Credit Card Fraud Detection using Transformer and Attention mechanism

Lecturer: Dr. Abdul Shahid
Foundations of Artificial Intelligence (MSCAI1)
National College of Ireland

Sachleen Singh Chani
ID: 22244778
MSc Artificial Intelligence
National College of Ireland

M Anis Mumtaz Chaudhary
ID: 23215011
MSc Artificial Intelligence
National College of Ireland

10 December 2023

Abstract

As online shopping has become more popular, people mostly use credit cards to buy items from online stores. With this rise in the usage of credit cards, some people try to buy items using fake credit cards and this is credit card fraud. To answer this issue there are many machine learning and deep learning algorithms developed to identify fraudulent and non-fraudulent transactions and prevent fraudulent activities. In this paper, our main aim is to use neural network transformers with attention layers for credit card fraud detection to increase the accuracy of credit card fraudulent and non-fraudulent transactions. The goal is to make online payments more secure and reliable for everyone.

Keywords— Credit card, Fraud Detection, Neural Network, Imbalanced data, SMOTE, Positional Encoding, Transformer architecture, Attention mechanism

1 Introduction

A credit card is a thin plastic or fiber card with your personal information like name signature etc. that allows us to buy things online by constantly adding money to card connected account [Berhane et al., 2023]. A credit card is from the bank which allows you to buy items without using physical money and the bank decides how much money you can spend from a card based on several factors like credit score, credit history and income. Credit cards can be used to borrow money while shopping, but you need to pay that money back within a specific date alongside some extra charges. So, it is like a small loan that needs to be paid back to the bank with some conditions.

With the rise of e-commerce stores, there is a huge increase in the number of online transactions and credit cards are one of the main sources of those transactions. More people are using technology for financial fraud and in 2021, the number of reported cases of credit and debit card fraud went up by 20% [Jessica et al., 2023]. These numbers show that we must take strong action against credit card fraud. One of the most common types of cybercrime is "card-not-present fraud" where your credit card can be misused. Stopping this kind of financial fraud has become more complicated due to the high volume of transactions that happen very quickly so we need smarter ways to identify these frauds quickly and efficiently. A report from the Consumer Financial Protection Bureau in 2019 highlighted that fraud continues to be a significant and expensive problem in the credit card industry, affecting large organisations worldwide. Even though many international transactions are marked as potentially fraudulent, a significant portion of these cases, around 65%, have been identified incorrectly, which is harming merchant sales [Amit Kumar and Kumar, 2022].

The use of credit cards worldwide has become more varied, with more than half of all annual payments made using credit cards. This has resulted in a substantial rise in transactions, totaling $3.6 trillion in 2017 [Salwa Al Balawi, 2021].

There are two types of card purchases, physical and online. The online type makes it easier to do a secret transaction without the cardholder knowing. To catch fraud, we need to investigate customers' past spending patterns and use machine learning algorithms to identify the status of transactions, and whether they are fraudulent or non-fraudulent. Credit card fraud detection using neural networks is the main objective of this project. We will use a transformer neural network model with multiple attention layers. We used a standard credit card detection dataset which was highly imbalanced, so we used some oversampling techniques to manage the imbalanced issue with the dataset. We will discuss the whole methodology and results in detail in the report with future work.

2 Related work

Credit cards allow card owners to buy goods or get services and pay at a specific time like the next billing cycle, this increases the chances of fraudulent transactions. These fraudulent transactions are the main concern for consumers and the financial sector. Nowadays sophisticated methods are used for credit card fraud, which requires up-to-date detection strategies and needs a strong system to prevent credit card fraud transactions.

[Jessica et al., 2023] use the machine learning method of decision tree and ensemble learning like stacking to get good results with an unbalanced dataset with only up to 5 % fraudulent transactions and use SMOTE (Synthetic Minority Over-Sampling Technique) to balance the dataset. [Tressa et al., 2023] use a random forest algorithm for credit card fraud detection by using the same SMOTE oversampling technique to cater for highly unbalanced dataset issues.

[Amit Kumar and Kumar, 2022] also work on dataset resampling using different oversampling and undersampling methods and focus on the importance and data preprocessing of some common machine learning methods. Using more enhanced methods instead of traditional rule-based methods for credit card fraud detection is very important nowadays to cater for sophisticated credit card fraudulent techniques. [Jain et al., 2022] in their analysis of various models like artificial neural networks, decision tree, gradient boosting, and logistic regression, evaluate these model performance using ROC (Receiver Operating Characteristics)

[Tanouz et al., 2021] also explored the shift from traditional methods to more complex learning methods for more efficient ways of credit card fraud detection. [Leevy et al., 2023] also investigate the issue of credit card fraud for online credit card transactions, by doing deep analysis of Xgboost, Neural Networks, random forest, and decision tree. The main aim of their study is to identify an effective algorithm and find out whether Xgboost, decision tree and random forest are better than other algorithms. Their study also highlights the significance of undersampling and oversampling.

In another study, [Salwa Al Balawi, 2021] use two commonly used deep learning techniques, CNN (Convolutional Neural Network) and ANN (Artificial Neural Network). They concluded that CNN (Convolutional Neural Network) without the pooling layer performed well and resulted in higher accuracy in classifying fraudulent and non-fraudulent transactions.

[Dornadula and Geetha, 2019] use a multi-step approach. They cluster the data into different groups based on the amount of transactions using range partitioning and apply different machine-learning algorithms to each group separately. The results of their paper indicate better classification results after applying SMOTE again highlighting the importance of a balanced dataset. They also suggest MCC (Matthew's Correlation Coefficient) is the better metric for measuring how well the model performed unbalanced dataset.

[Berhane et al., 2023], created a hybrid CNN-SVM model to check fraudulent and non-fraudulent credit card transactions. They use SMOTE for the oversampling of unbalanced dataset. They replace final output layer of CNN with SVM classifier, and their results shows that the CNN-SVM model performed better than fully connected convolutional neural network in terms of recall, precision, F1-score, and predicting whether the transaction is fraudulent or non-fraudulent.

3 Methodology

In developing a robust model in detecting credit card fraud we proposed a model that is based on the transformer architecture and attention mechanism. This way we dispense the use of RNN or CNN, this would simplify the model and help with the training time and resources [Vaswani et al., 2017]. At the core of the methodology is the innovative application of attention mechanisms and transformer architectures. MultiHead Attention layers are strategically deployed to capture intricate relationships within the data, enabling the model to focus on relevant features critical for fraud detection. To further improve this, feature selection is implemented. The incorporation of positional encoding provides essential information about the sequential order of transactions, improving the models performance and adaptability for unseen data.

3.1 Data Pre-processing

Pre-processing is a vital step in cleaning and making the data ready to be used to train a model. In case the dataset is unbalanced or have missing values, this would negatively impact the trained model, could potentially develop bias in the predictions [Delamaire et al., 2009].

In this paper the particular dataset we are using to train the model[1].

Since credit card information is highly sensitive personal data, due to security and ethical concerns the data in transformed using Principal Component Analysis (PCA) to anonymise the personal information from the dataset while also preserving the key attributes of the data which will help us in a better performance from the model. There are two features that have been kept the same, i.e. Time and Amount. Apart from these the last column of the data is the target class, "Class" which gives the boolean value, 1 for a fradulent transaction and 0 for a genuine transaction. The PCA transformation gave us features v0 to v28.

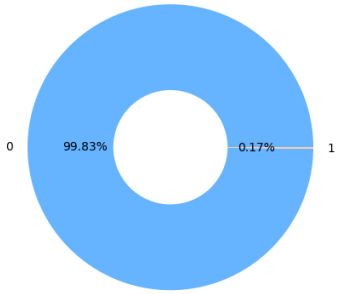

Figure 1: Fraudulent and Genuine data count.

Further inspection of the dataset was done to identify any outliers or missing values. There are no missing values or outliers in the data.

Moving on the next step, we observe that this dataset is highly unbalanced, with 99.83% non-fradulent transactions and only 0.17% fraud transactions. This would make it particularly hard to train the model without bias towards genuine transactions. To resolve this issue we use an oversampling technique called Synthetic Minority Over-sampling Technique (SMOTE). This balances the frequency on occurrence of the minority class by strategically generating synthetic data to increase the count of the minority class to match with the majority class. The way SMOTE generates synthetic data is to the k-NN (K nearest neighbore) is used to find the nearest neighbore and uses a statistical formula to place the generated data between itself and the nearest neighbore.

Correlation matrix helps us visualize the dependence of each feature to every other feature in the dataset. By observing the Pearson's Correlation Matrix for all the features from the data, Figure 2 we see that there are features that do not have any correlation with the "Class" target feature. To reduce the computation time, we have used feature selection to reduce the features. From this matrix and the Swarm intelligence plots done by [Benchaji et al., 2021], we conclude that V5, V6, V7, V8, V9, V13, V15, V16, V18, V19, V20,

[1]https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud/

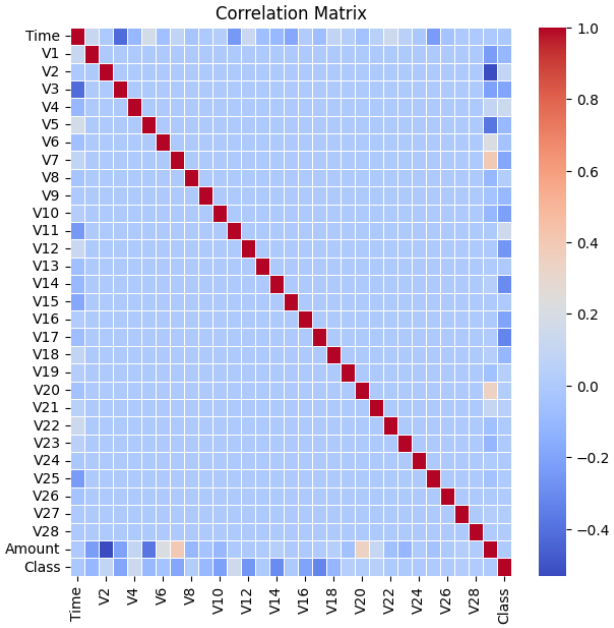

Figure 2: Pearson's Correlation Matrix of the features.

V21, V22, V23, V24, V25, V26, V27, V28, amount do not influence the target class and would not be helpful in training the model. Following this, the data was standardized, this a crucial step for deep learning. All the features are scaled to have the standard deviation of 1 and the mean of the data at 0. Finally, the dataset is split into three groups namely, training data, testing data and validation data. The split is performed to have 80% training data, 10% for validation and the remaining 10% for testing. The helps to check the performance of the model on new, unseen data.

3.2 Model Architecture

The transformer model was first proposed by [Vaswani et al., 2017] in their paper "Attention is all you need". The first layer in the model is the input layer, which is designed to accommodate the shape of (9,1). Since after feature selection, we are left with 9 most relevant features which are, V1, V2, V3, V4, V10, V11, V12, V14 and V17. A distinct feature used in this proposed model is the use of positional encoding through an embedding layer. This helps to enhance the model to capture sequential data.

Following this, there are two attention layers used. This attention layer plays a critical role in feature extraction. The first attention layer uses a Multi-head self attention mechanism with four heads. Dropout regularisation is used to prevent overfitting the model on the data. And the use of layer normalization ensures stable training of the model. Similarly, the final attention uses the same parameters but with the Multi-Head employing two heads.

Additionally, two dense layers are employed with 128 and 64 perceptrons respectively along with dropout layer being used to prevent overfitting.

Finally the last layer contains only a single neuron with the sigmoid activation function, which facilitates in binary classification.

Which the model architecture done, the model is exposed to the training data which has now been oversampled to balance the two classes, this would give us better unbiased performance of the model. The model is compiled using the Adam optimizer and binary crossentropy loss function.

4 Evaluation & Results

In this report we have used confusion matrix and classification report for the evaluation of the model. Confusion matrix provides a comprehensive performance evaluation for the classified problems. Confusion matrix consists of four key matrices: true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN).

True positives (TP) explains the number of times model correctly predicts fraudulent transaction. In other words transaction that are fraudulent and model correctly identified that transaction. True negatives (TN) explains the number of times model correctly predicts non fraudulent transaction. Transaction is non fraudulent and model predict that transaction correctly. False positives (FP) tells us the number of times when model predicts fraudulent transaction but actually transaction is non fraudulent, this is also known as false alarm. False negative (FN) tells us number of instances where model predicts non-fraudulent transaction but transaction is fraudulent this is also called miss.

From these values, different performance metrics can be calculated like accuracy, precision, specificity and recall (sensitivity). These are important part of classification report and can be calculated as follow:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$F1Score = \frac{TP}{TP + FN} \tag{2}$$

$$Specificity = \frac{TN}{FP + TN} \tag{3}$$

$$Recall/sensitivity = \frac{TP}{TP + FP} \tag{4}$$

From the Table 1 we can observe that the model has the accuracy of 94.63%, precision 97.57% and recall 91.37%. The link for the implementation code in python can be referenced from the github repository[2].

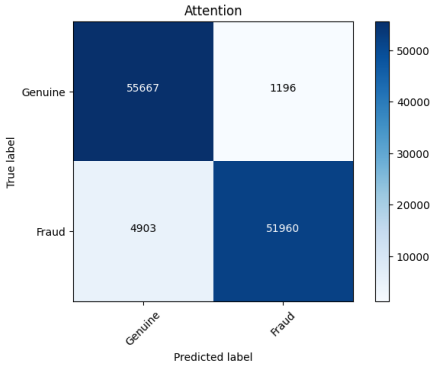

Figure 3: Confusion Matrix.

From the Figure 3 confusion matrix also confirms that the model is able to detect fraudulent transactions with 97.57%. The recall is notably high for both the classes, this indicates that the model is is able to capture a significant portion of true positives. The F1-scores for both classes are also impressive, indicating a balanced trade-off between precision and recall.

Table 2 compares difference existing model's performance in terms of the Accuracy, Precision and Recall. As we can see the porposed model performs

[2]https://github.com/dead96pool/FOAI-Project

Class	Precision	Recall	F1 Score	Support
Class 0	0.92	0.98	0.95	56863
Class 1	0.98	0.91	0.94	56863
Accuracy			0.95	113726
Macro avg	0.95	0.95	0.95	113726
Weighted avg	0.95	0.95	0.95	113726

Table 1: Classification Report

Table 2: Performance Comparison of Different Algorithms

Algorithm	Accuracy	Precision	Recall
GRU (2020) [Forough and Momtazi, 2021]	–	0.8626	0.7208
LSTM (2020) [Forough and Momtazi, 2021]	–	0.8575	0.7408
SVM (2021) [RB and KR, 2021]	0.9349	0.9743	0.8976
KNN (2021) [RB and KR, 2021]	0.9982	0.7142	0.0393
ANN (2021) [RB and KR, 2021]	0.9992	0.8115	0.7619
LSTM-attention [Benchaji et al., 2021]	0.9672	0.9885	0.9191
(Our-proposed Model) Transformer-Attention	**0.9463**	**0.9775**	**0.9137**

5 Conclusion

For this paper, we started by proposing a question to develop a model using the transformer architecture and self-attention mechanism to be applied for the detection of credit card fraud. With the pre-processing of the unbalanced fraud data using SMOTE and generating synthetic datapoints to balanced the dataset, we have trained the model without it being biased towards the minority "fraud" class. The model performs well with above 90% for both accuracy and recall, which means the model can successfully extract patterns from the data and performs well at classifying the data as fraudulent. In summary, the model exhibits strong predictive capabilities and generalizes well to unseen data.

As future work, we would like to use ensemble learning along with the attention mechanism.

References

[Amit Kumar and Kumar, 2022] Amit Kumar, Anant Jain, M. A. and Kumar, N. (2022). Credit card fraud detection using machine learning. *Journal of Pharmaceutical Negative Results*, pages 5717–5723.

[Benchaji et al., 2021] Benchaji, I., Douzi, S., El Ouahidi, B., and Jaafari, J. (2021). Enhanced credit card fraud detection based on attention mechanism and lstm deep model. *Journal of Big Data*, 8(1):151.

[Berhane et al., 2023] Berhane, T., Melese, T., Walelign, A., and Mohammed, A. (2023). A hybrid convolutional neural network and support vector machine-based credit card fraud detection model. *Mathematical Problems in Engineering*, 2023:Article ID 8134627, 10 pages.

[Delamaire et al., 2009] Delamaire, L., Abdou, H., and Pointon, J. (2009). Credit card fraud and detection techniques: A review. *Banks and Bank Systems*, 4.

[Dornadula and Geetha, 2019] Dornadula, V. N. and Geetha, S. (2019). Credit card fraud detection using machine learning algorithms. *Procedia Computer Science*, 165:631–641. 2nd International Conference on Recent Trends in Advanced Computing ICRTAC -DISRUP - TIV INNOVATION , 2019 November 11-12, 2019.

[Forough and Momtazi, 2021] Forough, J. and Momtazi, S. (2021). Ensemble of deep sequential models for credit card fraud detection. *Applied Soft Computing*, 99:106883.

[Jain et al., 2022] Jain, N., Chaudhary, A., and Kumar, A. (2022). Credit card fraud detection using machine learning techniques. In *2022 11th International Conference on System Modeling and Advancement in Research Trends (SMART)*, pages 1451–1455.

[Jessica et al., 2023] Jessica, A., Raj, F. V., and Sankaran, J. (2023). Credit card fraud detection using machine learning techniques. In *2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN)*, pages 1–6.

[Leevy et al., 2023] Leevy, J., Hancock, J., and Khoshgoftaar, T. (2023). Comparative analysis of binary and one-class classification techniques for credit card fraud data. *J Big Data*, 10:118.

[RB and KR, 2021] RB, A. and KR, S. K. (2021). Credit card fraud detection using artificial neural network. *Global Transitions Proceedings*, 2(1):35–41. 1st International Conference on Advances in Information, Computing and Trends in Data Engineering (AICDE - 2020).

[Salwa Al Balawi, 2021] Salwa Al Balawi, N. A. (2021). Credit-card fraud detection system using neural networks. *The International Arab Journal of Information Technology (IAJIT)*, 20(02):234 – 241.

[Tanouz et al., 2021] Tanouz, D., Subramanian, R. R., Eswar, D., Reddy, G. V. P., Kumar, A. R., and Praneeth, C. V. N. M. (2021). Credit card fraud detection using machine learning. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pages 967–972.

[Tressa et al., 2023] Tressa, N., Asha, V., M, G., Padanoor, S., Tabassum, R., Dharmesh, D. V., and Saju, B. (2023). Credit card fraud detection using machine learning. In *2023 3rd Asian Conference on Innovation in Technology (ASIANCON)*, pages 1–6.

[Vaswani et al., 2017] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Łukasz Kaiser, and Polosukhin, I. (2017). Attention is all you need. *CoRR*, abs/1706.03762.