Securing Finance System Using Deep Learning Techniques: Credit Card Fraud Detection

MSc Research Project
Cybersecurity

Sailee Wakhare
Student ID: 23100508

School of Computing
National College of Ireland

Supervisor: Niall Heffernan

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Sailee Sanjeev Wakhare

Student ID: 23100508

Programme: MSc Cybersecurity **Year:** 2024-25

Module: MSc Practicum Part 2

Supervisor: Niall Heffernan

Submission Due Date: 12/12/2024

Project Title: Securing Finance System Using Deep Learning Techniques: Credit Card Fraud Detection

Word Count: 6953 **Page Count:** 22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Sailee Sanjeev Wakhare

Date: 12/12/2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Securing Finance System Using Deep Learning Techniques: Credit Card Fraud Detection

Sailee Wakhare
23100508

Abstract

The increase of cyber-attacks in various sectors which leads to financial loss, data theft, identity theft, is raising concerns. It showcases the need to build secure, robust systems to prevent such attacks. Especially in the financial sector, the frauds occurring in credit card, need attention and application of advanced techniques to prevent them. These frauds involve complicated activities, complex patterns and large data. In this research, advanced methods are used for detecting such patterns. An effort is made to increase the capability of credit card fraud detection by implementing three robust deep learning algorithms which are Multilayer Perceptron, Long Short-Term Memory, Convolutional Neural Network and machine learning algorithm K-Nearest Neighbour. It also includes feature important analysis using SHapely Additive exPlanations technique of artificial intelligence. It helps to gain insights on contribution of individual features. These insights and the results obtained in this research align with previous research and contributes to securing financial systems by integrating advanced machine learning with eXplainable AI tools.

1 Introduction

The enhancement of technology has given rise to various inventions, discoveries and has helped humankind to grow. It has introduced convenience and ease in many regular activities specially in financial services with bill payments, bank transactions, digital sell and purchase. However, this advancement has also proven to be beneficial to the evil side. The attackers or the digital thieves have also upskilled themselves and are using innovative ways to scam people. Various types of fraud in the financial sector have been prevailing for a long period like identity theft, payment fraud, credit card fraud, etc. According to the U.S. Identity and Fraud Report by the Experian, around 61% of users have faced theft of credit card information Experian, P.L.C. (2023). Due to credit card fraud, there is a negative effect on the trust that customers have. It also leads to financial loss of the financial firms; it affects their reputation, and it becomes difficult to gain customer trust.

It has become necessary to identify and detect this fraud in early stages. The new enhancements in the field of artificial intelligence and machine learning can be used to detect patterns and predict the fraud. However, the use of traditional methods and the basic algorithms and techniques fall short when it comes to detecting and predicting fraud using the complex patterns. The datasets that are present are highly imbalanced, a smaller number of transactions are fraud among large number of non-fraud transactions. This can cause the model to perform in a biased way towards the transactions that are higher in number, which can impact the accuracy. Hence, the process of fraud detection using new techniques to address these issues is extremely crucial.

The advancement in machine learning algorithms and artificial intelligence gave rise to deep learning algorithms, techniques and methods. The challenges faced by the traditional methods were acknowledged and addressed and various balancing techniques, prediction algorithms have been developed to build effective fraud detection models.

These deep learning algorithms, specifically MLP (Multi-Layer Perceptrons), LSTM (Long Short-Term Memory), CNN (Convolutional Neural Network) along with a machine learning algorithm which is KNN (K-Nearest Neighbour) will be used to analyse their effectiveness and strengths. To make the model strong, the use of SHAP will be done to get deeper insights about the feature value.

The research questions addressed in this research are:

1. How can Machine Learning be used to secure digital finance systems by exploring different models and their techniques?
2. How can SMOTE balancing technique be used along with deep learning algorithms like MLP, CNN, LSTM for credit card fraud detection?
3. How can SHAP be used to perform feature importance analysis in credit card fraud detection?

For above research questions, below objectives will be fulfilled:

- Evaluating the performance of deep learning algorithms for credit card fraud detection using SMOTE balancing technique.
- The application of SHAP to perceive how model prediction is affected by which individual feature.
- The integration of machine learning and XAI to perform a feature importance analysis for credit card fraud detection.

In this research, a model is introduced in which, SMOTE is used for balancing the dataset, and is integrated with deep learning and machine learning algorithm for the purpose of credit card fraud detection. SHAP is incorporated with machine learning model to achieve transparency and interpretability.

The following section will include a literature review of the previous work conducted in this area of research. After that, the methodology section will contain the design and implementation done on the proposed research question. Next, the results will be evaluated in the results section and lastly, in discussion and conclusion, the research and its key findings will be summarized, and future scope will be discussed.

2 Related Work

The rise in fraud cases taking places, has made it necessary to find ways to detect these frauds in early stages. The technical teams have started to focus and rely on machine learning algorithms and artificial intelligence for this fraud detection mechanism. This literature review is a critical analysis on the past work of around twenty five research papers conducted in the area of credit card fraud detection using machine learning and artificial intelligence techniques. After conducting comparative analysis of the previous AI and ML techniques applied, results obtained, and research gaps, an approach was formed based on it. This comparative analysis consists of comparison of traditional machine learning algorithms and deep learning algorithms and usage of SHAP for explainability. This literature review is bifurcated into below sub-sections.

2.1 Fraud detection using Machine Learning Techniques

Machine learning algorithms have widely been used for the purpose of credit card fraud detection on various datasets. The algorithm like Random Forest, K-Nearest Neighbours (KNN) have proven to be simple yet powerful on smaller datasets, however, their performance quality can be questioned when it comes to huge and highly imbalanced datasets (Gayan, 2022). With respect to credit card fraud detection, four machine learning algorithms, random forest, decision tree, support vector machine (SVM) and naïve bayes were implemented on a dataset by (Sarker et al.,2024). To evaluate the results, five different metrices were used like precision, accuracy, recall, F1-score, Matthews Correlation. The results showed that random forest outperformed the other algorithms in four metrices with accuracy of 99.96%, precision of 97.43%, F1-score of 86.36% and MC of 86.36%. Another study was conducted on an imbalanced dataset with Naïve Bayes, K-Nearest Neighbour and Logistic regression. In this experiment, the imbalanced dataset was sampled using a hybrid way in which the positive class was addressed by oversampling it and negative was addressed by under-sampling it. In this case, K-NN performed well in all metrices expect accuracy. This study evaluated how hybrid sampling the performance of an algorithm Awoyemi, Adetunmbi and Oluwadare, 2017). This study shows importance of sampling of data; however, it does not address the challenges like high false positive rate or overfitting. The dataset containing transactions of customers using credit card in Europe was used in another study for credit card fraud detection using machine learning techniques like K-Nearest Neighbour, Naïve Bayes, Logistic Regression and Support Vector Machine. It showed in this case that after applying these algorithms, the Support Vector Machine gained accuracy of 99.94%, outperforming others with a very little margin (AIEmad, 2022). The experiments mentioned here, used the same dataset but applied different set of algorithms and received different results. Though these studies show that when these algorithms were applied on normal sized datasets, the challenge arises when we are dealing with huge datasets with large amount of data and real-life datasets are usually bigger. The algorithms mentioned above like Random Forest and Logistic Regression perform well with accuracy, precision but they pose disadvantage when it comes to huge datasets. These algorithms fall short when it comes to detecting non-linear patterns, or hidden patterns or temporal data. These challenges like data imbalance, model overfitting, need to handle large and complex data with different patterns, gives rise to the need of applying more advanced algorithms or deep learning algorithms.

2.3 Fraud detection using Deep Learning Techniques

The studies have shown that usage of Multilayer Perceptron (MLP) for fraud detection has proven to be beneficial when compared to other traditional machine learning algorithms. They have higher adaptability, and they can adapt to complex patterns in the dataset by learning from previous patterns observed. MLP when applied to a balanced dataset provides

better results than machine learning algorithms. There is some scope of research in cases of highly imbalanced dataset and how can different techniques be used to achieve better results.
Convolutional Neural Network (CNN) is related to image processing and is being widely used for credit card fraud detection since it is capable of detecting hidden patterns in data (Nguyen, *et al.* (2020)). On an imbalanced data, convolutional neural network has experienced high false positive rate, so the challenge of imbalanced dataset persists. The next deep learning algorithms is Long Short-Term Memory (LSTM) which is generally used for that data which is sequential in nature. Due to which, LSTM algorithm when used with the sampling method SMOTE has proven to provide better results which makes it better for considering the temporal patterns (Sulaiman, Nadher and Hameed, (2024)). LSTM is able to perform anomaly detection since it can process sequences that are long. The application of LSTM on smaller datasets seems challenging since the problem of overfitting arises. CNN and LSTM have been implemented in previous work and prove to be beneficial when used with balanced datasets, these studies can be tested further on real-time datasets. An experiment conducted using deep learning algorithms for credit card fraud detection suggested that deep learning algorithms like CNN, LSTM perform significantly better than traditional methods. In this study, three different sampling methods were compared which were Random Under Sampling (RUS), Near Miss Sampling and SMOTE (Synthetic Minority Oversampling Technique). It was observed that application of SMOTE improved the results of recall and fraud cases were detected. On comparing the algorithms, LSTM performed well by achieving a F1-score of 86.85%. The comparison was carried out three datasets. This study highlighted the future scope to include different hyperparameter techniques for more obtaining more optimized results (Nguyen, *et al.* (2020)). A research conducted on use of MLP for credit card fraud detection on imbalanced dataset, provided in depth insights about how can MLP be effectively leveraged in this field. It demonstrated how sensitivity was affected after increasing layers and using activation functions.

2.4 Challenges in Credit Card Fraud Detection

- The review of previous studies suggest that the crucial challenge is the imbalance present in datasets. In real life datasets, the actual fraud cases are almost less than 1%, which makes it a highly imbalanced dataset (Gayan, 2022). Due to this, models are unable to perform better since their accuracy is affected by the imbalance. In this case, various balancing techniques like SMOTE come into picture, which can be used to balance the dataset.
- The datasets of credit card fraud detection contain sensitive information like credit card details, owner name and personal details, details of transactions. Due to this, there are rarely any real-life datasets available to test the models created for fraud detection. To tackle this, generally synthetic datasets are used for training.
- The evolving technology around credit card fraud have been used by the attackers to use new and unique ways to carry out a fraud without getting detected in any patterns. It is necessary to use advanced techniques to detect such frauds.

2.5 Model Interpretability using SHAP

The use of deep learning methods like MLP, LSTM, CNN, have proven to give better results, however, the internal working of such advanced level algorithms is not always interpretable by entity like human, this removes transparency from the model, and it becomes difficult to understand how the results were obtained or which decision led to obtain which result. This is where Explainable AI (XAI) steps in. The use of XAI tools like LIME and SHAP helps to provide some ease in understanding the interpretability and decision-making process behind the results. It helps to highlight the impact different features have and criteria used to select the specific features like distance from home, lack of PIN number, chip usage, etc (Swetha, 2024). A framework was developed to evaluate different XAI methods like SHapley Additive exPlanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME), Anchors Explanations (Anchor) and Diverse Counterfactual Explanations (DiCE) for credit card fraud detection. This study concluded that there was not one single method that stood out, rather different methods worked differently based on metrices. It was revealed that SHAP outperformed the other methods in the metrices of identity and similarity (Raufi, Finnegan, C. and Longo, 2024). A research conducted involved six machine learning algorithms like Logistic Regression, Support Vector Machine, Random Forest, AdaBoost, AdaBoost with Random Forest as base estimator and AdaBoost with Logistic Regression as a base estimator. This study also involved SHAP to demonstrate explainability. Using SHAP the impact of features and their importance was highlighted to ensure trust in this research (Biswas et al., 2023).

This literature review contains study on importance of using machine learning algorithms in the field of credit card fraud detection, the challenges faced when suing traditional algorithms are addressed and studies. It also focuses on critical review of advanced deep learning algorithms, their applications and comparison of results to identify an approach for the implementation. It also addresses the challenge of interpretability which arises due to the black - box nature of deep learning algorithms that are difficult to interpret. This review includes research which focuses on different Explainable AI (XAI) tools to tackle this issue and increase transparency of models by finding out the decision-making elements or impact of different features on the results obtained.

3 Research Methodology

This section of Research Methodology includes the approach as well the design specifications considered for this research. It provides a structured approach which was refined and followed to achieve the end results of credit card fraud detection. The data mining methodology KDD (Knowledge Discover in Database) is utilized for here in which using data, information and patterns are extracted from that data. A comprehensive process including stages like data understanding, collection, preprocessing, etc is elaborated in this section.

Below is the diagram used to show the methodology approach and its phases for predicting credit card fraud detection in Figure 1:

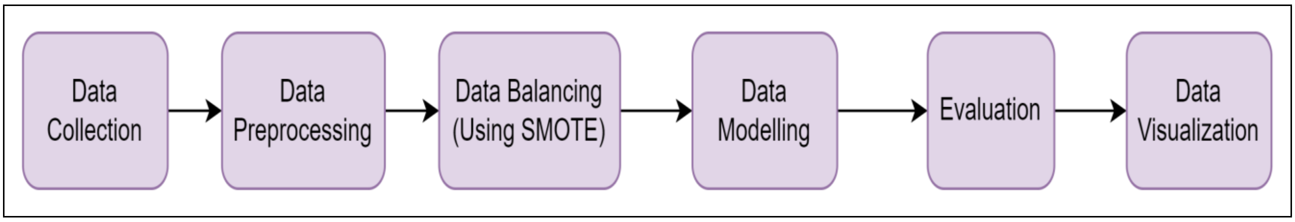


Figure 1. Methodology

Data Collection/Data Description – In this stage, various available datasets were studied and analysed to understand the importance of relevant features in a credit card dataset. After thorough research and consideration, the datasets with relevant features and size were shortlisted. A good quality dataset was selected which contains data of credit card frauds and its details.

Credit Card Fraud Detection Dataset

The dataset selected is available on Kaggle (Emma, no date). It is a highly imbalanced dataset with containing a smaller number of fraud cases, which makes it suitable for the approach of using balancing method SMOTE. The dataset contains various columns like:

Feature Name	**Meaning**
AcountNumber	The account number of the customer
CVV	Card Verification Value
CustomerAge	The age of the customer
Gender	The gender of the customer
Marital Status	The marital status of the customer
CardColour	The card color of the customer
CardType	The card type of the customer
Domain	The domain of the customer
Amount	The amount of the transaction
AverageIncomeExpendicture	The average of income expenditure done by the customer
Outcome	The target variable denoting if the transaction is fraud or not
Customer_City_Address	The city of the customer

Table 1: Details of Dataset

Data Preprocessing – The dataset selected has 12 rows. From these, six rows have numerical values, and 6 rows have categorical values. In this stage, data cleaning was carried out. By inputting the mean values in missing places, the missing and null values were addressed. The dataset was checked for duplicate values and if any duplicate values were present, they were removed from the dataset, to clean it. In order to get robust statistical results, the outliers were removed from the dataset using the Interquartile Range method. Then the data is normalized using the feature scaling method through StandardScalar. It makes sure that there is stability in dataset.

Data Balancing – In this approach, the challenge of data imbalanced is handled by using the balancing technique called Synthetic Minority Oversampling technique (SMOTE). In an imbalanced dataset, the model favours the class that is in majority, this affects the quality of results. To tackle this imbalance, SMOTE is used where, it creates or generates new values to match the feature space.

Modelling – This stage consists of the deep learning and machine learning algorithms to be applied on the dataset. For this research, Multilayer Perceptron, Convolutional Neural Network and Long Short-Term Memory model, these three deep learning algorithms and K-Nearest Neighbour machine learning algorithm is applied on the dataset obtained after applying SMOTE to it. After that, the hyperparameter tuning is done on the model with the help of random search optimization. The application of XGBoost is done and SHAP is applied on it for interpretability.

Evaluation – The assessment of the results obtained after applying algorithms on the dataset, is analysed in this stage. The performance of each algorithm is critically evaluated with respect to key metrices. The evaluation metrices considered for this research are accuracy, precision, recall, F1-score, confusion matrix. These values of each model will demonstrate the quality of results obtained.

Data Visualization – The data visualization performed in order to get the curve of accuracy and loss, arising after the epoch in the modelling phase. This is useful in plotting a graph of loss and accuracy.

This is the methodology followed which includes the above stages of data mining.

Setup and Tools Used – Below were the tools and libraries used for carrying out this implementation:

Environment – The Jupyter Notebook was used to code and execute all the deep learning and machine learning algorithms. These were executed in Python programming language.
In Jupyter notebook, virtual environment was created to prevent conflict of library versions, to maintain clean workspace and isolating the environment.

Libraries – The required libraries provided by Python for running these algorithms were downloaded, installed and used which included scikit-learn, Keras, TensorFlow, XGBoost. SMOTE uses certain libraries like imblearn or other libraries like numpy.

4 Design Specification

The architecture that is used for the execution is provided in this section. It includes the related requirements and end-to-end design for fraud detection which includes machine learning algorithms, deep learning algorithms and explainable AI methods. This framework is for predicting credit card fraud on an imbalanced dataset by applying these methodologies and applying SHAP on XGBoost.

The dataset available on Kaggle was downloaded, extracted and imported into Jupyter Notebook. The first data preprocessing step was carried on it by cleaning the data. The cleaning process involved filling the values that were absent, filling null values, removing the duplicate values, and shuffling the data. The outliers were removed to maintain data stability. After applying feature scaling on it, the data was normalized. Since the dataset that is selected here is an imbalanced dataset, the balancing techniques SMOTE was applied on it to generate new values and remove imbalance of class. After the pre-processing the method of feature importance analysis was applied through SHapley Additive explanation (SHAP) and this was done on XGBoost algorithm. This gave the information about the important features. Once the data was cleaned, processed, balanced, the next step was to split it. The data splitting took place, and dataset was split into two sets, training set and testing set. The dataset was divided such that training set was 80% and testing set was 20%. The model was tuned with the best parameters. The application of advanced Deep Learning Algorithms like Multilayer Perceptron, Convolutional Neural Network and Long Short-Term Memory and Simple machine algorithm K-Nearest Neighbour was applied to the training dataset. The testing data that was obtained before, was used to assess/evaluate the output received from these trained deep and machine learning algorithms and to measure their performance. The data visualization was carried out by plotting the loss and accuracy curve of each deep learning algorithm.

Below is the workflow diagram for the design specification:

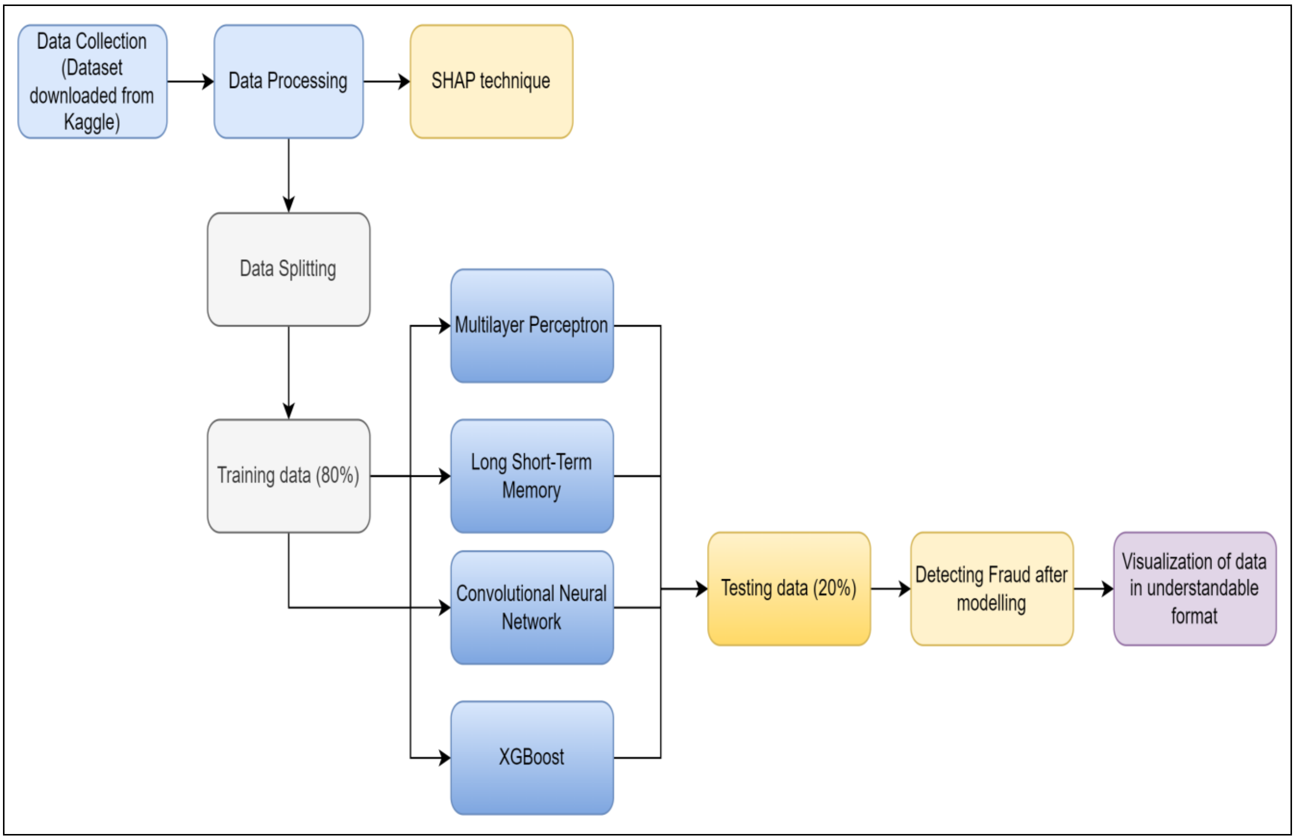

Data Collection (Dataset downloaded from Kaggle)
Data Processing
SHAP technique
Data Splitting
Training data (80%)
Multilayer Perceptron
Long Short-Term Memory
Convolutional Neural Network
XGBoost
Testing data (20%)
Detecting Fraud after modelling
Visualization of data in understandable format

Figure 2. Design Specification

5 Implementation

The implementation section provides the details of how the proposed models were applied to the selected dataset, what parameters were considered and provides elaborated explanation of the execution. The application of deep learning algorithms and machine learning algorithms are as follows:

5.1 Multilayer Perceptron

The deep learning algorithm MLP has proven to be efficient for credit card fraud detection. MLP is considered to be a neural network type which consists of multiple layers of neurons. The activation functions which are non-linear in nature are used for activation by neurons present the layers. This is how the network can learn about complex patterns in the dataset (Jaiswal, 2024). The data of credit card fraud contains complex patterns, anomalies which can be detected using Multilayer Perceptron, since MLP can detect hidden patterns, and can handle huge data when paired with balancing technique like SMOTE.

After the data is cleaned, pre-processed, hyperparameter tuning is performed on the dataset using the Randomized Search method, to select the best parameters and get better results. Through this, different parameters were analysed, and, in the implementation, the hidden_layer_sizes parameter was tested. This parameter shows how many neurons will be there in each hidden layer. There is one hidden layer which contains 50 neurons and another hidden layer consisting of 100 neurons. Adjustments were performed on activation functions like (tanh, relu) which are Hyperbolic Tangent and Rectifies Linear Unit respectively. This was for hyperparameter tuning to find the best parameter and to prevent overfitting. For model implementation, there are layers, in this case, the features that are obtained post preprocessing are with respect to the number of neurons. And the number of target class decides the output neuron number. One epoch in machine learning is when one a dataset passes through or completes one model. Such 50 epoch are set for the model to trained on selected parameter in this implementation. As the epoch proceeds, the parameter warm_start is used to trace the loss and accuracy curve which will be used for plotting in data visualization. After this the model is tested with test data and evaluation results are obtained in the form of evaluation metrices like accuracy, precision, F1-score, classification report, confusion matrix, etc

5.2 Long Short-Term Memory Model

The Long-Term Memory (LSTM) is an enhanced RNN (recurrent neural network), the long-term dependency challenge of RNN is addressed by the LSTM model. LSTM model can hold data for a longer period with the help of cell for memory it has (Aakarshachug, 2019). This advantage of long-term dependency makes it suitable for temporal data. Here, it is used for classification of transactions. LSTM generally has three layers.

The pre-processing of dataset including all the activities for cleaning it, removing outliers using IQR, standardization, is carries out. The encoding is performed on the variables that are categorical. SMOTE is applied to address and manage data imbalance. In this implementation of LSTM, there are three layers which are used to capture the temporal dependency. Here, in the code, LSTM has 128 units its first layer and the return_sequences=True signifies that output will be used to next layer. In order to make stabilize training even more, the process of batch normalization is applied in this case after every LSTM layer, and to address and prevent the issue of overfitting, 50% neurons are deactivated and long with that dropout layers are implemented after every long term-short memory layer. In the MLP implementation, for training the model, the techniques used here are early stopping and learning rate schedular. If there is a case that even after 10 rounds, the performance is not increasing then the training is stopped. This is done to prevent the risk of overfitting. The second technique, learning rate schedular is responsible for decreasing the learning rate after 20 rounds to half. For this model, the epoch is considered to be 50. After the implementation is done, the evaluation is carried out by testing it with test data and using evaluation metrices.

5.3 Convolutional Neural Network

The deep learning algorithm Convolutional Neural Network (CNN) is a special type of algorithm. It deals with the field related to object recognition, detection, classification in images. CNN is beneficial compared to other traditional algorithms because it does not rely on manual intervention for feature engineering, rather it is capable of large-scale feature extraction with efficiency. It has the ability to detect and extract patterns no matter what if there are variations in position or formation or scale (Keita, 2023).

The preprocessing of dataset is completed which includes cleaning of data, normalizing it, performing label encoding, data imbalanced is addressed with SMOTE. Coming to the CNN model implemented here, it uses sequential form in which the 1D convolutional layer is the first layer and it contains 64 filters and has a kernel of size 2. For the model to be able to capture complex data and its patterns, the activation function ReLU is applied. ReLU is also used for the purpose of introducing non-linearity. After this, the next layer is of MaxPooling 1D layer which is necessary for decreasing the spatial dimensions, but at the same time preserving the features that are required and essential. Like LSTM, overfitting is addressed here by randomly deactivating out neurons, this is achieved by a dropout layer of 0.5. The output obtained here is then flattened using flatten function and this is given to dense layer which has ReLU function and 50 neurons. The last layer consists of an activation function called softmax which gives the output. For achieving optimized results, the Adam optimizer is used for compiling the model. The model is trained with batch size of 32 and it is trained for 100 epochs. Multiple iterations will allow the model to learn better. The data visualization is carried out by plotting the loss and accuracy curve at every epoch and evaluation is carried out using various metrices like confusion matrix and accuracy.

5.4 K-Nearest Neighbour

KNN is a supervised learning algorithm. It is a simple algorithm which uses the proximity to form decision and classify, predict the outcome (What is the k-nearest neighbors algorithm?, 2024). It is used for classification and regression. A data known as training data is taken where the co-ordinates are classified into groups which are known by an attribute. The next data known as test data is taken, with the help of training set, the points of test data are allocated to a group. These points when plotted on graph, form clusters. A point can be classified into a group by considering which group is nearest to this point (Geeksforgeeks, 2017). After the initial phase of preprocessing, cleaning, standardizing data, the splitting of data is done into 80% for training and 20% for testing. SMOTE is applied to obtained a balanced dataset. For K-NN implementation, here the n_neighbors are considered as 5 which indicates that five nearest neighbors are considered in proximity to classify a given data point. Uniform weightage is applied to the 5 neighbors. The training dataset is resampled. The testing dataset is sued to form classification. Here, the evaluation metrices used to evaluate the performance of the algorithm are precision, recall, F1-score, in the form of a classification report. Confusion matrix is used which give the details about true positives, true negatives, false positives and false negatives.

5.5 SHAP for Feature Importance Analysis

SHapely Additive exPlanations also known as SHAP is a method which is used to explain the output of an algorithm, it is used in feature importance analysis. SHAP is used to find out values contributed by each feature to the results of a model. It is used along machine learning model, and it provides information on how and which feature had an impact on the final

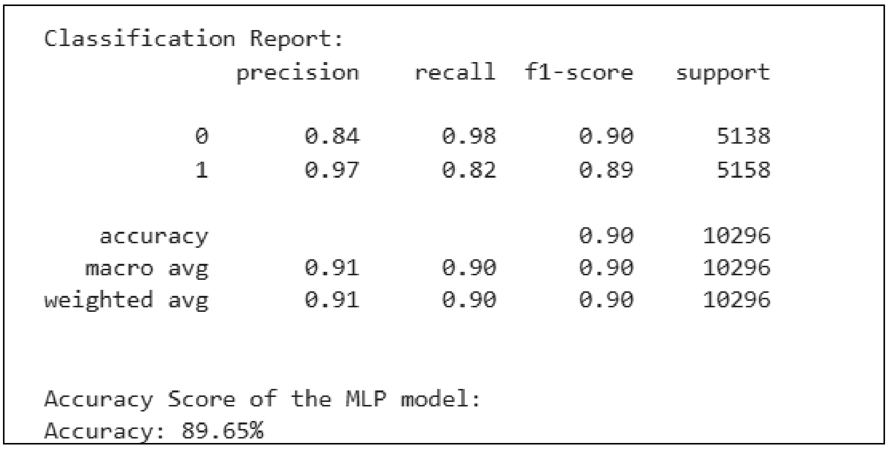
result obtained. It performs complex tasks that are difficult for humans to carry out (Mikesuperman, 2023).
In this implementation, the SHAP is applied on XGBoost algorithm which is a tree-based algorithm. The data preprocessing carried out by cleaning of data, encoding, removing outliers is completed. The data is split into training and testing set and SMOTE is applied for maintaining data balance. The dataset obtained after this is used for XGBoost. The predictions made by XGBoost are interpreted using SHAP. The trained model obtained is used along with training data and using this the shap.Explainer is initialized, after which value for every feature is calculated by SHAP. These values are used to indicate how much each feature contributed to the result. After getting the overall contribution by taking mean, the percentage contribution for each feature is calculated which is easy to understand the decision-making process of the model. A DataFrame containing percentage contribution of features in an organized manner is formed. Visualization tools like graph is used to highlight the importance of features and contribution percentage.

6 Evaluation

Evaluation is used to quantify the quality and performance of the models applied to the dataset. The results and evaluation of each algorithm is discussed in below sub-sections.

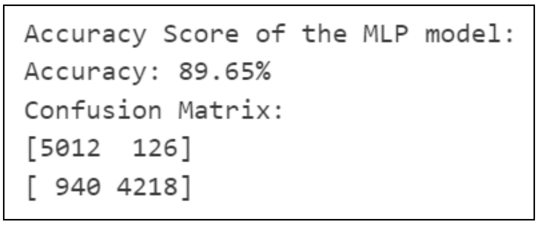
6.1 Evaluation for Multilayer Perceptron

The evaluation is performed based on evaluation metrices. The accuracy obtained was the important one to consider. For MLP, an accuracy of 89.65% was obtained. The classification report is calculated for fraud as well as non-fraud cases. Considering class 0 first which indicates the non-fraudulent transactions, here the precision obtained is 0.84 recall is 0.98 and F1-score is 0.90. This reveals that 84% of transactions were predicted correctly as non-fraud from all transactions predicted to be non-fraud. And recall reveals that out of actual non-fraud transactions, 98% were correctly predicted. For class 1 which denotes the fraudulent transaction, the precision is 0.97, recall is 0.82 and F1-score is 0.89. This reveals that 09% transactions predicted as fraud were actually fraud and 82% of fraud transactions were identified. A confusion matrix is used here to get elaborate results about the true positives, true negatives, false positives and false negatives. The confusion matrix and classification report and its values are as below in figure.



```
Classification Report:
              precision    recall  f1-score   support

           0       0.84      0.98      0.90      5138
           1       0.97      0.82      0.89      5158

    accuracy                           0.90     10296
   macro avg       0.91      0.90      0.90     10296
weighted avg       0.91      0.90      0.90     10296


Accuracy Score of the MLP model:
Accuracy: 89.65%
```

Figure 3: Classification Report



```
Accuracy Score of the MLP model:
Accuracy: 89.65%
Confusion Matrix:
[5012  126]
[ 940 4218]
```

Figure 4: Confusion Matrix

The training accuracy curve obtained shows that with each passing epoch, there is a consistent increase in the accuracy, which means the model is successfully learning from the iterations and is providing better results with each epoch. The training is stable since there are no drops in the curve. The training loss curve shows a drop at the start indicating that the model is learning from the data and is getting adapted to it. Optimization is performed effectively which is depicted by the steady decline. The absence of drops or highs maintaining smoothness of curve shows that there is no over fitting.

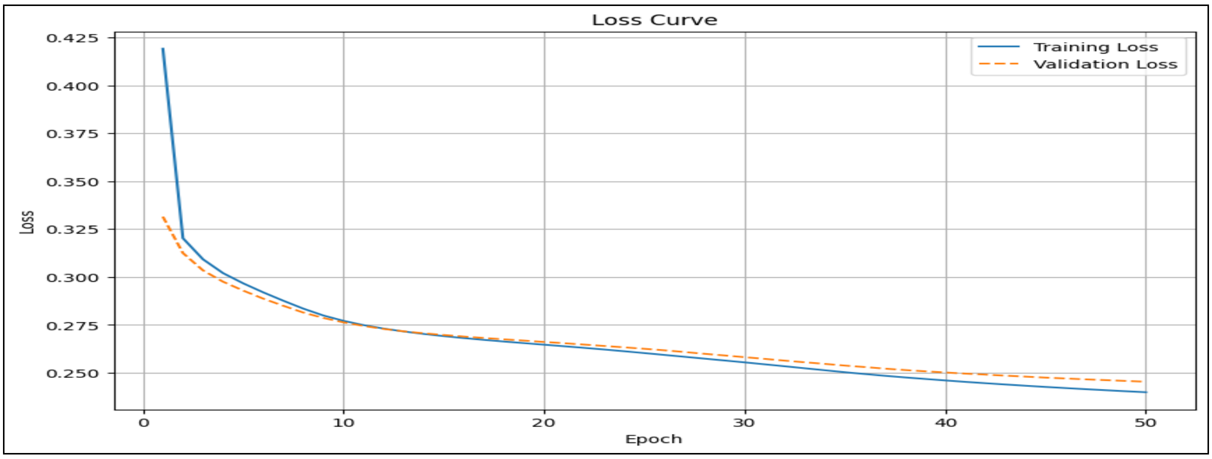


Figure 5: Loss Curve

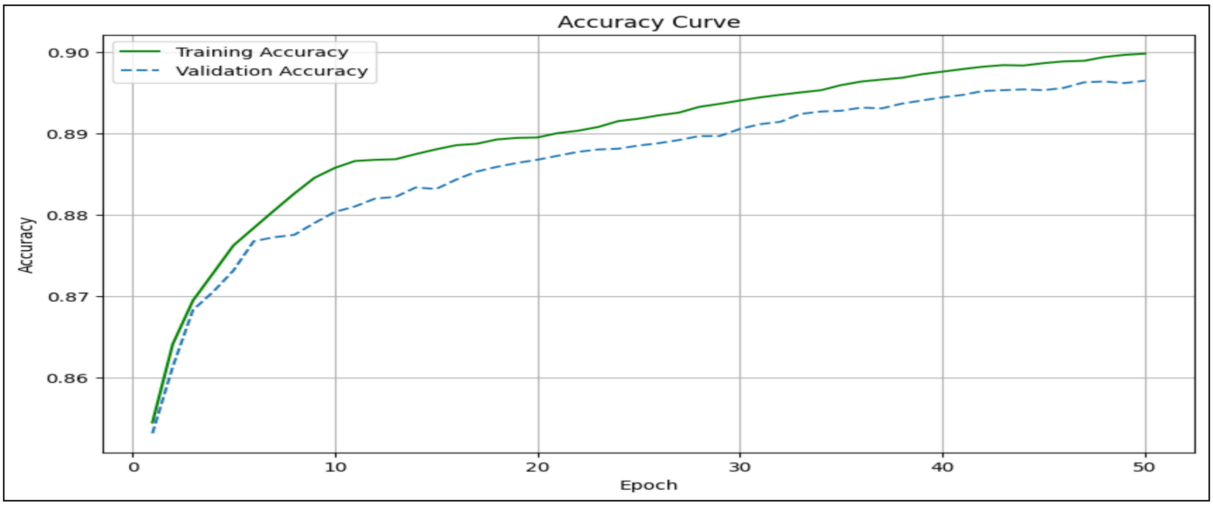
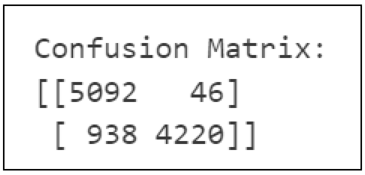


Figure 6: Accuracy Curve

6.2 Evaluation of Long Short-Term Memory

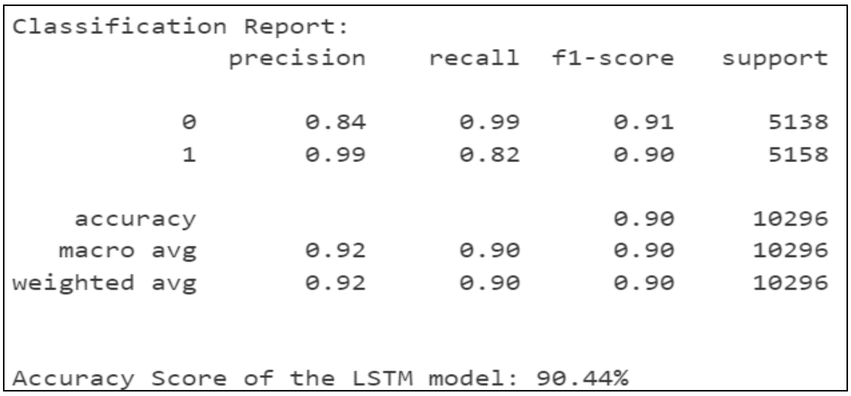
The accuracy obtained for LSTM is 90.44%, which is the highest accuracy, indicating that this is a robust model in predicting fraud and non-fraud transactions. In the classification report, lets consider class 0 first. Here the precision is 0.84, recall is 0.99, F1-score is 0.91. High value of recall indicates that this model successfully found 99% of non-fraudulent transactions and there was lesser number of false positives. For class 1 of fraudulent transactions, precision of 0.99 shows that model is efficient in detecting fraudulent transaction. The confusion matrix is depicted in below fig.



```
Confusion Matrix:
[[5092   46]
 [ 938 4220]]
```

Figure7: Confusion Matrix

The classification report is shown in below figure.



```
Classification Report:
              precision    recall  f1-score   support

           0       0.84      0.99      0.91      5138
           1       0.99      0.82      0.90      5158

    accuracy                           0.90     10296
   macro avg       0.92      0.90      0.90     10296
weighted avg       0.92      0.90      0.90     10296


Accuracy Score of the LSTM model: 90.44%
```

Figure 8: Classification Report for LSTM

Across the epochs, the training accuracy curve shows gradual rise in improvement with respect to training and validation. The end of the curve faces stable line, which shows that the model is optimized and reached optimal performance., achieving accuracy of almost 90%. The model is successful in learning patterns in the data. The loss curve with initial steep decline and stability after several epochs demonstrates absence of overfitting and it shows the effectiveness of LSTM model for credit card fraud detection.

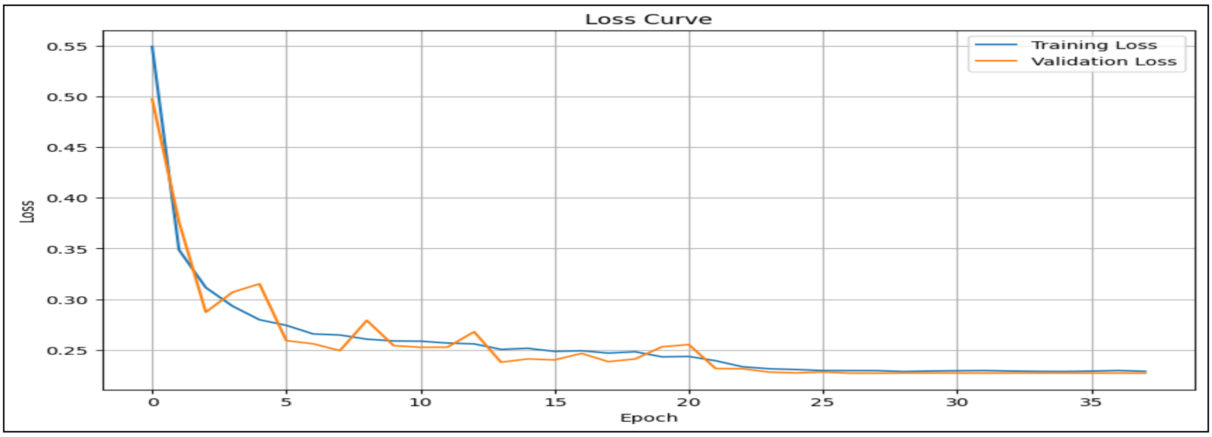


Figure 9: Loss Curve

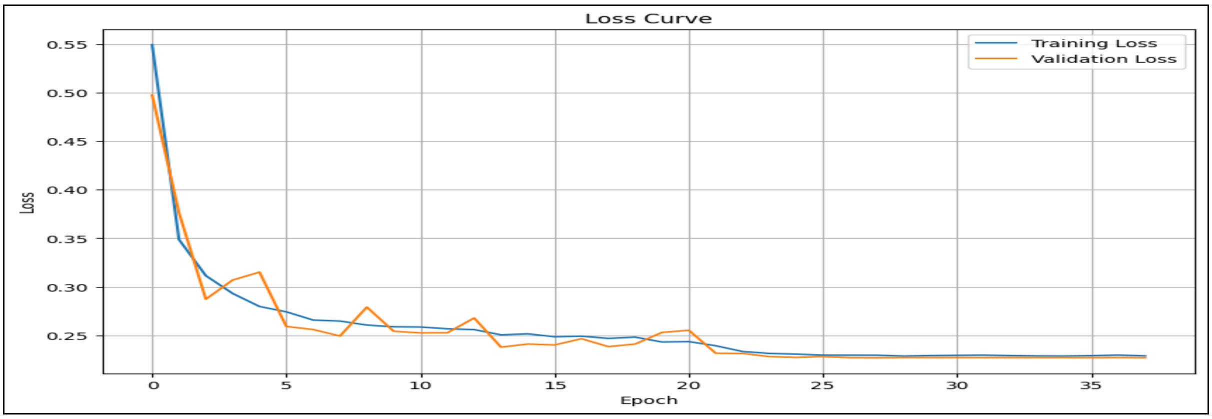
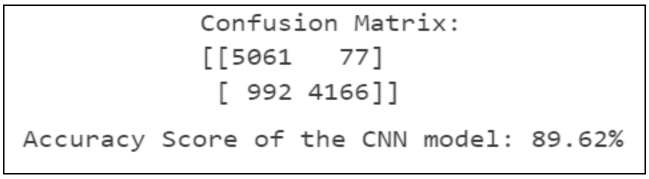


Figure 10: Accuracy Curve

6.3 Convolutional Neural Network

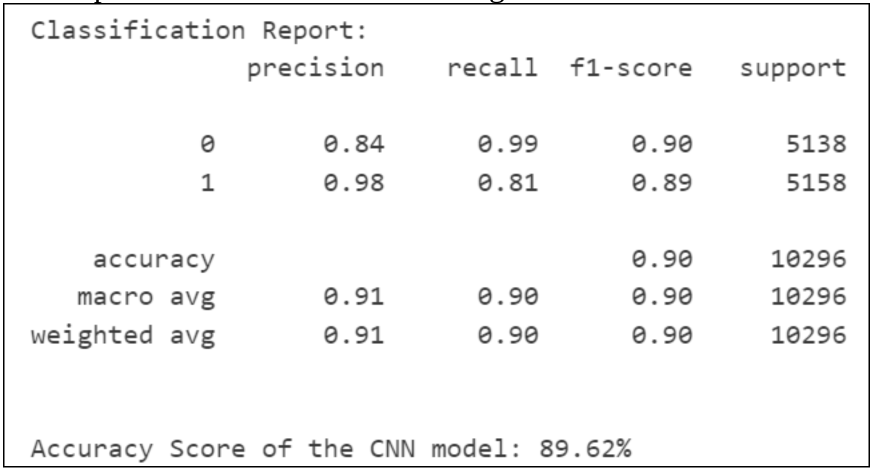
For CNN, it is observed that for class 0, the precision value is 0.84, recall value is 0.99, F1-score is 0.90. For class 1, the precision value is 0.98, recall value is 0.81 and F1-score is 0.89. The higher precision value for class 1 reveals that the model was able to correctly identify the fraudulent transactions from overall transactions. The recall value for both the classes depicts the ability of model in detecting non-fraud transactions but the reduced efficiency in detecting fraud transactions. The accuracy of CNN is 89.62%. The confusion matrix is depicted in below fig.



```
Confusion Matrix:
[[5061   77]
 [ 992 4166]]

Accuracy Score of the CNN model: 89.62%
```

Figure 11: Confusion Matrix

The classification report CNN is shown in below figure:



```
Classification Report:
              precision    recall  f1-score   support

           0       0.84      0.99      0.90      5138
           1       0.98      0.81      0.89      5158

    accuracy                           0.90     10296
   macro avg       0.91      0.90      0.90     10296
weighted avg       0.91      0.90      0.90     10296


Accuracy Score of the CNN model: 89.62%
```

Figure 12: Classification report for CNN

In the accuracy curve the gradual increase is observed with every epoch. Also, less gap between training and validation suggests that the model is robust since the overfitting is minimized. Further, there are some changes which might suggest need of optimizing the model. The loss curve also shows stability in later iterations with initial decrease, this shows normal behaviour of model to first learn from new data and then optimize. The model learns from data.

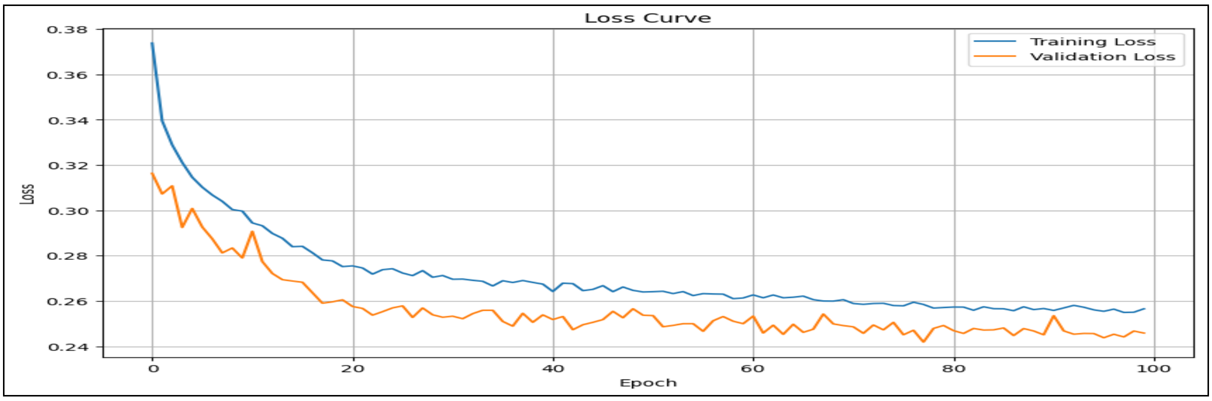


Figure 13: Loss Curve

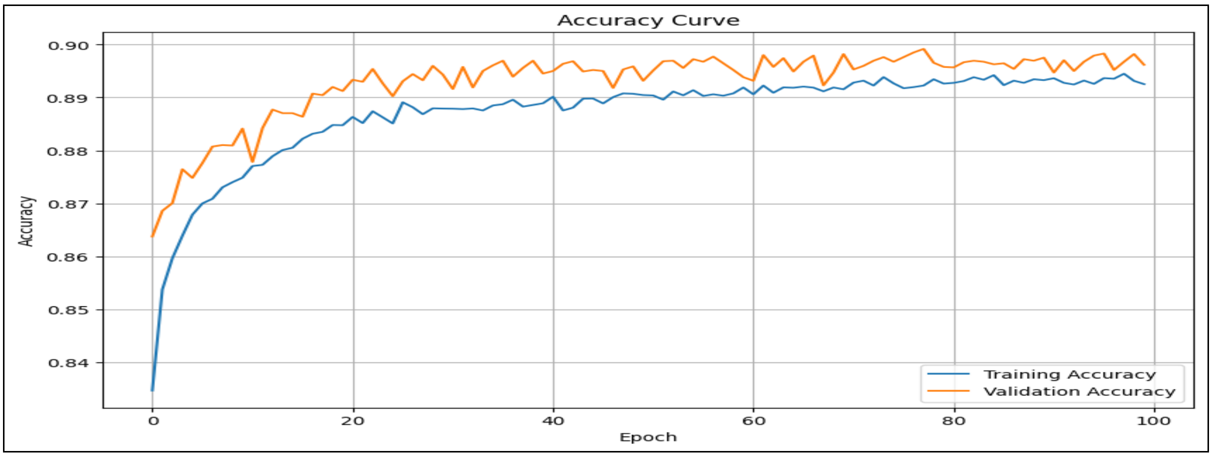
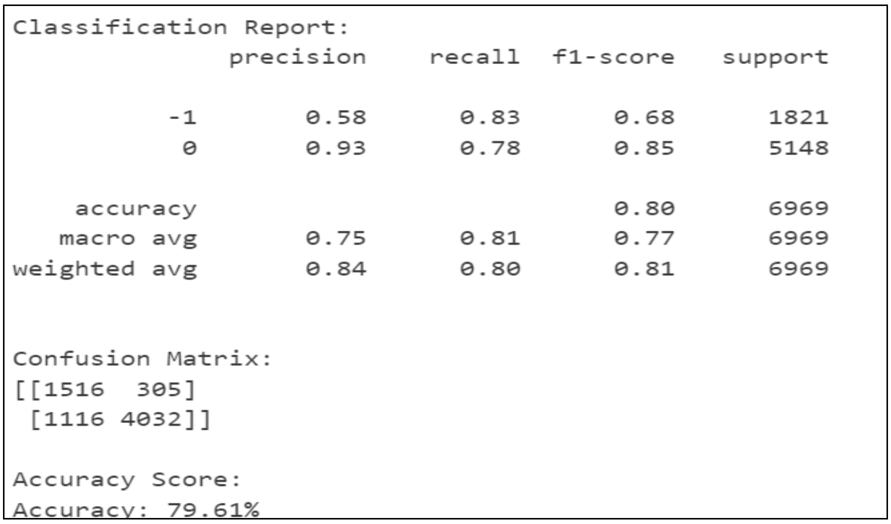


Figure 14: Accuracy Curve

6.4 Evaluation of K-Nearest Neighbour

The KNN is a machine learning algorithm for which the evaluation metrices considered are accuracy, classification report including precision, recall, F1-score, and the confusion matrix. The accuracy obtained for KNN is 79.61%. It is low when compared to deep learning algorithms. In the classification report, for class 1 the precision value is 0.58, the recall is 0.83 and for class 0 the precision is 0.93, recall is 0.78. This shows that the model is not effective in detecting the fraud transactions and is able to detect non-fraud transactions. Below is the classification report for the same.

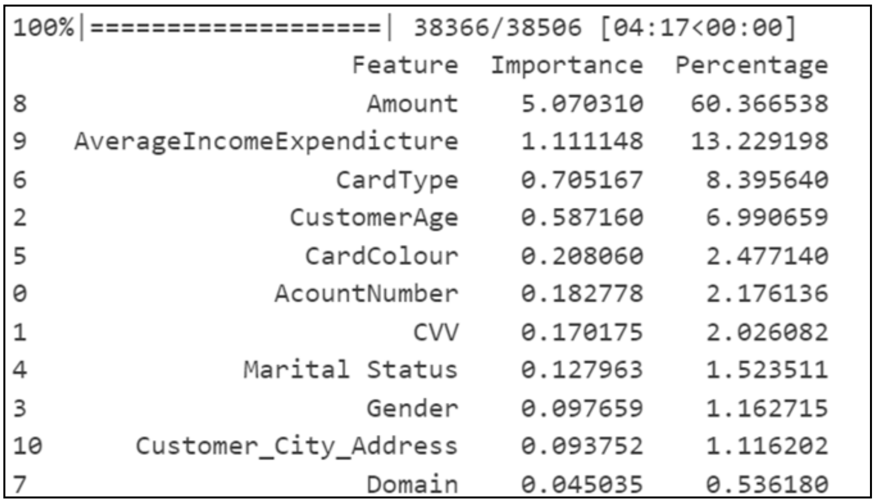


```
Classification Report:
              precision    recall  f1-score   support

          -1       0.58      0.83      0.68      1821
           0       0.93      0.78      0.85      5148

    accuracy                           0.80      6969
   macro avg       0.75      0.81      0.77      6969
weighted avg       0.84      0.80      0.81      6969


Confusion Matrix:
[[1516  305]
 [1116 4032]]

Accuracy Score:
Accuracy: 79.61%
```

Figure 15: Classification Report

6.5 SHAP - Feature Importance Analysis

The feature important analysis performed using the method SHAP on XGBoost revealed valuable insights on which features contributed towards the output and the decision-making process. In this analysis, it was revealed that the feature Amount has the highest contribution percentage with 60.37%. It denotes that the amount of transaction has strong influence on the model. It shows that fraud transactions contain patterns that are detectable by the transaction amount. The next feature with second highest percentage is the AverageIncomeExpenditure with 13.23%. It depicts that, patterns observed in expenditure of an individual and his finances, prove to be important in detecting fraud transaction. The CardType has 8.40% contribution and CustomerAge has 6.99% contribution. The CustomerAge may denote that there is certain demographic patterns are also related to fraud transactions. The contribution percentage for CardColor is 2.48% and for AccountNumber it is 2.18%. These features have less impact on the results. Significantly less contribution is showcased by the features CVV, Marital Status and Gender with 2.03%, 1.52%, and 1.16% respectively. Two features have shown least percentage which are Customer_City_Address and Domain with 1.12% and 0.54% respectively. The percentage contribution obtained is shown in below fig.



```
100%|===================| 38366/38506 [04:17<00:00]
                      Feature  Importance  Percentage
8                      Amount    5.070310   60.366538
9   AverageIncomeExpendicture    1.111148   13.229198
6                    CardType    0.705167    8.395640
2                 CustomerAge    0.587160    6.990659
5                  CardColour    0.208060    2.477140
0                AcountNumber    0.182778    2.176136
1                         CVV    0.170175    2.026082
4              Marital Status    0.127963    1.523511
3                      Gender    0.097659    1.162715
10      Customer_City_Address    0.093752    1.116202
7                      Domain    0.045035    0.536180
```

Figure 16: Percentage Contribution of individual feature

A graph is plotted which contains the percentage contribution of each feature for better understanding and readability. It has Features on its X-axis and Percentage Contribution calculated from SHAP values on Y-axis. Below is the graph for the same.

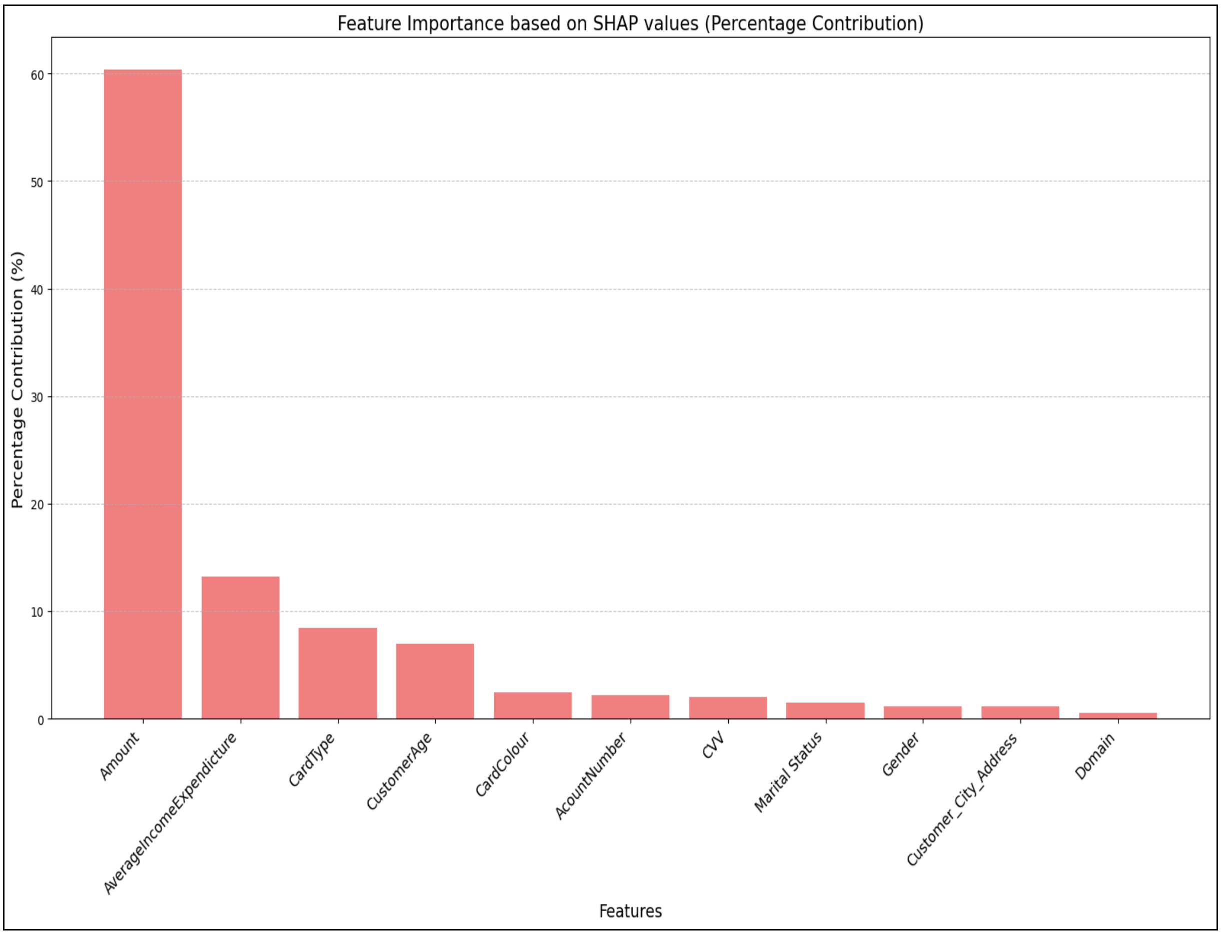


Figure 17: Graph for feature contribution

6.6 Discussion

The results evaluated from the experiments performed, showcase how effective deep learning models like Multilayer Perceptron, Long Short-Term memory, Convolutional Neural Network and machine learning model like K-Nearest Neighbour are in the context of credit card fraud detection and depicts the importance of explainability tools like SHAP for interpretability. This gave rise to some notable observations and analysis. The Multilayer Perceptron (MLP) performed notably well with higher accuracy. Its suitability with balanced dataset and its strength to capture non-linear patterns, both were witnessed in this research.

The Long Short-Term Memory (LSTM) exhibited its capability of identifying complex patterns and temporal relations. The Convolutional Neural Network (CNN) model performed effectively while depicting its ability in finding complicated patterns. Nevertheless, the high false positive rate given by CNN, underlined the need to focus on making improvements in feature engineering. The accuracy showcased by all deep learning algorithms, proves them to be robust and effective in the context of credit card fraud detection.

The results of machine learning algorithm K-Nearest Neighbour (KNN) when integrated with SMOTE, showcased that it strived to gain results even after balancing the data. Through it is a robust, simple algorithm, when compared to results of deep learning, it depicted it was susceptible to distribution of data.

Below is the comparison of deep learning algorithms versus machine learning algorithm:

Model	Accuracy
Multilayer Perceptron	89.65%
Long Short-Term	90.44%
Convolutional Neural Network	89.62%
K-Nearest Neighbour	79.61%

Table 2: Comparison of Results for machine learning and deep learning

For performing feature importance analysis, the XGBoost algorithm was incorporated with Explainable AI method SHapley Additive exPlanations (SHAP). It gave deeper insights on how much each feature contributes. The results revealed that two features that contributed the most were amount and average income expenditure in the context of credit card fraud detection. It critically depicted the feature importance and decision-making process for model predictions. This approach exhibited the use of incorporating XAI for the purpose of performing detection and feature important analysis.

The evaluation metrices represent that the deep learning algorithms showed good quality performance. However, the need to use data balancing techniques and lack of real-life dataset availability, highlight need for improvement.

7 Conclusion and Future Work

This research work explored the field of securing financial system using machine learning and artificial intelligence tools and techniques. The investigation was performed on analysing the efficiency of three deep learning algorithms MLP, LSTM, CNN and machine learning algorithm KNN for credit card fraud detection. Along with that, explainable AI tool SHAP was used to carry out feature importance analysis using XGBoost algorithm. The results and findings revealed that the models performed well through regressive experiments and were successful in achieving nearly 90% accuracy. The results gave rise to some limitations like the reliability on data balancing techniques, the need to use hyperparameter tuning, model being sensitive to distribution of data and high cost of implementing AI tools.

The models and methodology used satisfied the research question and answered on how machine learning and artificial intelligence tools can be used in this context. The discussions demonstrated that the results obtained were in alignment with the previous work showing the significant efficiency of deep learning model over traditional method.

Future work should involve getting some real-world datasets involved and testing the efficacy of these models on those datasets, exploring different balancing techniques that can be suitable for various dataset types and sizes, developing different framework for strong detection of anomalies and patterns to increase accuracy of models. Advanced artificial intelligence techniques like LIME and Anchor can be integrated to investigate the features and their impact in the field of credit card fraud detection.

Acknowledgement

I would like to express sincere appreciation to the National College of Ireland for providing the required guidance. I would like to thank my supervisor Niall Heffernan, for the assistance, support and help provided. It played a huge role in improving the quality of my research thesis.

References

Aakarshachug (2019) *What is LSTM - long short term memory?*, *GeeksforGeeks*. Available at: https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/ (Accessed: December 12, 2024).

AlEmad, M. (2022) *Credit card fraud detection using machine learning*. Rochester Institute of Technology. Available at: https://repository.rit.edu/theses/11318/ (Accessed: December 12,2024)

Awoyemi, J.O., Adetunmbi, A.O. and Oluwadare, S.A. (2017) "Credit card fraud detection using machine learning techniques: A comparative analysis," in *2017 International Conference on Computing Networking and Informatics (ICCNI)*. IEEE, pp. 1–9. Available at: https://ieeexplore.ieee.org/document/8123782 (Accessed: December 12,2024)

Biswas, J. *et al.* (2023) "Interpretable credit card fraud detection using machine learning leveraging SHAP," in *2023 IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT)*. IEEE, pp. 1206–1211. Available at: https://ieeexplore.ieee.org/document/10245439 (Accessed: December 12,2024)

Emma (no date) *Credit Card Fraud Detection DataSet*, *Kaggle.com*. Available at: https://www.kaggle.com/datasets/emmamichael101/credit-card-fraud-detection-dataset (Accessed: December 12, 2024).

Experian, P.L.C. (2023) Over half of consumers feel they're more of a fraud target than a year ago, Experianplc.com. Available at: https://www.experianplc.com/newsroom/press-releases/2023/over-half-of-consumers-feel-they-re-more-of-a-fraud-target-than-a-year-ago (Accessed: December 12,2024)

Gayan, K.K. (2022) "Challenges and complexities in machine learning based credit card fraud detection," *https://doi.org/10.48550/arXiv.2208.10943*. (Accessed: December 12, 2024).

Geeksforgeeks (2017) *K-nearest neighbor(KNN) algorithm*, *GeeksforGeeks*. Available at: https://www.geeksforgeeks.org/k-nearest-neighbours/ (Accessed: December 12, 2024).

Jaiswal, S. (2024) *Multilayer Perceptrons in Machine Learning: A Comprehensive Guide*, *Datacamp.com*. Available at: https://www.datacamp.com/tutorial/multilayer-perceptrons-in-machine-learning (Accessed: December 12, 2024).

Keita, Z. (2023) *An Introduction to Convolutional Neural Networks (CNNs)*, *Datacamp.com*. Available at: https://www.datacamp.com/tutorial/introduction-to-convolutional-neural-networks-cnns (Accessed: December 12, 2024).

Mikesuperman (2023) *SHapley Additive exPlanations or SHAP : What is it ?*, *Data Science Courses | DataScientest*. DataScientest. Available at: https://datascientest.com/en/shap-what-is-it (Accessed: December 12, 2024).

Nguyen, T.T. *et al.* (2020) "Deep learning methods for credit card fraud detection," *arXiv [cs.LG]*. Available at: http://arxiv.org/abs/2012.03754. (Accessed: December 12,2024)

Pillai, T.R. *et al.* (2018) "Credit card fraud detection using deep learning technique," in *2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA)*. IEEE, pp. 1–6. Available at: https://ieeexplore.ieee.org/document/8776797 (Accessed: December 12,2024)

Raufi, B., Finnegan, C. and Longo, L. (2024) "A comparative analysis of SHAP, LIME, ANCHORS, and DICE for interpreting a dense neural network in credit card fraud detection," in *Communications in Computer and Information Science*. Cham: Springer Nature Switzerland, pp. 365–383. Available at : https://link.springer.com/chapter/10.1007/978-3-031-63803-9_20 (Accessed: December 12,2024)

Sarker, A. *et al.* (2024) "Credit card fraud detection using machine learning techniques," *Journal of computer and communications*, 12(06), pp. 1–11. Available at: https://doi.org/10.4236/jcc.2024.126001. (Accessed: December 12,2024)

Sulaiman, S.S., Nadher, I. and Hameed, S.M. (2024) "Credit card fraud detection using improved deep learning models," *Computers, materials & continua*, 78(1), pp. 1049–1069. Available at: https://doi.org/10.32604/cmc.2023.046051. (Accessed: December 12,2024)

Swetha, P. (2024) "Credit card fraud detection with LIME and SHAP," *International journal for research in applied science and engineering technology*, 12(6), pp. 1677–1684. Available at: https://doi.org/10.22214/ijraset.2024.63375. (Accessed: December 12,2024)

"What is the k-nearest neighbors algorithm?" (2024) *Ibm.com*, 28 October. Available at: https://www.ibm.com/topics/knn (Accessed: December 10, 2024).