International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 10 | Oct 2024 www.irjet.net p-ISSN: 2395-0072

Analysis of Different Machine Learning Models for Credit Card Fraud Detection

Harsh Mehta

School of Computer Science, Presidency University, (UGC), Bangalore, Karnataka, India

---

Abstract*— The increase in number of online transactions has led to a significant amount of credit card fraud over the past decade. Unauthorized use of one's credit card information by stealing the information through dark web or scam calls, poses a major risk to both customer and businesses, particularly in e-commerce setting. This paper presents a comparative analysis of multiple machine learning models for credit card fraud detection, including logistic regression, isolation forest, K – mean clustering, and convolutional neural networks. With a highly unbalanced dataset we aim to evaluate these models' performance in differentiating between genuine and fraudulent transactions based on features such as transaction history, user details, and merchant information. Our experiment results will help provide insights into effectiveness of each model for finding patterns to distinguish between real and fake that can be applied to real world data. This research contributes to the field of financial security by offering guidance on model selection for credit card fraud detection and related applications. View this project here.

Keywords – Credit card fraud, machine learning, logistic regression, isolation forest, k-mean clustering, convolutional neural network, financial security.

I. INTRODUCTION

The rapid growth of online financial transactional methods are seen in the recent times and adopted widely because it's easy, reliable, and faster in multiple aspects compared to traditional payment methods. Among this online credit card fraud has been a concerning issue that challenges the security and integrity of information that can be circulated through internet. This paper will help future peers in understanding and choosing models according to their build requirements.

A. Background on credit card frauds

Credit card frauds have become a significant threat in the coming digital age, possessing an enormous financial risk to individual, businesses and the global financial system. As e-commerce and digital transactions grow with time so does the fraudulent activities. Credit card fraud generally occurs when unauthorized individual gain access to card information through various means like data breach, skimming devices, or phishing attacks. These scammers then use the stolen information to make unauthorized purchases or even cash withdrawals, often resulting in financial losses for cardholder and merchants. The problem goes beyond the loss of money as it affects the trust in digital payment systems, and potentially leads to long term economic instability if left unchecked.

B. Current challenges in detection

The detection and prevention of credit card fraud presents several challenges for developers and organizations trying to deal with it. One of the primary obstacles is working with high dynamic nature of fraudulent activities, with scammers always changing and adapting new methods to cheat the detection system. This makes it necessary to keep evolving our detection methods to stay ahead of emerging threats and avoid before it even takes place. The number of genuine transactions vastly outnumber fraudulent one, this results in having a dataset where fraud transactions represent very minute number of the whole dataset. This imbalance creates biased models that prioritize the majority class, which might miss critical fraud transaction. Additionally, the sensitive nature of financial data often limits access to real world datasets, making it very difficult for researchers and developers to build and test a model.

C. Our approach and its significance

Our approach to address this issue involves a performance analysis of multiple machine learning model applied to credit card fraud detection. Using a dataset from Kaggle named "Credit Card Transactions Fraud Detection Dataset" (Brandon, 2022) which mimics real world transaction pattern while preserving user's privacy, we implemented a unique methodology where we evaluate the effectiveness of different models like: regression model, decision tree model, clustering model and convolutional neural network (CNN). We compare the performance of these models across multiple metrics, such as classification report, confusion metrics, AUC-ROC scores and feature importance analysis, through this we aim to find relative strengths and weaknesses in the context of credit card fraud detection. This performance evaluation contributes to providing help in ongoing efforts for improvement in fraud detection systems and offers valuable guidance to future peers in selecting and implementing appropriate model according to needs for similar security applications.

II. LITERATURE REVIEW

A. Credit Card Fraud Detection using Machine Learning and Data Science, DOI: ISSN: 2278-0181, (S P Maniraj, 2019)

Fraud detection in credit card transactions has been a subject of extensive research due to its significant financial implications. Previous studies have explored various data mining applications and machine learning techniques for automated fraud detection. Supervised and unsupervised learning methods have been applied to this domain, with varying degrees of success. Some researchers have utilized outlier mining and distance sum algorithms to predict fraudulent transactions in emulated credit card transaction datasets. While these methods have shown promise in certain areas, they have not provided a consistent and permanent solution to the fraud detection problem.

More recent approaches have incorporated advanced techniques such as hybrid data mining/complex network classification algorithms. These methods have demonstrated effectiveness in detecting illegal instances in real card transaction datasets, particularly for medium-sized online transactions. Efforts have also been made to improve the alert feedback interaction in fraudulent transaction detection systems. Artificial Genetic Algorithms have been explored as a novel approach, showing accuracy in identifying fraudulent transactions while minimizing false alerts. However, these methods often face challenges related to classification problems with variable misclassification costs. The ongoing research in this field continues to seek more robust and adaptable solutions to address the evolving nature of credit card fraud.

B. A Research Paper on Credit Card Fraud Detection, (BORA MEHAR SRI SATYA TEJA, 2022)

The paper explores various techniques used in credit card fraud detection, including outlier detection, unsupervised outlier detection, peer group analysis, and breakpoint analysis. Outlier detection identifies abnormal transactions that deviate from a user's typical behaviour, but it may misclassify legitimate unusual transactions. Unsupervised outlier detection focuses on understanding customer transaction patterns without predicting specific outcomes. Peer group analysis compares entities with similar characteristics to identify anomalies. Breakpoint analysis examines structural changes in data to detect anomalies.

The authors note that while supervised learning methods are commonly used in fraud detection, they may fail in certain cases. The paper highlights the challenge of class imbalance in fraud detection datasets, where genuine transactions significantly outnumber fraudulent ones. This imbalance can lead to difficulties in accurately identifying fraudulent activities. The researchers also discuss the concept of "concept drift," where transaction patterns change over time, further complicating the fraud detection process. To address these challenges, the paper proposes using machine learning algorithms such as Decision Trees and Random Forests, along with techniques like oversampling to mitigate class imbalance issues.

C. A machine learning based credit card fraud detection using the GA algorithm for feature selection, DOI: 10.1186/s40537-022-00573-8, (Emmanuel Ileberi, 2022)

The literature survey on credit card fraud detection reveals a growing interest in machine learning techniques to address this critical issue in financial security. Researchers have explored various approaches, including supervised and unsupervised learning methods, to improve the accuracy and efficiency of fraud detection systems. Several studies have focused on the application of traditional machine learning algorithms such as Support Vector Machines (SVM), Decision Trees, and Neural Networks. These methods have shown promising results in identifying fraudulent transactions, although they often face challenges related to imbalanced datasets and the dynamic nature of fraud patterns.

Recent research has increasingly turned towards ensemble methods and hybrid approaches to enhance fraud detection capabilities. Random Forest and Gradient Boosting algorithms have gained popularity due to their ability to handle complex, high-dimensional data and their robustness against overfitting. Additionally, some studies have explored the potential of deep learning techniques, including Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, to capture intricate patterns in transaction data. These advanced methods have demonstrated improved performance in detecting subtle fraud patterns that may be missed by traditional approaches.

A significant trend in the literature is the focus on feature engineering and selection techniques to improve model performance. Researchers have employed various methods, including Principal Component Analysis (PCA), Genetic Algorithms, and domain-specific feature extraction, to identify the most relevant attributes for fraud detection. Moreover, there is a growing emphasis on developing real-time fraud detection systems that can adapt to evolving fraud patterns and provide timely alerts. Despite these advancements, the literature highlights ongoing challenges in credit card fraud detection, including the need for more representative and up-to-date datasets, addressing class imbalance issues, and developing interpretable models that can provide insights into fraudulent behaviour patterns.

D. Review of Machine Learning Approach on Credit Card Fraud Detection, DOI: 10.1007/s44230-022-00004-0, (Rejwan Bin Sulaiman, 2022)

This review examines various machine learning techniques for credit card fraud detection (CCFD), focusing on their effectiveness, limitations, and privacy considerations. The paper discusses several algorithms, including Random Forest (RF), Artificial Neural Networks (ANN), Support Vector Machines (SVM), and K-Nearest Neighbors (KNN). Each method demonstrates unique strengths and weaknesses in handling CCFD tasks. For instance, Random Forest shows promise in handling large datasets but may be slower in real-time scenarios. ANN, particularly when used in unsupervised learning, demonstrates high accuracy and fault tolerance, making it a strong contender for CCFD applications. SVM performs well with smaller feature sets but struggles with larger volumes of data, while KNN offers high accuracy and efficiency but faces challenges with memory usage and performance degradation on extensive datasets.

The review highlights a critical challenge in CCFD: balancing effective fraud detection with data privacy and confidentiality. Traditional centralized approaches to fraud detection face limitations due to data sharing restrictions imposed by regulations like GDPR. Even anonymized datasets stored locally on servers' risk being reverse-engineered, potentially compromising user privacy. This privacy concern is a recurring theme across various machine learning approaches discussed in the paper, emphasizing the need for more secure and privacy-preserving methods in CCFD.

To address these challenges, the paper proposes a hybrid approach combining Federated Learning (FL) with Artificial Neural Networks. This innovative model aims to train data locally on edge devices, sharing only the trained model among participating institutions. This approach potentially enhances fraud detection accuracy while maintaining strict privacy standards. By allowing banks and financial centres to collaborate without directly sharing sensitive customer data, the proposed method offers a promising solution to the privacy-accuracy trade-off in CCFD. The authors suggest that this hybrid model could significantly improve fraud detection capabilities while ensuring compliance with data protection regulations, marking a potential advancement in the field of credit card fraud detection.

E. A Review Paper on Feature Selection in Credit Card Fraud Detection, (Surbhi Bansal, 2024)

Credit card fraud detection has been a subject of extensive research due to its significant economic impact. Researchers have compared the performance of various machine learning techniques such as Support Vector Machines, Random Forests, and Logistic Regression in detecting credit card fraud, highlighting the importance of feature selection in improving model accuracy. The challenge of class imbalance in fraud detection has also been addressed, with proposed methods combining techniques like SMOTE and random under sampling. These works have emphasized the need for adaptive learning techniques in handling evolving fraud patterns.

Feature selection in fraud detection has seen increasing attention, with researchers exploring various approaches. The effectiveness of transaction aggregation for creating behavioural features has been demonstrated, significantly improving fraud detection rates. Scalable real-time fraud detection systems using feature engineering and hybrid methods have been proposed, showcasing the importance of both domain expertise and machine learning in feature creation. More recently, Swarm Intelligence techniques have been applied for feature selection in fraud detection, demonstrating improved model performance and interpretability compared to traditional methods.

F. Credit card fraud detection using machine learning, (Mr. Thirunavukkarasu.M, 2021)

Credit card fraud detection has been an active area of research due to its significant economic impact. Previous studies have compared the performance of various machine learning techniques such as Support Vector Machines, Random Forests, and Logistic Regression for detecting credit card fraud, with Random Forests often outperforming other methods. Research has also demonstrated the effectiveness of transaction aggregation combined with Random Forests for fraud detection, showing improved results over single transaction analysis.

In recent years, machine learning approaches have gained prominence in fraud detection. Researchers have addressed the challenge of class imbalance in credit card fraud detection datasets, proposing methods that combine under sampling with different algorithms to improve overall performance. Comprehensive reviews of intelligent fraud detection techniques have highlighted the potential of ensemble methods like Random Forests in handling complex, high-dimensional data typical in financial transactions.

The application of deep learning to credit card fraud detection has also emerged as a promising direction. Studies have explored the use of Long Short-Term Memory (LSTM) networks for sequence classification in credit card fraud detection, showing that incorporating transaction sequences can enhance detection accuracy compared to traditional methods. However, while deep learning models can offer improved performance, they often lack the interpretability of simpler models like Random Forests, which remains an important consideration in the financial industry.

III. OBJECTIVES

A. Understanding various ML models with respect to credit card fraud detection

We aim to explore and analyze different machine learning models, specifically logistic regression, isolation forest, k-means clustering, and convolutional neural network, with respect to credit card fraud detection. We will understand the principle of each model and how are they used to identify fraud transactions.

B. Performance analysis of ML models

We will evaluate each model performance in detecting credit card fraud. This includes assessing their ability to correctly identify fraudulent transactions while minimizing false positives. This analysis is based on factors like accuracy, precision, and recall to provide an overall view of each models effectiveness.

C. Assessing the effectiveness of each model using different metrics

To ensure our model is performing well we will use various performance metrics beyond basic accuracy. This includes confusion matrices, AUC-ROC curves and F1 scores, by using these factors we will aim to find out more about the strengths and weaknesses of each detection models.

D. Provide recommendation for the ML model

Based on our analysis we will provide insights and recommendation on which model perform best for credit card fraud detection. These recommendations will consider factors such as model performance, computational requirements and ease of implementation providing guidance to future peers.

E. Understanding features that affect the model development

We will understand the importance of different features in the dataset and their impact on the performance of each model. This involves conducting feature importance analysis to identify which transaction characteristics are most crucial in determining whether a transaction is fraudulent or legitimate.

IV. PROPOSED METHODOLOGY

To develop this credit card fraud detection project using various machine learning models we have taken the following steps that helps us understand this project from scratch:

A. System overview

Our credit card fraud detection system follows the given workflow:

- Data Ingestion: Raw data that's downloaded from Kaggle is fed to the system without any preprocessing or scaling.
- Preprocessing: The data undergoes cleaning through various methods and techniques to modelling can be done on the data that makes sense.
- Data Scaling: The numerical features are normalized in the data so the model can ensure to provide consistent outputs.
- Applying Pretrained Models: We use four different pre trained machine learning models on the preprocessed data.
- Classification Report and Metrics: Performance metrics and reports are produced for each model.

B. Dataset Description

The dataset used in this project is downloaded from Kaggle the dataset originally belongs to Brandon Harris and generated using a simulator (Brandon, 2022). This data consists of legitimate and fraud transactions details from Jan 2019 till Dec 2020, and consist of card details of over 1000 customers and 800 merchants. This data generated creates easy to use fraud transaction dataset which is a representation of real-life transactions it contains two files named "fraudTrain" and "fraudTest" both of them combining contains over 1.5 million various transactions.

C. Data Preprocessing

In our preprocessing pipeline we:

- Convert date to datetime: The time features is converted to datetime for better interpretability.
- Extracting features from datetime: We extract additional features like hour, day and month to capture temporal patterns.
- Dropping unnecessary columns: Removing redundant and non-informative columns are always helpful for better model interpretability.
- Scaling the data: Numerical features are scaled using standard scalers to ensure all features contribute for model development.

D. Model Description

We are using four different types of models and they all work and train themselves using the data differently:

- Logistic Regression Model: In statics the logistic regression model helps in estimating the probability of an event taking place provided on the provided dataset, and helps analyze the relationship between factors. This would fit well as the model can mark the fake detection as odds and log them for future predictions.

- Isolation forest model: This algorithm is used for anomaly detection in the data with the help of binary trees. This algorithm is ideal for credit card fraud detection as it has a low time complexity and memory use that works well with huge amount of data too.
- K-Mean clustering: This is an unsupervised machine learning algorithm, which helps group unlabeled data into multiple groups or clusters. It creates a centroid in the data and based on the distance it classifies or categorize the data. This model will theoretically fit well as the model and create two cluster of real and fake and predict using their centroids.
- Convolutional Neural Network: CNN comes under deep learning and is a type of neural network that usually creates 3 layers: input, hidden, and output. It will help in Local Pattern Detection, and Feature Extraction and generally works well with large volume of data.

E. Training Process

Even though the preprocessing method for all the four models is the same but each one of them will undergo a different training process:

- Logistic Regression and Isolation Forest: They will be directly trained on the pre-processed data with default hyperparameter.
- K-Means: Here the number of clusters would be determined using elbow method before training.
- CNN: The network architecture would be modified according to the tabular data with multiple convolutional layers. The training would go on for 10 rounds with early stopping to prevent overfitting.

F. Evaluation and Analysis

We will evaluate the model using metrics such as:

- Accuracy: Overall correctness of the model.
- Precision and Recall: To access model performance on minority class.
- F1-score: The harmonic mean of precision and recall.
- ROC-AUC: To check models' ability to distinguish between different classes.
- Confusion Matrix: To visualize model performance across all outcomes.
- Feature Importance Analysis: To check which feature in the dataset is most important for fraud detection.

Finally, we will note down all the results and check how each model performs in various metrics and also note down the time and computational power that was required for each model to give the final predictions.

TABLE I

Hardware Requirements	**Software Requirements**
Graphic Card (Recommended): - NVIDIA GPU with CUDA support (Optional but recommended). **Compute Resources:** - 8 core CPU. - Adequate RAM (8GB or above) **Storage:** - SSD with at least 20GB free space **Network Infrastructure:** - High-speed Internet Connection.	**Operating System:** - Windows 10/11, macOS, or Linux (Ubuntu 18.04 or later recommended) **CUDA Toolkit:** - Version compatible with PyTorch and GPU (Optional: works only with graphic cards) **Necessary Library:** - numpy - scikit-learn - matplotlib - seaborn - pandas **Development Tools:** - Anaconda - Jupyter Notebook

Requirements for deepfake detection model

These are general requirements

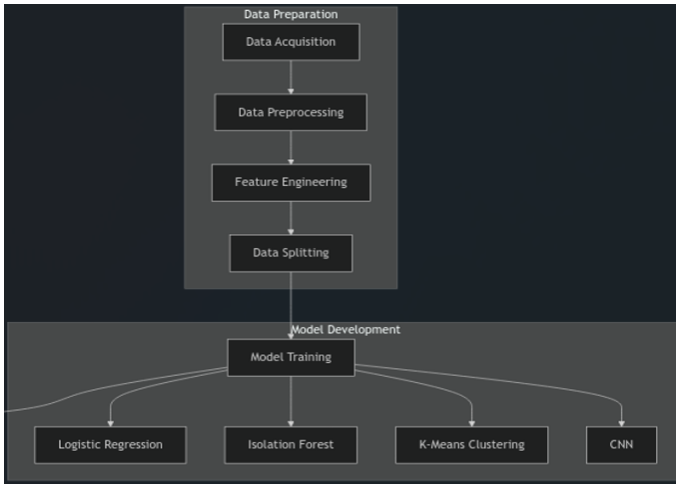

Fig. 1.1 Workflow diagram of credit card fraud detection model.

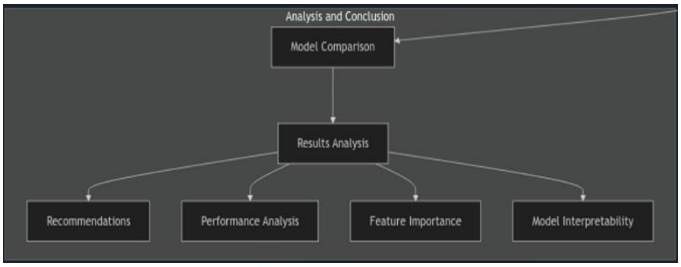

Fig. 1.2 Workflow diagram of credit card fraud detection model.

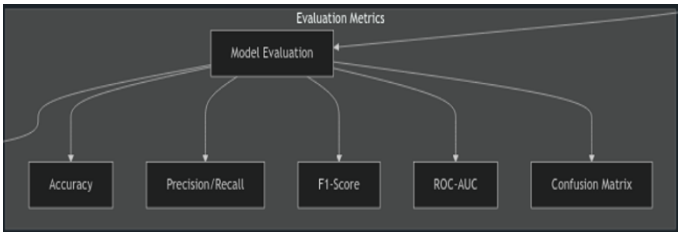

Fig. 1.3 Workflow diagram of credit card fraud detection model

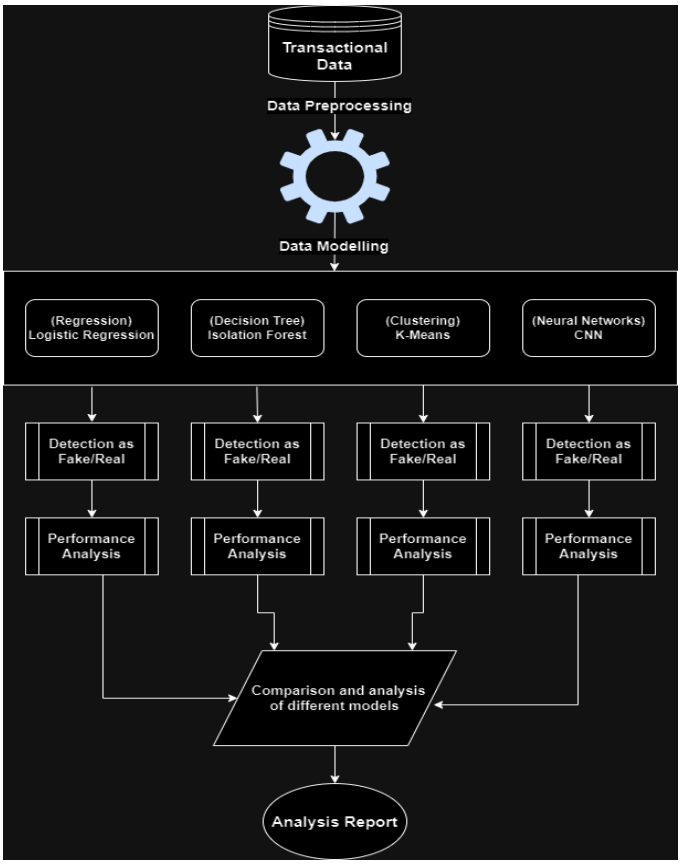

Fig. 2. Architecture of credit card fraud detection model.

V. RESULTS AND DISCUSSION

Evaluation of each model is necessary to understand and rank the models accordingly. As discussed earlier we will evaluate all the four model on different metrics:

- Training and Testing accuracy: It's the proportion of correct prediction by total number of cases. It's used check training vs testing set to assess overfitting.
- Classification Report: It's a summary of the key classification's metrics including precisions, recall score, and F1-score for each class. It helps us provide a comprehensive view of model's performance
- AUC-ROC and Average Precision Score: The AUC-ROC measures the model's ability to differentiate between classes among different threshold. The average precision scores summarize the precision-recall curve as the weighted mean of precisions. achieved at each threshold
- Confusion Matrix: It's a table showcasing the number of correct and incorrect predictions made by the model. This helps us provide a breakdown of model's performance and understand error types.
- Feature Importance: This is the measure of the features which contributes to the prediction of the model. This helps us provide transaction characteristics and that provides insights for feature engineering and model interpretation.

A. Logistic Regression Model Performance

```
Training Accuracy: 0.8167
Testing Accuracy: 0.8799
```

Fig. 3.1. Training and Testing scores using Logistic regression model.

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

           0       1.00      0.88      0.94    368549
           1       0.03      0.76      0.06      1930

    accuracy                           0.88    370479
   macro avg       0.52      0.82      0.50    370479
weighted avg       0.99      0.88      0.93    370479
```

Fig. 3.2. Classification report of Logistic regression model.

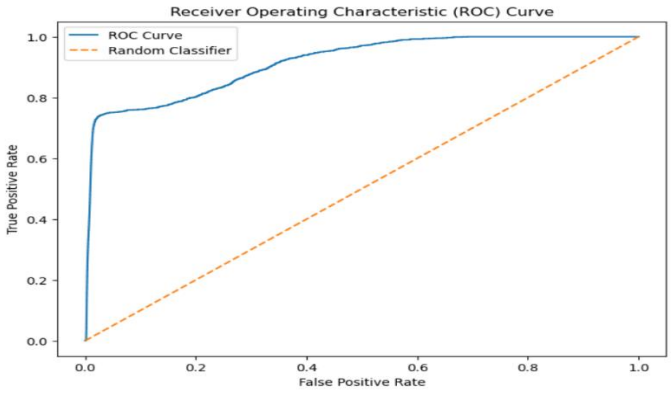

Fig. 3.3. ROC Curve of Logistic regression model.

```
AUC-ROC Score: 0.9160
Average Precision Score: 0.1553
```

Fig. 3.4. AUC-ROC and Average Precision score of Logistic regression model.

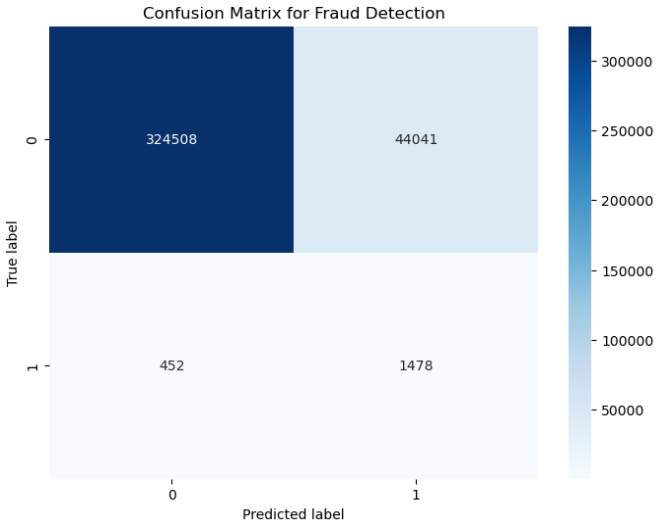

Fig. 3.5. Confusion Matrix of Logistic regression model.

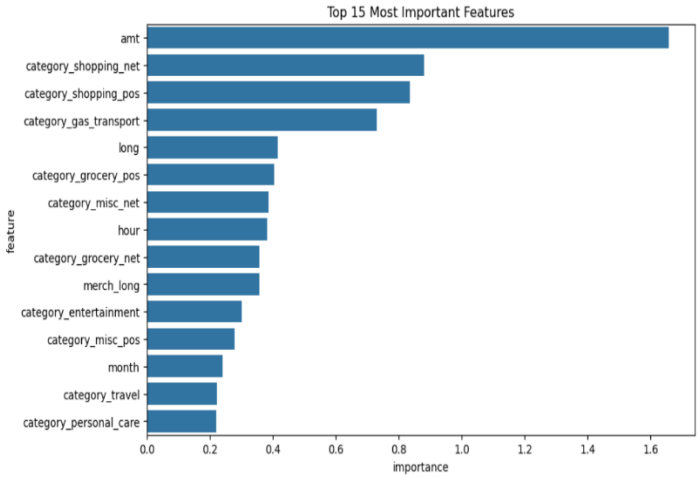

Fig. 3.6. Feature Importance Analysis of Logistic regression model.

B. Isolation Forest Model Performance

```
Training Accuracy: 0.9854944376964159
Testing Accuracy: 0.9843949909936497
```

Fig. 4.1. Training and Testing scores using Isolation Forest model.

```
Classification Report (Testing Set):
              precision    recall  f1-score   support

           0       1.00      0.99      0.99    553574
           1       0.03      0.10      0.05      2145

    accuracy                           0.98    555719
   macro avg       0.51      0.55      0.52    555719
weighted avg       0.99      0.98      0.99    555719
```

Fig. 4.2. Classification report of Isolation Forest model.

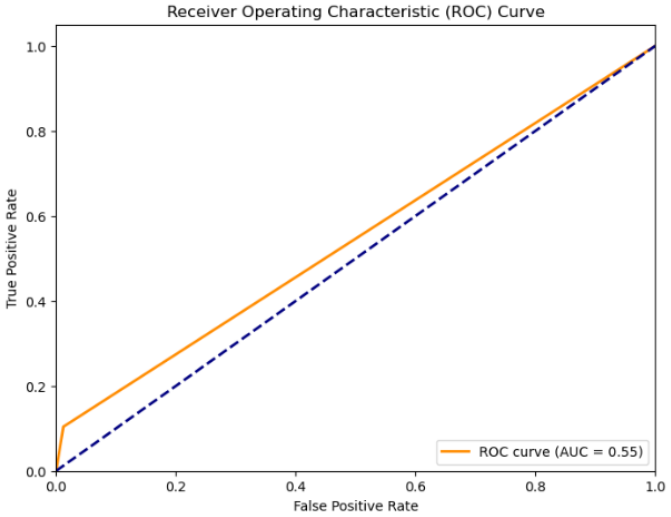

Fig. 4.3. ROC Curve of Isolation Forest model.

```
AUC-ROC Score: 0.5459
Average Precision Score: 0.0068
```

Fig. 4.4. AUC-ROC and Average Precision score of Isolation Forest model.

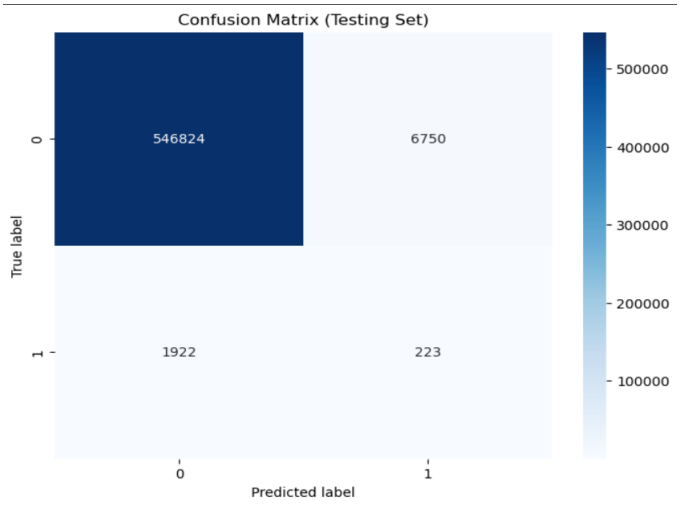

Fig. 4.5. Confusion Matrix of Isolation Forest model.

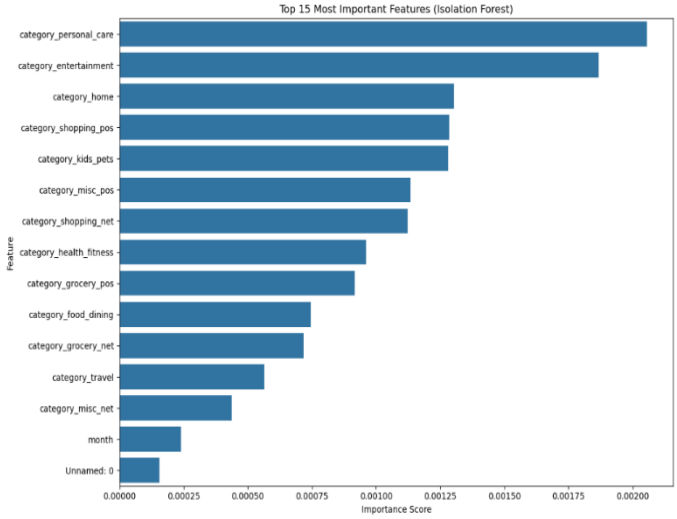

Fig. 4.6. Feature Importance Analysis of Isolation Forest model.

C. K-Mean Model Performance

```
Training Accuracy: 0.5473
Testing Accuracy: 0.5493
```

Fig. 5.1. Training and Testing scores using K-Means model.

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

           0       1.00      0.55      0.71    368549
           1       0.01      0.50      0.01      1930

    accuracy                           0.55    370479
   macro avg       0.50      0.52      0.36    370479
weighted avg       0.99      0.55      0.70    370479
```

Fig. 5.2. Classification report of K-Means model.

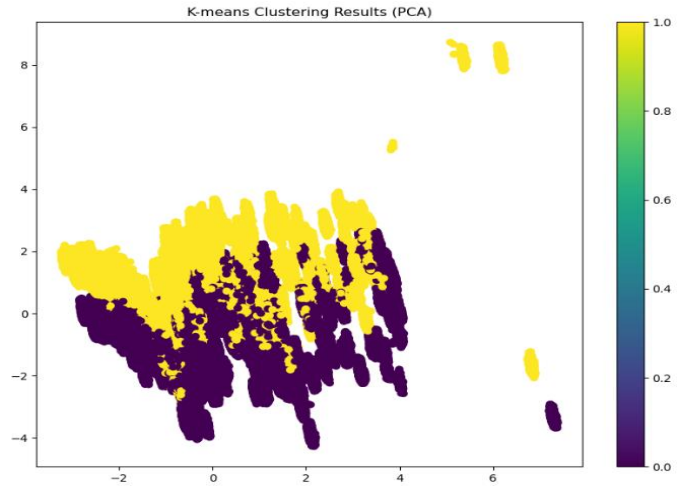

Fig. 5.3. Clustering of K-means model

```
AUC-ROC Score: 0.5248
Average Precision Score: 0.0055
```

Fig. 5.4. AUC-ROC and Average Precision score of K-means model

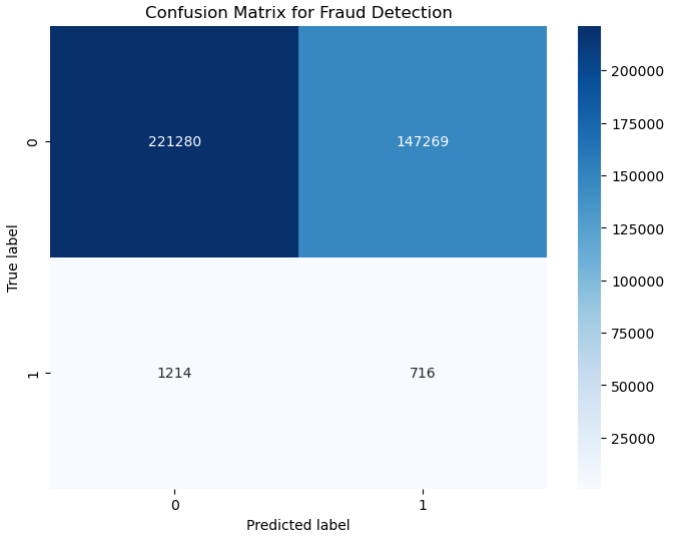

Fig. 5.5. Confusion Matrix of K-Means model

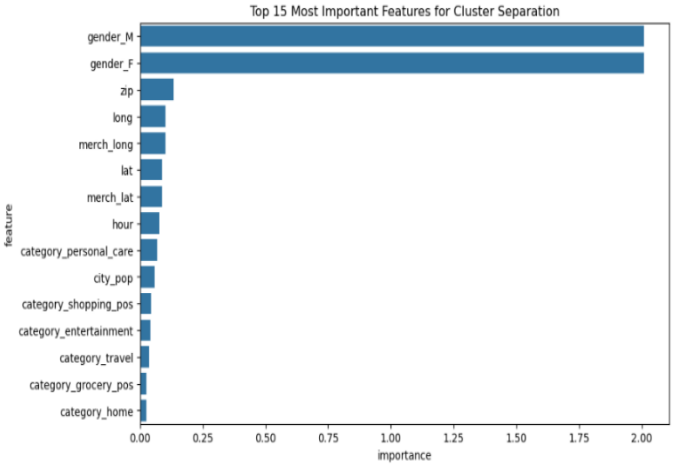

Fig. 5.6. Feature Importance Analysis of K-Means model

D. Convolutional Neural Network Model Performance

```
Training Accuracy: 0.9913
Testing Accuracy: 0.9847
```

Fig. 6.1. Training and Testing scores using CNN model.

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

           0       1.00      0.99      0.99    368549
           1       0.24      0.91      0.38      1930

    accuracy                           0.98    370479
   macro avg       0.62      0.95      0.69    370479
weighted avg       1.00      0.98      0.99    370479
```

Fig. 6.2. Classification report of CNN model.

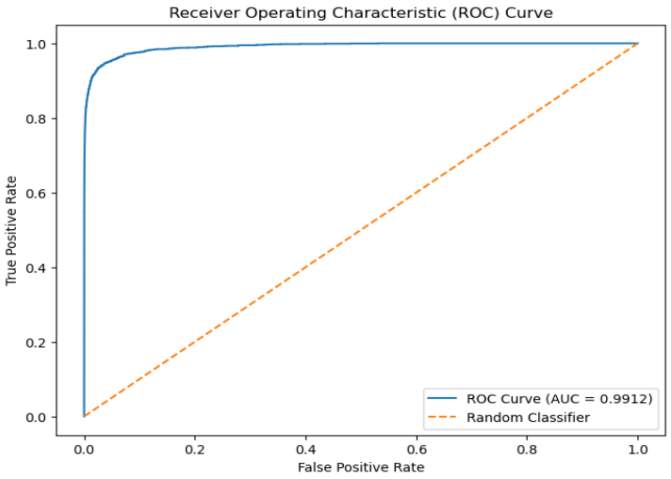

Fig. 6.3. ROC Curve of CNN model

```
AUC-ROC Score: 0.9912
Average Precision Score: 0.8051
```

Fig. 6.4. AUC-ROC and Average Precision score of CNN model

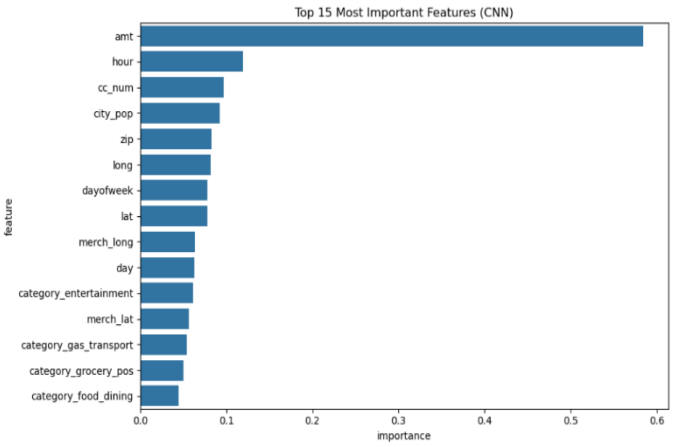

Fig. 6.6. Feature Importance Analysis of CNN model

Final Analysis of all the 4 models and their performance.

Logistic Regression	K- Mean	Isolation forest	CNN
- Accuracy: 0.88 - F1 score: 0.94 - Recall: 0.88	- Accuracy: 0.55 - F1 score: 0.71 - Recall: 0.55	- Accuracy: 0.98 - F1 score: 0.99 - Recall: 0.99	- Accuracy: 0.98 - F1 score: 0.99 - Recall: 0.99
Computational Power: Low	Computational Power: Low	Computational Power: Medium	Computational Power: High
- AUC-ROC score: 0.91 - Average Precision Score: 0.15	- AUC-ROC score: 0.52 - Average Precision Score: 0.005	- AUC-ROC score: 0.54 - Average Precision Score: 0.006	- AUC-ROC score: 0.99 - Average Precision Score: 0.80
Top Feature: Amount	Top Feature: Gender_M	Top Feature: Category_personal_care	Top Feature: Amount

VI. CONCLUSION

This study compares four machine learning models- Logistic Regression, K-Means clustering, Isolation Forest, and convolutional Neural Network (CNN) for credit card fraud detection. We evaluated these models using various set of metrics including accuracy, F1 score, recall, AUC-ROC score.

Our results reveal the performance of various models. The CNN model is able to generate a model with accuracy of 0.98 F1 score of 0.99, and AUC-ROC of 0.99. However, this superior performance comes at cost of high computational power. The logistic regression comes out as a good model with good performance showcasing scores with accuracy of 0.88, F1 score of 0.94, AUC-ROC of 0.91 and also has low computational power. Therefore, these two models emerge as a viable option for real-time fraud detection as well where accuracy is important and computational power is optional.

Interestingly, the Isolation Forest model achieves a high accuracy of .98 compared to CNN, but its low AUC-ROC score shows that there might be some potential issue with class separations. This tells us that it's important to consider multiple metrics in evaluating model performance, particularly in imbalanced classification problems like fraud detection. The K-Means clustering performs poorly across all metrics showcasing its not an ideal model to predict credit card fraud detection, this also indicates that unsupervised learning methods may not fit well with problems like credit card fraud detections.

These models explain the trade-off between models' complexity and performance in credit card fraud detection. Where models like CNN provides higher detection capability but models like Logistic regression offer strong balance between accuracy and computational efficiency. Finally, the choices of the models should be done based on specific requirements and constraints of fraud detection system that is needed to be developed.

This study helps contributing into the ongoing studies and development that happening around credit card fraud detection. Future Work can explore ensemble modelling techniques that uses strength of different models to improve the detection mechanism and develop computational efficient models that can run on any device with minimal requirements.

ACKNOLODGEMENT

I would like to extend my sincere and heartfelt thanks to my professor Dr. Pallavi who guided me by reviewing and providing feedback throughout this research and actively encouraged me to complete this work. I would also take a moment to appreciate Mr. Arun Samanta for taking out his time in reviewing my work and providing me a plagiarism report on the paper to keep this work original. The journey would not have been completed without the resources and knowledge provided by these faculties at presidency university.

I am also very grateful to the services provided by OpenAI and Anthropic for their resources and tools for various help that includes resources collection, content paraphrasing and debugging coding errors. Finally, I would like to thank Google scholar and Research gate for providing relevant articles which helped later during the development of project.

REFERENCE

[1] AlEmad, M. (2022). *Credit Card Fraud Detection Using Machine Learning.* RIT Digital Institutional Repository.

[2] BORA MEHAR SRI SATYA TEJA, B. M. (2022). A Research Paper on Credit Card Fraud Detection. *International Research Journal of Engineering and Technology*, 1-4.

[3] Brandon, H. (2022). *Synthetic Credit Card Transaction Generator used in the Sparkov program*. Retrieved from GitHub: https://github.com/namebrandon/Sparkov_Data_Generation

[4] Emmanuel Ileberi, Y. S. (2022). A machine learning based credit card fraud detection using the GA algorithm for feature selection. *Journal of Big data*, 2-15.

[5] Harris, B. (2020). *Credit Card Transactions Fraud Detection Dataset.* Retrieved from Kaggle: https://www.kaggle.com/datasets/kartik2112/fraud-detection

[6] Mr. Thirunavukkarasu.M, A. N. (2021). CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING. *International Journal of Computer Science and Mobile Computing*, 2-7.

[7] Rejwan Bin Sulaiman, V. S. (2022). Review of Machine Learning Approach on Credit Card Fraud Detection. *Human-Centric Intelligent Systems*, 1-12.

[8] S P Maniraj, A. S. (2019). Credit Card Fraud Detection using Machine Learning and Data Science. *International Journal of Engineering Research & Technology*, 2-4.

[9] Surbhi Bansal, R. H. (2024). A Review Paper on Feature Selection in Credit Card Fraud Detection. *International joint conference on computing sciences*, 1-5.

[10] Vaishnavi Nath Dornadulaa, G. S. (2019). Credit Card fraud Detection using Machine Learning Algorithms. *International Conference on Recent Trends in Advanced Computing*, 3-9.