fraudulent transactions), the dominant class before the sampling, had the same intensity (equal presence) after the under-resampling. RUS removed the bias from the data by randomly eliminating some observations from the dominant class to ensure an equal class distribution. Given this large dataset, SMOTe RUS did not lead to information loss as the final dataset comprises 52,000 observation points[26].

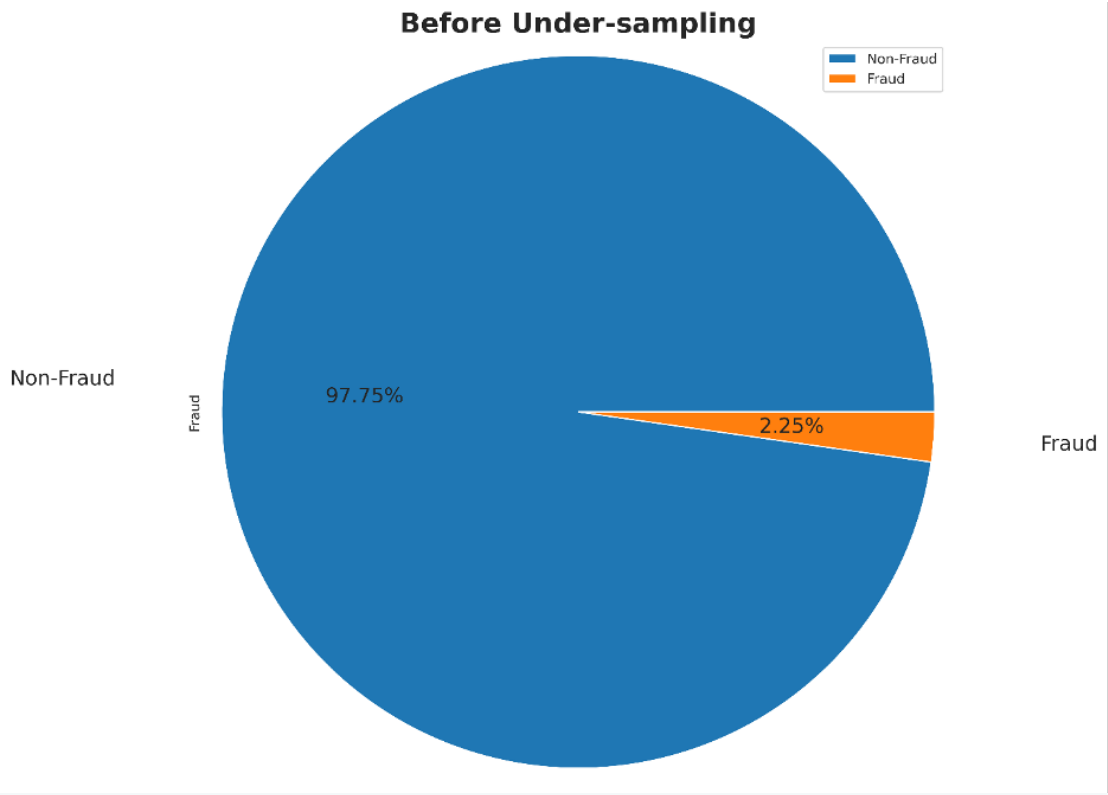

Figure 3. *Data before using Under Sampling*

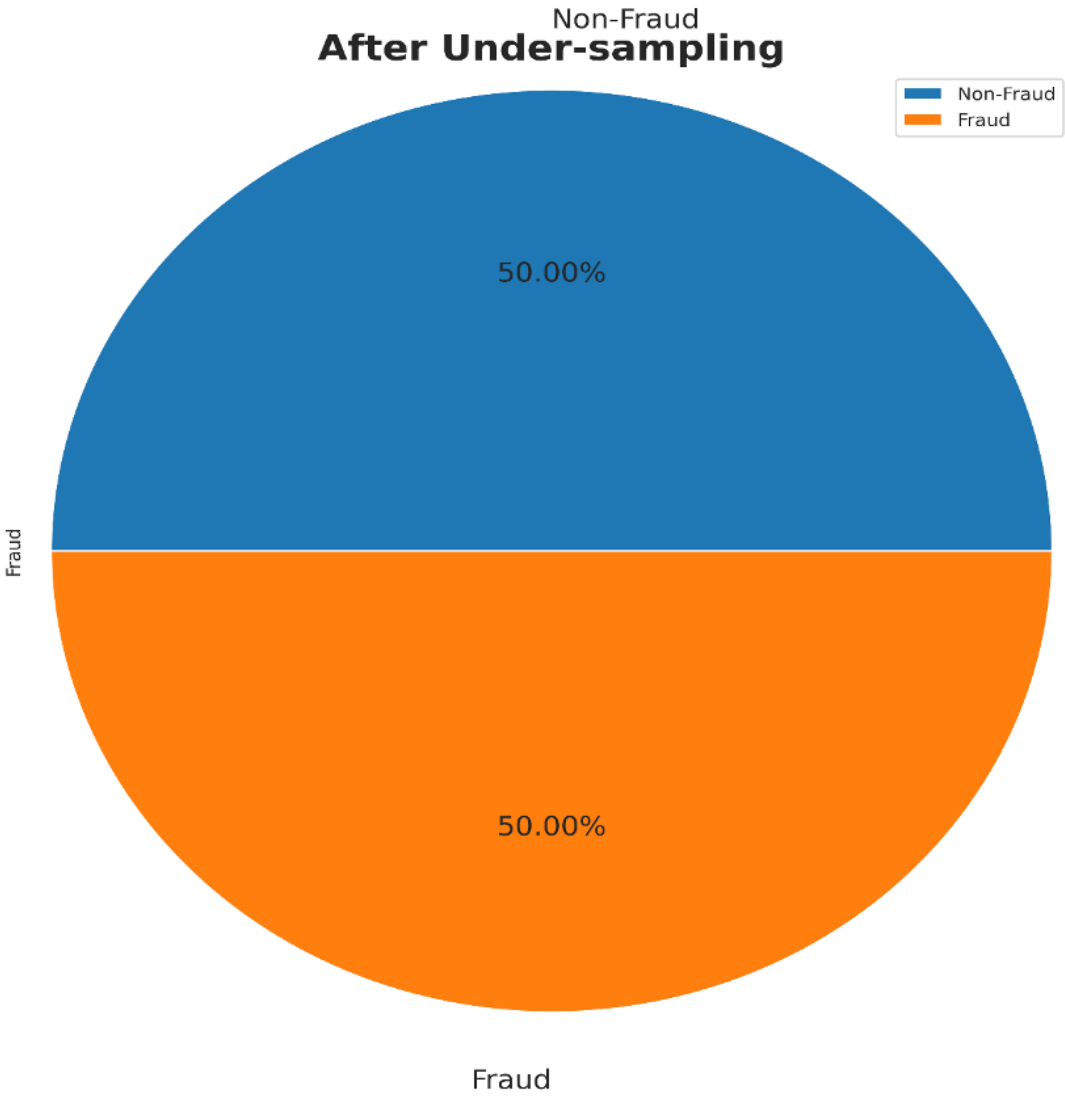

Figure 4. *Data after using Under Sampling*

Descriptive Findings on Fraudulent Observations

To gain more insight into the experimental results, it is important first to understand the instances of fraud throughout the supply chain. Figure 5 shows the instances of fraud based on transfer types. Fraudulent transactions occur only in the "Transfer" type of payment. These findings mean that companies must be extra vigilant and screen all transfer transactions for red flags of fraud.

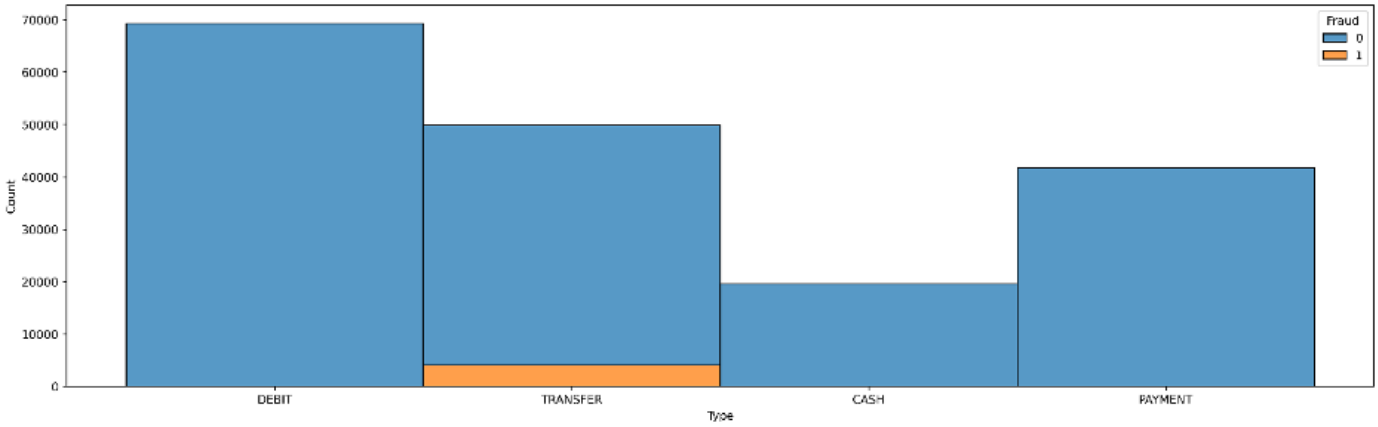

Figure 5. *Fraudulent Transaction with Payment Type*

As can be seen in Figure 6, most fraudulent transactions occur on time. When a transaction is carried out within the typical period for that kind of purchase, it is considered an instance of on-time fraud. Note also that none of the fraudulent transactions are ever late (i.e., all fraudulent items are delivered on time). The complexity of on-time fraud suggests that internal control mechanisms are not effectively spotting red flags of fraud. More critically, on-time fraud transfer might also imply that the organization's employees are not trained to identify fraud from normal transactions. These findings suggest a need to automate the fraud prediction system so that suspicious transactions may be identified in real time.

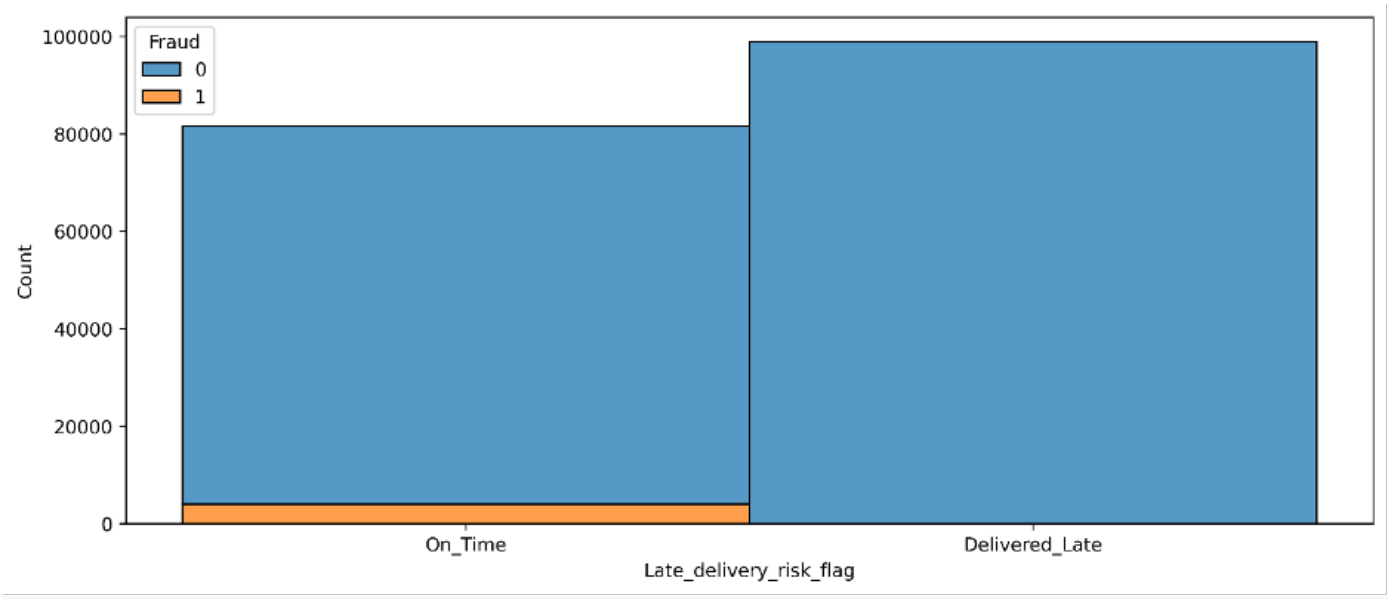

Figure 6. *Fraudulent Transaction with delivery risk*

Figure 7 presents the results of the department most affected by fraud. Most of the fraudulent transactions occur in three departments, namely, apparel, golf, and fans. Note from Table 2 that the Fan Shop, Apparel, and Golf departments contributed to more than 82% of illegal transactions. The Fan Shop department had the most fraudulent transactions (37%). Further research should be done to determine the reasons behind the 82% of fraudulent transactions that take place in these three departments.

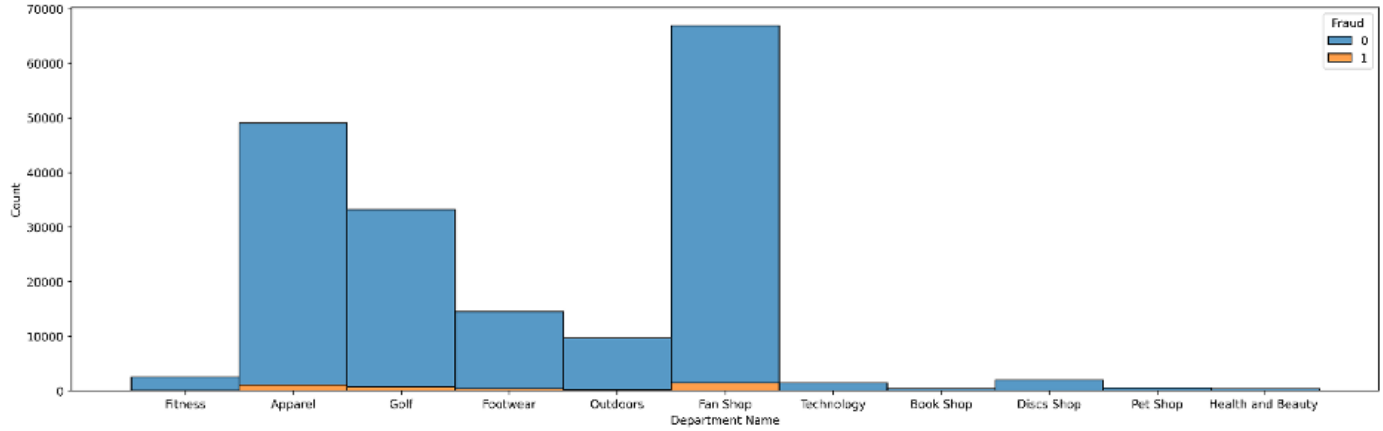

Figure 7. *Fraudulent Transaction with Department name*

Table 2. Fraud Transactions

Department	Fraud Transactions
Fan Shop	36.51%
Apparel	27.74%
Golf	18.32%

Algorithms Considered

For the purpose of detecting fraud in SCM, three different algorithms were considered. The algorithms include logistic regression, random forest, and an AI-based sequential model. These ML algorithms were chosen because they have a track record of successfully detecting fraudulent transactions. Additionally, these algorithms have been tried and tested over time across various domains and datasets.

Logistic regression

Logistic regression is a well-known classification algorithm in ML. The essential premise is that a logistic function assigns a conditional probability to one of two possible output variables. S(t) is a logistic sigmoid function that modifies and categorizes the linear function $b + b'X$ into two or more discrete categories[4]. Logistic regression is very beneficial for detecting fraud since it may offer a ranking order of the classified data set based on the likelihood of fraud versus no-fraud observations[27]. Logistic regression examines each transaction and assigns a probability to it, which is then used to determine whether the transaction is fraudulent. If the probability exceeds the threshold (0.5 by default), the transaction is considered a fraud; otherwise, it is considered not fraud[28]. The logistic regression equation is represented by the formula in equation 3.

where,

P is the probability,

X is the input set, and

*b and b′*are the corresponding coefficients calculated using maximum-likelihood estimation while training.

$$P(x)^n = \left(\frac{(e)^{b+b'X}}{1 + (e)^{b+b'X}}\right) \quad \text{eq. 3}$$

CatBoost

CatBoost is a machine learning algorithm that can be used for both regression and classification tasks. It is well-suited for dealing with categorical features and can also handle missing values[29]. Furthermore, it is insensitive to the order of categorical features, which makes it robust to potential data leakage[30]. The CatBoost algorithm has been designed specifically to achieve high accuracy on tabular data. In general, CatBoost outperforms other machine learning algorithms on regression and classification tasks[16][29][30]. CatBoost is also popular in the industry due to its ease of use and ability to scale to large datasets. In conclusion, CatBoost is a powerful machine learning algorithm that can be used to get state-of-the-art results on both classification tasks.

Random forests Classifier Tuned with GridSearchCV

A Random forest Classifier (RFC) is a supervised ML technique often used to solve classification and regression tasks[31]. The RFC is made up of a collection of tree classifiers, each of which is produced using a random vector sampled separately from the input vector. Each tree in a random forest is created independently of the others, contributing to the RFC's computational efficiency. The vast number of trees in the ensemble makes the RFC resistant to overfitting and noise in the data[32]. A decision tree is constructed using an RFC by employing a random selection of characteristics or a combination of features at each node. The RFC works by constructing a collection of decision trees from a large number of samples using a method known as sample with replacement. This method considers which trees have received the most votes for classification problems and which trees have received an average number of votes for regression tasks. The RFC is represented by the formula in equation 4.

Where

h_i corresponds to single-decision-tree model trees,

Y is the target output,

I is the indicator function, and

arg max returns the value of the outcome variable *Y*

$$H(x) = arg\ max\ Y \left(\sum_{i=1}^{nl} (h_i x = Y) \right) \quad \text{eq. 4}$$

GridSearchCV (gridsearch) was used to optimize the RFC model in this project. Gridsearch is a robust ML technique for optimizing models by modifying their hyperparameters (Lokanan and Sharma, 2022). The gridsearch algorithm searches exhaustively through a defined parameter grid, using a cross-validation scoring metric, to discover the best set of parameters for the model[33]. This technique may significantly increase performance, mainly if the hyperparameter default values are inappropriate. Gridsearch is very good at finding fraudulent transactions because it looks at all possible parameter combinations and chooses the best model based on a tested approach[34].

AI Sequential model

A sequential approach is appropriate for a simple stack of layers with exactly one input and one output tensor[25][35]. It is constructed with only Relu, Batch Normalization, and Softmax layers. As shown in Figure 8, the tensor of feature-encoding features is present in the input, output, and weight layers. The model, which consists of an input layer, a hidden layer, and an output layer, is optimized using gradient descent and the Adamax algorithm[35][36]. One tensor in the input layer accepts a feature vector and outputs the class of the related layer, in this case, fraud or no-fraud. We utilize the sequential model in this data set to forecast the type of an observed transaction based on its attributes. The AI sequential classification model accepts various activation functions: Tanh, Sigmoid, Linear, and ReLu. In this paper, ReLu was used because it is appropriate for different classification problems for neural network architectures and is useful to avoid vanishing gradient problems[37][38]. Using a variety of parameters, different activation functions and parameter values were used to test how well the models worked.

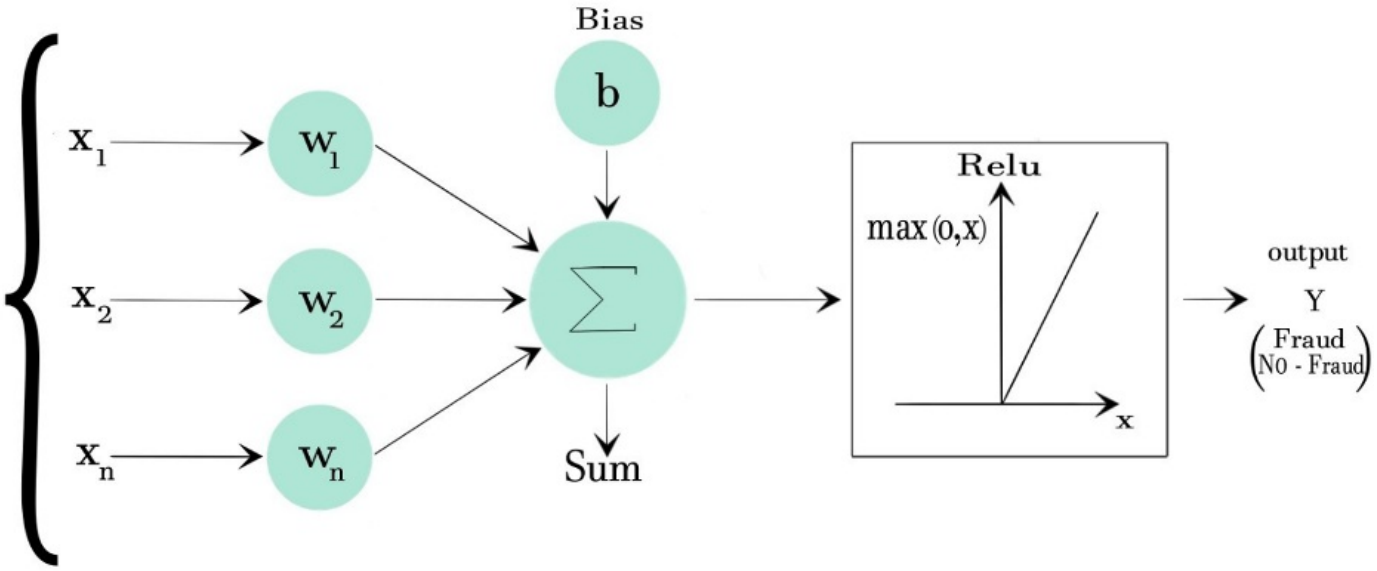

Figure 8. *AI Sequential Model with Relu Activation Function*

Experimental Results

The confusion matrix is used to assess the effectiveness of a classification model (or "classifier") on a set of test data for which the actual values are known. The percentage of correctly predicted classes is the simplest method to interpret a confusion matrix. The entries in the confusion matrix are actually counted as predictions, which are thought of as predicted class labels. As shown in Figure 9, the matrix rows correspond to the actual class labels while the columns correspond to the predicted class labels. The count of correct predictions appears along the diagonal of the matrix, while the counts of incorrect predictions appear in the off-diagonal elements. The confusion matrix is decomposed into the 2x2 binary classification diagram in Figure 9:

Where

True Positive (TP): Correct accepted – The algorithm predicts that it is a fraudulent online transaction, and the transaction is actually fraudulent.

True Negative (TN): Correct reject – The algorithm predicts that it is not a fraudulent online transaction and the transaction is actually not fraudulent.

False Positive (FP): False Alarm – The algorithm predicts that it is not a fraudulent transaction, but the transaction is actually fraudulent.

False Negative (FN): Miss – The algorithm predicts that it is a fraudulent online transaction, but it was actually not a fraudulent transaction.

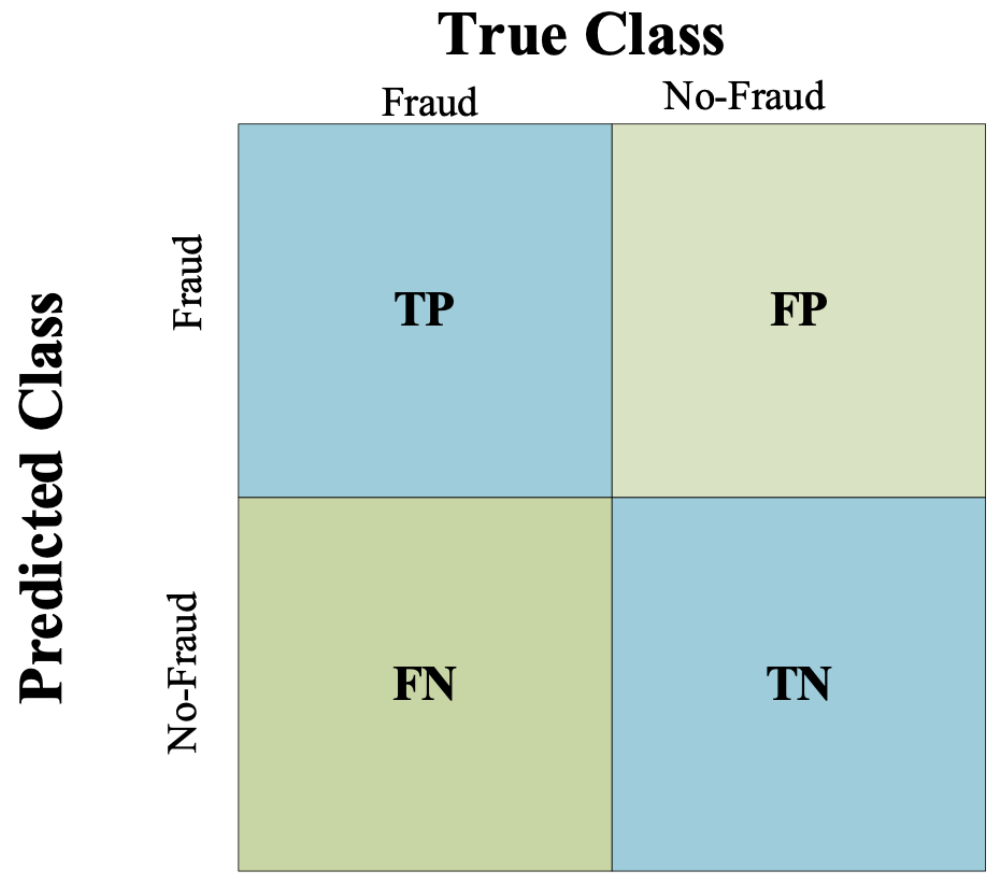

Figure 9: Confusion Matrix

Performance Accuracy

There are several ways to compute classification accuracy measures from a confusion matrix. The most basic measure is the ratio of correct predictions to total predictions, also known as accuracy. The confusion matrix can be used to calculate the performance accuracy of the models. The formula for computing the accuracy score is shown in equation 5:

$$= (+) / (+ + +) \quad \text{eq. 5}$$

Logistic regression was used as the baseline model to compare the other classifiers. One of the hallmarks of accurate modelling is to prevent model overfitting. As shown in Table 3, none of the models suffers from overfitting problems because there is not a huge gap between the training and test scores. The AI model built on the Keras sequential classification and the CatBoost classifiers had the highest (99%) performance accuracy on the test set. The random forest (97%) and logistic regression classifier (96%) also performed exceptionally well. The accuracy of all three models on the test dataset is very close, which signifies that our models are consistent with each other.

However, accuracy is not the best metric to use when measuring performance with an imbalanced dataset because it does not differentiate between the number of correctly classified observations of the two classes in binary classification

models[39]. With an imbalanced dataset, accuracy favours the majority class and only considers the false positive and false negative. In such cases, the sensitivity specificity, precision, and F1-scores are considered more robust measures of performance for imbalance datasets[25][40].

Table 3. Classification Performance Accuracy

Type of Algorithm	Accuracy Score	
	Training Score	Testing Score
Logistic regression	96.6%	96.0%
Random forest with GridSearchCV	97.1%	97.0%
CatBoost	99%	99.0%
AI Sequential Classifications	99.3%	99.0%

Classification Scores

The classification scores (i.e., sensitivity, specificity, and F1-measure) provide insightful information about how well the models perform on a given task[4]. For example, a model designed to detect fraud would want to maximize its sensitivity or true positive rate (TPR) while maintaining a high specificity or true negative rate (TNR). The F1-score is an excellent overall performance measure because it combines sensitivity and specificity. The goal is to achieve high scores on all three measures. However, the trade-offs between sensitivity and specificity can vary depending on the application. For instance, in some cases, having a high specificity or low false positive rate (FPR) may be more important than having a high sensitivity or higher false negative rate (FNR). It is ultimately up to the analysts to determine what trade-offs are acceptable. Classification sensitivity and specificity are complementary measures to accuracy. A classifier's sensitivity indicates how likely it is to give the correct answer to unseen data[35].

Sensitivity

A classifier's sensitivity is the probability that a randomly chosen example from a dataset is classified as the positive class. Sensitivity is calculated as the inverse of specificity and measures the proportion of actual positive classes that were correctly predicted as positive. Thus, if the probability of being classified as positive is $P(y = 1)$ and the probability of being classified as negative is $P(y = 0)$, then sensitivity equals $P(y = 1)/P(y = 0)$. Following this formula, the results indicate that all models did an excellent job of correctly classifying the fraudulent observations in the supply chain. The number of fraudulent transactions captured in fraud detection is more important than the model's accuracy[41]. An arbitrary high score for the sensitivity parameter is not necessarily desirable in real-world applications. For example, a financial institution is less likely to approve a loan to an individual with a high sensitivity score because there is a greater chance that the individual will default on their loan. As can be seen in Table 4, the sensitivity scores were high for all of the models. The random forest with gridsearch and CatBoost classifiers had the highest sensitivity scores (100%), followed by the logistic regression and the AI

models (99%). These findings indicate that all of the models were better balanced for Type I (FP) and Type II errors (FN) when attempting to predict whether the transaction involved fraudulent activity. However, because the FN would result in the transactions being labelled as fraudulent when they are not fraudulent, it is a significant challenge to reduce the amount of Type II error.

Specificity

A classifier's specificity is the proportion of test data that is not classified incorrectly. Like sensitivity, specificity is calculated as the inverse of recall. Thus, if the probability of being classified as positive is $P(y=1)$ and the probability of being classified negatively is $P(y=0)$, then specificity equals $P(y=0)/P(y=1)$. According to Bayes' logic, conditional probabilities often refer to the conditional probability of some event occurring, given the occurrence of another event (Zhang and Gao 2011). Specificity is crucial when the aim is to reduce the number of negative classes incorrectly classified by the model. Specifically, the specificity metric measures the degree to which non-fraudulent transactions are misclassified. When a classification has a high false positive rate, its specificity diminishes[42]. Table 4 shows that the AI sequential model and CatBoost have the highest specificity scores (98%), followed by Random forest with gridsearch (94%), and logistic regression has the lowest score (93%). These findings indicate that the AI Model correctly predicts a 98% chance of detecting non-fraudulent transactions in the supply chain.

Precision

Precision is another crucial parameter for assessing performance in classification models. Precision measures the proportion of predicted positive classes that are actually positive. Like sensitivity and specificity, precision is calculated as the inverse of recall and is represented in equation 6:

$$P(y = 1 \mid x) / (P(y = 1 \mid x) + P(y = 0 \mid x)) = P(y = 1 \mid x) / (1 - p(y = 0 \mid x)) \qquad \text{eq. 6}$$

Table 4 shows that the AI sequential model and the CatBoost classifiers have the highest precision ratio (98%), followed by Random forest with gridsearch (94%) and logistic regression (93%). These findings indicate that the AI model did an excellent job of correctly predicting and classifying fraudulent transactions in the supply chain.

The F1-measure

The F1-score combines the measures of precision and recall into a single score and is calculated as the geometric mean of these two metrics. It is a trade-off between precision and recall, and as such, it can be a useful measure of performance in circumstances where accuracy is not an appropriate measure. The F1-measure is useful for comparing the performance of different classifiers because it is robust to changes that affect precision but do not affect recall and vice

versa. As a result, the F1-Measure is the "favoured" metric to measure model performance in ML applications. The formula for the F1-measure is shown in equation 7:

$$- = 2 * (*) / (+) \quad \text{eq. 7}$$

Table 4 shows that the classification model has excellent F1-scores. The F-value for the CatBoost classifiers is 99%, followed by the AI sequential model with 98%, Random forest with gridsearch (97%), and logistic regression (96%). Because the F1-score is a weighted average of precision and recall, the FP and FN are treated equally. Since the FPs and FNs have different costs, the F-score is useful for this study to avoid misleading interpretations of performance accuracy when working with an imbalanced dataset. The higher the F1 score, the better the models are performing in predicting fraud.

Table 4. Classification Performance Metrics

Performance Measures	Logistic Regression	Random forest	CatBoost	AI Sequential
Sensitivity (Recall)= TP/(TP+FN)	99%	100%	100%	99%
Specificity= TN/(TN+FP)	93%	94%	98%	98%
Precision =TP/(TP+FP)	93%	94%	98%	98%
F1-Score- Calculated	96%	97%	99%	98%

ROC Curve

Another statistic for measuring classification success is the receiver operating characteristic curve (ROC). The ROC plots the TPR, the FPR, and specificity against the TNR at various thresholds. An area under the ROC curve (AUROC) greater than 0.5 indicates that the classifier performs better than random guessing[31]. The ROC curve can be used to help make decisions about the appropriateness of a classifier for a given task. By evaluating the AUROC, we can determine the model's overall performance and how well it separates the observations in the dataset into fraud and no-fraud classes. A higher AUROC indicates better classification performance and is better for decision-making. A ROC analysis demonstrates how sensitivity (TPR) varies with specificity (TNR or 1- the FPR) for various thresholds[43].

Figure 10 shows the AUROC scores for the various algorithms. Note that all the classifiers covered the maximum area of the AUC plot. However, the random forest with gridsearch classifier model score (99.8%) was slightly higher than the other three classifiers. The AUROC score of all these different models indicates that all three algorithms are robust and consistent in predicting fraud in the supply chain. In ranking the observations, the AUROC identifies the probability that a randomly selected transaction will have a higher predicted risk of being fraudulent than a randomly selected transaction that is not fraudulent. The results of the AUROC back up the results of the classification metrics because they show that the models did an excellent job of spotting transactions that were not legitimate.

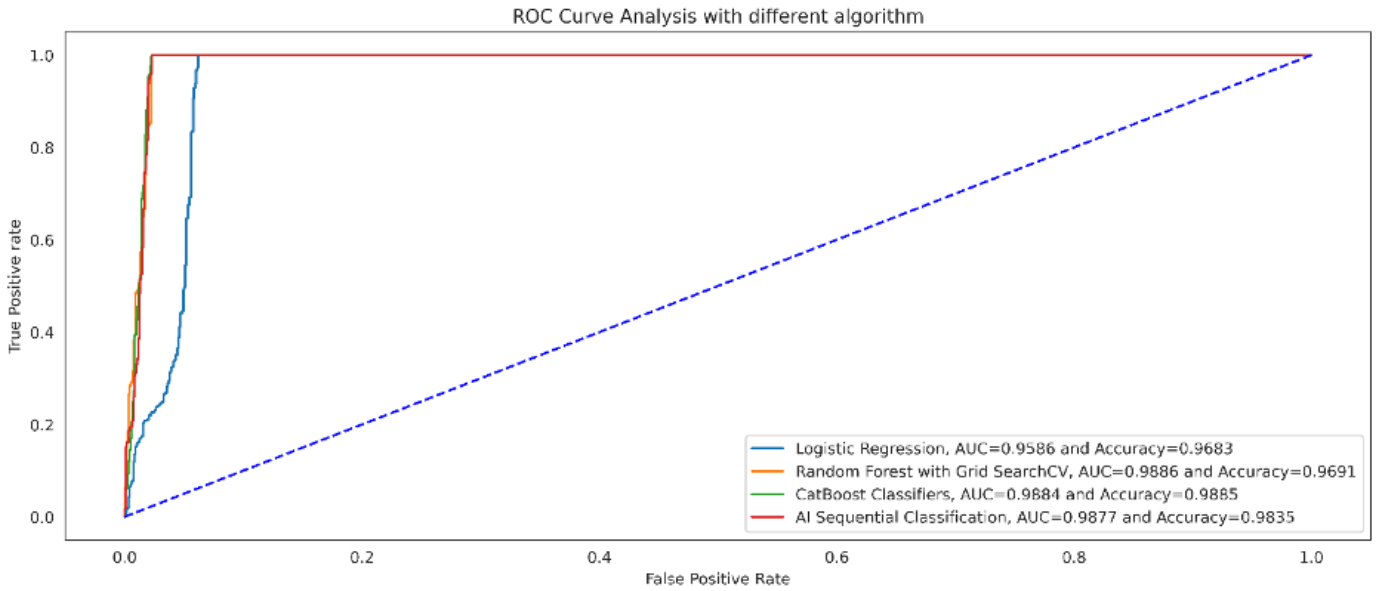

Figure 10. *AUROC Scores for all the Algorithms*

Feature importance

In ML, feature importance is a technique that can be used to determine which predictor variables are most important in determining the outcome of the dependent variable. This technique is often used with other ML methods, such as decision trees or support vector machines. Feature importance can be calculated in several ways, but one common method is to look at the coefficients of the predictor variables in the final model. The larger the coefficient, the more significant the impact of that particular predictor on the outcome variable. In many cases, feature importance can be used to identify which predictors are most important for a given problem. These predictors can then be used to build a more accurate model. Understanding which predictors are most important can make it easier to tweak the algorithm to achieve better results[10]. Figure 11 shows the features importance of the gridsearch random forest model. Note that Delivey_Status, Payment type, and Late_delivery_Risk had the highest standardized coefficients and were the most relevant predictors of supply chain fraud. The larger the coefficient, the more significant the impact these predictors have in predicting fraudulent transactions.

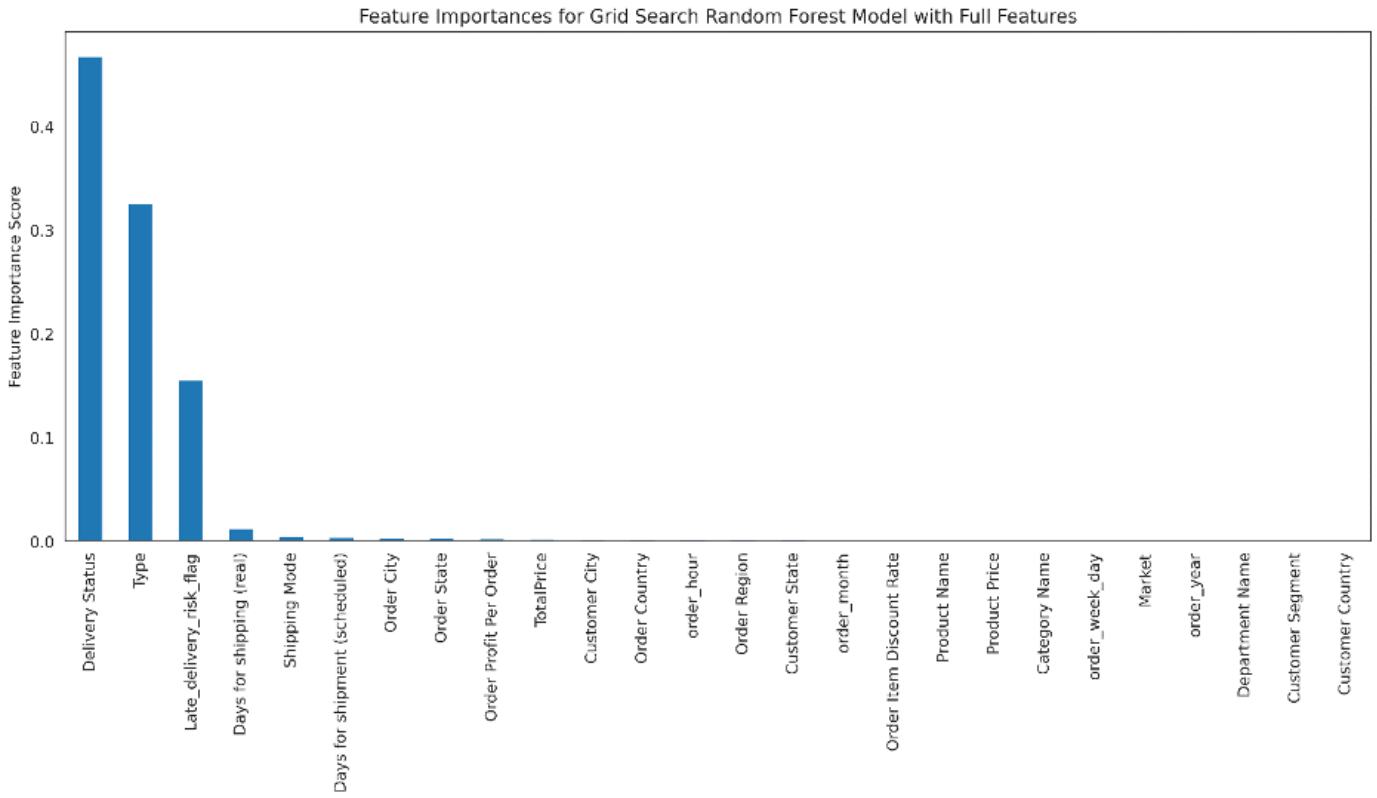

Figure 11. *Variable Importance Estimates of the model*

Conclusion

This study provides valuable insights into ML and AI for predictive modelling and highlights the potential of these techniques for fraud detection in SCM[5][12][21]. In recent years, the use of ML in supply chain management has increased due to its ability to detect patterns or anomalies from data[2]. ML algorithms can automatically identify patterns in data and use this knowledge to make predictions or recommendations about potentially fraudulent activities in the supply chain[15][20]. In this regard, ML can be used to detect anomalies in the data, such as unusual patterns and activities that may indicate instances of fraud. This paper has shown that by harnessing the power of ML and AI, organizations can improve the efficiency of their supply chains and protect against risks such as fraud and human error. As the capabilities of these technologies continue to grow, they will likely have an even more significant impact on how supply chains are managed[22]. By using the power of ML and AI, companies can make their supply chains more efficient and protect themselves from risks like fraud.

SCM is a critical part of any business, and it is important to ensure that it is free from fraud. The increasing sophistication of supply chain fraud calls for new methods of detection[3]. This paper explores the use of ML algorithms and AI to predict supply chain fraud. A logistic regression model was used as the baseline classifier to predict fraud. The other classifiers were an RFC with grind search and an AI sequential model. The results indicated that all three models consistently predicted fraudulent transactions. However, the sequential neural network model had the highest performance accuracy (99%), followed by the gridsearch RFC model (97%). The logistic regression model had the lowest accuracy (96.6%). These

results suggest that ML can be an effective tool for detecting supply chain fraud. However, further research is needed to improve the accuracy of predictions.

Model Limitations and Future Research

While ML and AI have several potential advantages for supply chain management, these technologies also have serious drawbacks. The need for a lot of data for ML and AI to function effectively is one of the main challenges. Obtaining data can be problematic in the supply chain, where data is frequently siloed and fragmented. AI and ML can also be costly to set up and maintain. Finally, there is always a chance that these technologies will not catch fraud or that they will produce false positives. Given these difficulties, it is crucial to consider the advantages and dangers of applying ML and AI to the supply chain before going forward.

The AI Sequential Model has some limitations that should be considered for the performance evaluation of a classification matrix. First, the order of the layers of the model is essential. If the order of the layers is incorrect, then the model will not work as intended. Second, the sequential model does not work well with large feature vectors. This is because the computational cost required to train a neural network increases exponentially with the size of the data. Therefore, if the data being classified are very large, then the sequential model could not classify them effectively.

As the world digitizes, the supply chain becomes more susceptible to fraud. Specifically, the proliferation of digital commerce has enabled criminals to conduct sophisticated attacks that are difficult to detect. As a result, there is a growing need for research on how ML and AI can be used to enhance supply chain fraud detection. Most fraud detection systems are based on rule-based models that cannot keep up with the rate of change in the digital world. On the other hand, ML and AI can create dynamic models that are continuously updated as new data is collected. This enables the detection of emerging fraud patterns and a rapid response to supply chain changes. In addition, ML and AI can be used to create predictive models that aid organizations in preventing fraud through proactive measures. As machine learning (ML) and artificial intelligence (AI) are used more and more in the supply chain, future research must focus on how these technologies can be used to improve fraud detection.

Declaration

Ethics approval and consent to participate

Not Applicable

Consent for publication

Not applicable

Availability of data and material

Data will be made available upon request

Competing interests

Not applicable

Funding

There is no funding for this project

Authors' contributions

Mark Lokanan prepares the main manuscript and guides the methodology and the coding sections.

Vikas Maddhesia wrote the methodology and performed the analysis.

Acknowledgements

Not Applicable

References

1. [a, b] *Baryannis G, Dani S, Antoniou G. Predicting supply chain risks using machine learning: The trade-off between performance and interpretability. Future Gener Comput Syst. 2019 Dec 1;101:993–1004.*
2. [a, b, c, d, e, f, g, h] *Constante-Nicolalde FV, Guerra-Terán P, Pérez-Medina JL. Fraud Prediction in Smart Supply Chains Using Machine Learning Techniques. In: Botto-Tobar M, Zambrano Vizuete M, Torres-Carrión P, Montes León S, Pizarro Vásquez G, Durakovic B, editors. Applied Technologies. Cham: Springer International Publishing; 2020. p. 145–59. (Communications in Computer and Information Science).*
3. [a, b, c, d] *Mao D, Wang F, Hao Z, Li H. Credit Evaluation System Based on Blockchain for Multiple Stakeholders in the Food Supply Chain. Int J Environ Res Public Health. 2018 Aug;15(8):1627.*
4. [a, b, c] *Ileberi E, Sun Y, Wang Z. A machine learning based credit card fraud detection using the GA algorithm for feature selection. J Big Data. 2022 Dec;9(1):24.*
5. [a, b, c, d] *Abbas K, Afaq M, Ahmed Khan T, Song WC. A Blockchain and Machine Learning-Based Drug Supply Chain Management and Recommendation System for Smart Pharmaceutical Industry. Electronics. 2020 May 21;9(5):852.*
6. [^] *Ni D, Xiao Z, Lim MK. A systematic review of the research trends of machine learning in supply chain management. Int J Mach Learn Cybern. 2020 Jul 1;11(7):1463–82.*
7. [a, b] *Zhou Y, Song X, Zhou M. Supply Chain Fraud Prediction Based On XGBoost Method. In: 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE). 2021. p. 539–42.*
8. [^] *Schroeder M, Lodemann S. A Systematic Investigation of the Integration of Machine Learning into Supply Chain Risk Management. Logistics. 2021 Sep;5(3):62.*
9. [^] *Wan F. XGBoost Based Supply Chain Fraud Detection Model. In: 2021 IEEE 2nd International Conference on Big Data,*

Artificial Intelligence and Internet of Things Engineering (ICBAIE). 2021. p. 355–8.

10. [a, b]*Zhang Y, Tong J, Wang Z, Gao F. Customer Transaction Fraud Detection Using Xgboost Model. In: 2020 International Conference on Computer Engineering and Application (ICCEA). 2020. p. 554–8.*
11. [^]*Rodriguez-Aguilar R, Marmolejo-Saucedo JA. Structural Dynamics and disruption events in Supply Chains using Fat Tail Distributions. IFAC-Pap. 2019 Jan 1;52(13):2686–91.*
12. [a, b, c]*Camossi E, Dimitrova T, Tsois A. Detecting Anomalous Maritime Container Itineraries for Anti-fraud and Supply Chain Security. In: 2012 European Intelligence and Security Informatics Conference. 2012. p. 76–83.*
13. [^]*Zhang W, Gao F. An Improvement to Naive Bayes for Text Classification. Procedia Eng. 2011 Jan 1;15:2160–4.*
14. [^]*Lo SK, Xu X, Wang C, Weber I, Rimba P, Lu Q, et al. Digital-Physical Parity for Food Fraud Detection. In: Joshi J, Nepal S, Zhang Q, Zhang LJ, editors. Blockchain – ICBC 2019. Cham: Springer International Publishing; 2019. p. 65–79. (Lecture Notes in Computer Science).*
15. [a, b, c]*Shahbazi Z, Byun YC. A Procedure for Tracing Supply Chains for Perishable Food Based on Blockchain, Machine Learning and Fuzzy Logic. Electronics. 2021 Jan;10(1):41.*
16. [a, b, c]*Bagga S, Goyal A, Gupta N, Goyal A. Credit Card Fraud Detection using Pipeling and Ensemble Learning. Procedia Comput Sci. 2020;173:104–12.*
17. [^]*Lezoche M, Hernandez JE, Alemany Díaz M del ME, Panetto H, Kacprzyk J. Agri-food 4.0: A survey of the supply chains and technologies for the future agriculture. Comput Ind. 2020 May 1;117:103187.*
18. [a, b, c]*Herland M, Khoshgoftaar TM, Bauder RA. Big Data fraud detection using multiple medicare data sources. J Big Data. 2018 Dec;5(1):29.*
19. [a, b]*Johnson JM, Khoshgoftaar TM. Medicare fraud detection using neural networks. J Big Data. 2019 Dec;6(1):63.*
20. [a, b]*Zhang G, Zhang X, Bilal M, Dou W, Xu X, Rodrigues JJPC. Identifying fraud in medical insurance based on blockchain and deep learning. Future Gener Comput Syst. 2022 May 1;130:140–54.*
21. [a, b, c, d]*Dua P, Bais S. Supervised Learning Methods for Fraud Detection in Healthcare Insurance. In: Dua S, Acharya UR, Dua P, editors. Machine Learning in Healthcare Informatics [Internet]. Berlin, Heidelberg: Springer; 2014 [cited 2022 Aug 22]. p. 261–85. (Intelligent Systems Reference Library). Available from: https://doi.org/10.1007/978-3-642-40017-9_12*
22. [a, b, c, d]*Bordoloi D, Singh V, Sanober S, Buhari SM, Ujjan JA, Boddu R. Deep Learning in Healthcare System for Quality of Service. J Healthc Eng. 2022 Mar 8;2022:e8169203.*
23. [^]*Li H, Li W, Pan X, Huang J, Gao T, Hu L, et al. Correlation and redundancy on machine learning performance for chemical databases: Correlation and Redundancy on Machine Learning Regressions. J Chemom. 2018 Jul;32(7):e3023.*
24. [^]*Wang E, Alp N, Shi J, Wang C, Zhang X, Chen H. Multi-criteria building energy performance benchmarking through variable clustering based compromise TOPSIS with objective entropy weighting. Energy. 2017 Apr 15;125:197–210.*
25. [a, b, c]*Ganguly S, Sadaoui S. Classification of Imbalanced Auction Fraud Data. In: Mouhoub M, Langlais P, editors. Advances in Artificial Intelligence. Cham: Springer International Publishing; 2017. p. 84–9. (Lecture Notes in Computer Science).*
26. [^]*Ma T, Song F. A Trajectory Privacy Protection Method Based on Random Sampling Differential Privacy. ISPRS Int J Geo-Inf. 2021 Jul;10(7):454.*
27. [^]*Lokanan M, Liu S. Predicting Fraud Victimization Using Classical Machine Learning. Entropy. 2021 Mar;23(3):300.*

28. ^Rushin G, Stancil C, Sun M, Adams S, Beling P. Horse race analysis in credit card fraud—deep learning, logistic regression, and Gradient Boosted Tree. In: 2017 Systems and Information Engineering Design Symposium (SIEDS). 2017. p. 117–21.
29. a, b Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. J Big Data. 2020 Dec;7(1):94.
30. a, b Zuech R, Hancock J, Khoshgoftaar TM. Detecting web attacks using random undersampling and ensemble learners. J Big Data. 2021 Dec;8(1):75.
31. a, b Lokanan ME, Sharma K. Fraud prediction using machine learning: The case of investment advisors in Canada. Mach Learn Appl. 2022 Jun 15;8:100269.
32. ^Bhattacharyya S, Jha S, Tharakunnel K, Westland JC. Data mining for credit card fraud: A comparative study. Decis Support Syst. 2011 Feb 1;50(3):602–13.
33. ^Botchey FE, Qin Z, Hughes-Lartey K. Mobile Money Fraud Prediction—A Cross-Case Analysis on the Efficiency of Support Vector Machines, Gradient Boosted Decision Trees, and Naïve Bayes Algorithms. Information. 2020 Aug;11(8):383.
34. ^Rtayli N, Enneya N. Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization. J Inf Secur Appl. 2020 Dec 1;55:102596.
35. a, b, c Bao Y, Hilary G, Ke B. Artificial Intelligence and Fraud Detection. In: Babich V, Birge JR, Hilary G, editors. Innovative Technology at the Interface of Finance and Operations [Internet]. Cham: Springer International Publishing; 2022 [cited 2022 Aug 22]. p. 223–47. (Springer Series in Supply Chain Management; vol. 11). Available from: https://link.springer.com/10.1007/978-3-030-75729-8_8
36. ^Mehbodniya A, Alam I, Pande S, Neware R, Rane KP, Shabaz M, et al. Financial Fraud Detection in Healthcare Using Machine Learning and Deep Learning Techniques. Secur Commun Netw. 2021 Sep 11;2021:e9293877.
37. ^Bhat AZ, EMA TK, Asim F. Evaluation of Neural Network Model for Better Classification of Data and Optimum Solution of Real-World Problems. J Stud Res [Internet]. 2022 Jun 1; Available from: https://www.jsr.org/index.php/path/article/view/1483
38. ^Hahnloser RHR, Sarpeshkar R, Mahowald MA, Douglas RJ, Seung HS. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. Nature. 2000 Jun;405(6789):947–51.
39. ^Dal Pozzolo A, Caelen O, Le Borgne YA, Waterschoot S, Bontempi G. Learned lessons in credit card fraud detection from a practitioner perspective. Expert Syst Appl. 2014 Aug 1;41(10):4915–28.
40. ^Shamsudin H, Yusof UK, Jayalakshmi A, Akmal Khalid MN. Combining oversampling and undersampling techniques for imbalanced classification: A comparative study using credit card fraudulent transaction dataset. In: 2020 IEEE 16th International Conference on Control & Automation (ICCA). 2020. p. 803–8.
41. ^Sahin Y, Duman E. Detecting credit card fraud by ANN and logistic regression. In: 2011 International Symposium on Innovations in Intelligent Systems and Applications. 2011. p. 315–9.
42. ^Barse EL, Kvarnstrom H, Johnson E. Synthesizing test data for fraud detection systems. In: 19th Annual Computer Security Applications Conference, 2003 Proceedings [Internet]. Las Vegas, Nevada, USA: IEEE; 2003 [cited 2022 Aug 22]. p. 384–94. Available from: http://ieeexplore.ieee.org/document/1254343/
43. ^Muschelli J. ROC and AUC with a Binary Predictor: a Potentially Misleading Metric. J Classif. 2020 Oct 1;37(3):696–708.