***Coursework 1 – Artificial intelligence***

***Dr Zied M’nasri***

**An insight into how machine learning is being used to detect and prevent fraud.**

**Haris Mahmood: 23012399**

**Adam Akhtar: 23008644**

**Abstract**

This report critiques the use of AI in online banking, specifically in fraud detection. Included is a real-life example as well as an application of machine learning and the advantages/disadvantages of using this method.

**Introduction**

Fraud is defined as the act of deception which rewards the perpetrator and denies the rights of a victim. According to (Kesseler, 2024), scammers will manipulate their victims by using methods that will deceive them, such as making them believe they are authorised personnel and therefore, approve payments. Statistics show fraud is having a destructive impact in the world of online banking; according to (Kesseler, 2024) the ‘Nasdaq Verafin 2024 global financial Crime Report’ claims there has been a projected $485.6 billion in losses in 2023 alone through digital fraud schemes. The report will explore how AI methods, such as supervised machine learning, are used to detect, and therefore, prevent fraud in the online banking space.

**Background Learning**

According to (Khokkar, 2024), twenty percent of online customers have been victim to a phishing or identity theft attempt, meaning many customers are cautious of participating in online banking. Banks rely on online banking as it puts less pressure on the physical buildings to provide for their customers in case of an event where it is not possible to utilise the building, such as the COVID pandemic. It also helps financially, as it means less wages are needed to pay staff. Therefore, it is important to bankers that the trust between the customers and the bank is not eroded. However, fraud is frequent and often successful. According to (Khokkar, 2024), Data analysed by ‘NICE Actimize’ suggests that attempted global fraud transactions jumped by 92% and attempted fraud amounts surged by 146% in 2022 compared with 2021.

AI, or artificial intelligence, is defined by (Khokkar,2024) as machines that can simulate human learning, comprehension, problem solving, decision making, and autonomy. Within fraud detection, users utilise supervised learning. This is defined as the use of labelled data sets to train algorithms that classify data or predict outcomes accurately. An example of supervised learning is labelling large amounts of data as either fraudulent or legitimate.

Detecting suspicious activity is crucial in preventing fraud. For example, banks analyse transaction patterns, and flag potential fraud when cards are used for high-value transactions. Moreover, bank will use chatbots who can detect scams. The chatbots analyse language patterns and user behaviour to detect scams, such as identifying phishing attempts and warning users about suspicious links. AI can also be used to quickly analyse real-time transaction data, identifying unusual patterns, such as multiple transactions from different locations. However, there are some downsides when it comes to implementing fraud detection method. There often are high false positive rates, which is seen as inconvenient for the customer.

**Methodology and Data:**

Financial institutes may use various methods to identify fraudulent credit card transactions. In this section, the methods and actions that financial institutions can take to prevent fraudulent activities in real time will be highlighted. Supervised learning, the earliest method of machine learning, plays a significant role for fraud detection and prevention. One of the supervised machine learning methods that can be used for fraud detection are decision trees and logistic regression. A decision tree can learn to classify transactions as either fraudulent or non-fraudulent based on features such as transaction amount, location, time of day, user behaviour, and the frequency of the transaction.

**Example – supervised learning in financial services / fraud detection**

The method trains a model with labelled information to detect the patterns that correspond to fraud. One real-life example is PayPal; they use fraud detection models to recognize fraudulent patterns and activity. PayPal also uses fraud detection as a way of scoring risk of fraudulent activities using historical data including user behaviour.

They can then apply those models to suspect or normalize new transactions for safe real-time fraud detection, which reduces the number of fraudulent activities over time. Before a transaction is flagged by the AI model, a security check will need to be passed to identify whether it has the features of a fraudulent transaction or not.

A diagram of a algorithm

Description automatically generated

Below is a representation on how the risk of a transaction is calculated and categorised into low risk or high-risk fraudulence.

A diagram of a flowchart

Description automatically generated

The process uses featured extraction of transaction data. After all the factors have been processed, the data is now prepared and can be used by the AI system to calculate if the transactions are fraudulent or non-fraudulent. The developed AI system uses this transaction data.

**Creating and Training the ML fraud detection Model**

The dataset will be processed and trained using Logistic regression.

The dataset is from Kaggle and contains 284,808 rows of credit card information, this dataset presents transactions that occurred in a two-day timespan and is highly unbalanced.

Below is the code using logistic regression. Logistic regression is a binary classifier that fits well with this dataset as we can categorise the transactions as fraudulent (1) or non-fraudulent. (0)

**Importations**

A screenshot of a computer program

Description automatically generatedFor this model we first must import the libraries needed for this dataset such as pandas for data analysis, scikit-learn for supervised learning model and in this case, importing logistic regression. seaborn and matplotlib are used for graphs and data visualization tools.

**Preprocessing the dataset**

A screen shot of a computer program

Description automatically generated

The database ‘creditcard.csv’ is loaded onto the python code. Rows with missing values are dropped from the class column using ‘data.drop.’

X – input data

Y – output data

Class 0 – non-fraud

Class 1 – fraud

**Split Data into Training and Test Sets** A black background with white text

Description automatically generated

The data is split into X and Y training and testing sets. ‘test\_size = 0.5’ refers to how much of the dataset is being trained which in this case, is 50 percent. ‘Stratify=y’ ensures the same proportion of fraud and non-fraud cases in both training and test sets.

**Standardization/scaling**

A screen shot of a computer code

Description automatically generated

This calculates the mean and standard deviation for each feature in the dataset. These values will be used to transform the features into a standardised format. X/input is used to scale then train and test the input.

**Training - Logistic Regression Model**

A computer code on a black background

Description automatically generated

This is a classification algorithm that predicts the probability of a binary outcome being, fraud vs. non-fraud using the sigmoid function.

The output of the sigmoid function is used to make predictions to calculate the probability of the credit card information being fraudulent or non-fraudulent. If the output is greater than 0.5, the outcome will be classified as 1 and below 0.5 will be classified as 0.

‘Y\_pred = model.predict(X\_test\_scaled)'

Uses the trained model to predict the class (fraud or non-fraud) for the test data.

**Data Visualization**

A computer screen with text on it

Description automatically generated

After the model has been trained, the confusion matrix is imported to show the actual vs predicted outcome. The Matrix presents the binary outcome of the data and is split into:

True Negatives - Correctly predicted non-fraud transactions.

False Positives - Non-fraud transactions wrongly classified as fraud.

False Negatives - Fraud transactions wrongly classified as non-fraud.

True Positives - Correctly predicted fraud transactions.

**Classification Report**

A black screen with colorful text

Description automatically generated

This presents the overall performance of the trained model measuring precision, accuracy and recall of the model.

**Analysis and Discussions:**

**Results**

**A graph with numbers and a number on it

Description automatically generated with medium confidence**

Above are the results of the datasets shown in a confusion matrix. A confusion matrix is a type of statistical classification, which summarises the performance of a machine learning model on a set of test data. Out of the 142,404 rows of CC information provided 222 is labelled as fraudulent. 138,794 rows are predicted to be non-fraudulent by the model. 3364 are predicted to be fraud but are not fraud and 24 are seen as actual fraud but are predicted as non-fraud.

Below is the classfication report showing the accuracy of the model. It indicates that out of the 142158 CC data, 98 percent was predicted accuretly by the model. The implemented AI model shows promising outcomes for fraud detection in real life applications due its high accuracy and percision in detecting fraud.

A screenshot of a computer screen

Description automatically generated

**Strengths**

1. **Real-time detection:** The system demonstrates an efficient approach to categorising transaction data, such as credit card information, into fraudulent and non-fraudulent, enabling rapid analysis for fraud teams to prioritise cases effectively. This not only streamlines fraud counter measure procedures but also optimises resource allocation within financial services.

**2. Fraud Prevention Accuracy:** Machine Learning identifies deviations from normal behaviour before fraud occurs at an accuracy of 98 percent. With high accuracy, the likelihood of incorrect classifications (false positives or false negatives) decreases, improving decision-making.

**Limitations**

**1. Dependence on Data Quality**: One significant criticism is that there is a high reliance on data. Predictions are only accurate if the data it is tied to is comprehensive and of the right calibre. This means the system has a vulnerability to inaccuracies when the input data is below par. Therefore, it can be said that this model has an overreliance on data.

**2. handling of Imbalanced Data:** Datasets tend to be disproportionate, with fraudulent transactions making up a very small minority of the total sets of data. Logistic regression models struggle to handle an imbalance of such magnitude which in turn leads to poor accuracy.

**Conclusions and Suggestions for Future Work:**

In conclusion, after exploring the potential of AI and its assistance in detecting and preventing fraud, there are certainly promising signs for the future. The 98% accuracy score shows there is a high accuracy in terms of results. Moreover, the results are completed much more swiftly than human detection. Overall, this makes for a more efficient program. However, the operation is heavily dependent on input data quality. Additionally, there are ethical issues assigned to this such as using credit card information for datasets which can be seen as sensitive.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**