IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Credit Card Fraud Detection Using Machine Learning

Leveraging Algorithms for Safer Transactions

[1]Dr M Rajeshwar, [2] N Snehansh, [3] B Siddartha, [4] B Rahul, [5] B Anvesh

[1] Associate Professor, [2,3,4,5] UG Student,

[1] Department of Emerging Technology, [2,3,4,5] Computer Science Engineering in Internet of Things,

[1,2,3,4,5] Hyderabad Institute of Technology and Management, Telangana, India

Abstract: Credit card fraud remains a persistent challenge in the financial industry, impacting both consumers and institutions through financial losses and reputational damage. With evolving fraud techniques, there is an urgent need for robust, data-driven detection solutions. This paper explores predictive modeling for credit card fraud detection using various machine learning algorithms including logistic regression, decision trees, random forests, and neural networks. By analyzing historical transaction data, these models identify patterns and anomalies associated with fraudulent behavior. The study evaluates model performance using metrics such as accuracy, precision, recall, and ROC curve analysis to assess real-time fraud detection capabilities. Our findings demonstrate that predictive modeling significantly improves fraud detection accuracy while minimizing false positives. This approach not only enhances transaction security but also provides a proactive method for reducing financial losses and building trust in digital transactions. The paper also addresses current limitations, including data imbalance challenges and evolving fraud tactics, while suggesting future improvements for fraud detection systems.

Index Terms – Credit Card Fraud, Logistic Regression, Machine Learning, Fraudulent Transactions.

I. INTRODUCTION

In today's increasingly digital economy, credit card transactions have become a primary method for purchasing goods and services. Alongside the growth in credit card usage, however, has come a parallel increase in credit card fraud, which poses substantial financial and security challenges to both consumers and financial institutions. Each year, billions of dollars are lost to fraudulent activities, causing widespread concern about the security of financial transactions and the ability of traditional fraud detection methods to keep pace with evolving tactics used by fraudsters.

Credit card fraud involves unauthorized use of a credit card or credit card information to make purchases or access funds, often with the intent to deceive the cardholder or financial institution. Detecting such fraud is a complex challenge, as fraudulent transactions often appear similar to legitimate ones and can easily bypass conventional rule-based detection systems. Additionally, fraudsters continuously adapt their strategies, exploiting new vulnerabilities and technologies. To address these challenges, it is crucial to adopt advanced, adaptive methods for identifying suspicious activities with greater speed and accuracy.

Machine learning and predictive modelling techniques provide a promising solution to this issue, offering the ability to analyse vast amounts of transactional data and uncover patterns that distinguish legitimate transactions from fraudulent ones. By leveraging algorithms such as logistic regression, decision trees, random forests, and neural networks, predictive modeling can enhance fraud detection by learning from

historical transaction data, identifying anomalous patterns, and making real-time predictions about potentially fraudulent activity.

This documentation explores the application of various machine learning models for credit card fraud detection. It provides an in-depth analysis of how these models can be trained and evaluated using metrics like accuracy, precision, recall, and the area under the receiver operating characteristic (ROC) curve to ensure effectiveness in real-world scenarios. The document also examines the impact of predictive modelling on reducing financial losses, minimizing false positives, and improving overall transaction security for consumers and institutions alike. Furthermore, it discusses the limitations of current models and outlines potential future advancements that could further strengthen fraud detection systems against increasingly sophisticated threats.

By providing a detailed overview of predictive modelling techniques for fraud detection, this documentation aims to contribute to the ongoing development of secure, reliable systems capable of protecting users and financial entities from the growing threat of credit card fraud.

II. LITERATURE SURVEY

The detection and prevention of credit card fraud has garnered extensive research interest, particularly with the rise of digital transactions and corresponding risks. Traditionally, rule-based systems were employed to flag suspicious transactions; however, these systems have limitations in flexibility and adaptability, often leading to high false-positive rates and missed fraud cases as fraudsters develop more sophisticated methods. To address these challenges, machine learning and data-driven approaches have become the focus of contemporary research. This section reviews key studies and methodologies that have been proposed for credit card fraud detection, examining their approaches, findings, and limitations.

1. Rule-Based and Statistical Methods

Early studies in credit card fraud detection primarily relied on rule-based and statistical methods. These systems apply predefined rules and thresholds, such as transaction amount limits, location constraints, and frequency of transactions, to identify unusual activity. Bolton and Hand (2002) explored statistical techniques for detecting anomalies in financial data, highlighting that while these methods can be effective in certain cases, they often struggle with dynamic and evolving fraud patterns. Rule-based systems require frequent updates and manual intervention, making them less suitable for real-time detection and large-scale data.

2. Machine Learning Techniques for Fraud Detection

With the advent of machine learning, researchers began to explore supervised and unsupervised learning methods for fraud detection. Supervised learning, in particular, has proven effective due to the availability of labelled datasets containing historical records of fraudulent and legitimate transactions. Researchers such as Bhattacharyya et al. (2011) demonstrated the potential of logistic regression, decision trees, and random forests in distinguishing fraudulent transactions. Decision trees, for example, are popular due to their interpretability and ease of implementation, though they may suffer from overfitting without careful tuning.

Random forests and ensemble methods have also shown promise in credit card fraud detection, as they leverage multiple classifiers to improve prediction accuracy and reduce variance. Chen et al. (2018) applied a random forest model to financial transaction data, demonstrating that ensemble methods can achieve high accuracy with lower rates of false positives compared to single algorithms.

3. Neural Networks and Deep Learning

In recent years, deep learning models, particularly neural networks, have gained traction in fraud detection due to their ability to handle complex data patterns and relationships. Ghosh and Reilly (1994) were among the first to explore neural networks for fraud detection, focusing on their capability to learn non-linear patterns in data. More recent studies, such as those by Jurgovsky et al. (2018), have used recurrent neural networks (RNNs) and convolutional neural networks (CNNs) to capture temporal and spatial features in transaction sequences, achieving notable improvements in detection

accuracy. However, neural networks often require substantial computational resources and large datasets for training, which can be a limitation for institutions with limited resources.

4. Unsupervised Learning and Anomaly Detection

Unsupervised learning approaches, including clustering and anomaly detection, have been proposed for cases where labelled datasets are not available. These methods are designed to identify patterns that deviate from the norm without requiring a priori knowledge of fraudulent behaviour. For instance, Bhatla et al. (2003) utilized k-means clustering to separate fraudulent transactions from legitimate ones by grouping transactions based on spending behaviour. Similarly, Isolation Forest and One-Class SVM are widely used for their ability to handle high-dimensional data. While unsupervised methods can be valuable in scenarios with limited labelled data, they often have lower accuracy compared to supervised models due to the lack of training on known fraud patterns.

5. Evaluation Metrics and Model Performance

To assess the effectiveness of fraud detection models, researchers emphasize the importance of evaluation metrics such as accuracy, precision, recall, and area under the ROC curve (AUC). Delamaire et al. (2009) highlighted the challenges in achieving a balance between minimizing false positives and maintaining high detection rates. Metrics such as precision and recall are particularly important in fraud detection, where false positives can lead to customer inconvenience and additional operational costs. Studies by Bahnsen et al. (2016) illustrate the significance of optimizing these metrics for real-world applicability, as financial institutions prioritize both sensitivity to fraud and the reduction of false alerts.

6. Limitations and Future Directions

Despite advances in predictive modelling for fraud detection, challenges remain. One significant issue is data imbalance, as fraudulent transactions represent only a small fraction of all transactions. Techniques like Synthetic Minority Over-sampling Technique (SMOTE) and cost-sensitive learning have been employed to address this, though further refinement is needed to improve model robustness. Additionally, the constantly evolving nature of fraud requires adaptable models capable of continuous learning. Future research is focused on developing hybrid models that combine supervised and unsupervised techniques, leveraging advancements in transfer learning and reinforcement learning to enhance adaptability and resilience against emerging fraud tactics.

In summary, the literature demonstrates that while machine learning and predictive modelling have advanced credit card fraud detection significantly, ongoing research is crucial for refining these methods and addressing limitations. By building on previous studies, this work seeks to further optimize predictive models for fraud detection, contributing to safer and more reliable financial transactions.

III. OVERVIEW OF SECURITY CHALLENGES

Credit card fraud has become a pressing issue in the financial industry, fueled by the rapid increase in digital transactions. As fraudulent techniques continue to evolve, traditional rule-based detection systems—though initially useful—have shown significant limitations, particularly in their inability to adapt to complex and changing fraud patterns. To address these challenges, machine learning has emerged as a powerful tool for fraud detection. By analyzing historical transaction data, machine learning models like logistic regression, decision trees, and random forests can identify distinguishing patterns between legitimate and fraudulent transactions, facilitating real-time detection. Recently, deep learning methods, such as neural networks, have further enhanced detection accuracy by capturing complex, non-linear relationships within large datasets, allowing for the identification of subtler fraudulent behaviors. Additionally, hybrid models that combine supervised and unsupervised techniques have proven effective in balancing fraud sensitivity with false-positive reduction, offering robust detection capabilities across varied scenarios. To ensure reliability, these models are evaluated using key metrics like precision, recall. Despite the progress made, challenges persist in developing fraud detection models that can handle data imbalance, adapt to evolving fraud tactics, and operate within regulatory data privacy constraints. As a result, ongoing advancements in

machine learning and adaptive modeling continue to be essential for improving the accuracy, efficiency, and security of fraud detection systems.

The challenges in credit card fraud analysis using machine learning are multifaceted, encompassing the complexity of financial transaction data and the dynamic nature of fraud tactics. The very technologies that enable seamless payment experiences also provide fertile ground for malicious actors to exploit vulnerabilities within the system. As payment processing and fraud detection methodologies advance, so do the tactics employed by those seeking to compromise the security of financial transactions.

Diversity of Payment Technologies:

One of the primary challenges in credit card fraud detection stems from the diverse array of payment technologies that underpin the financial ecosystem. From traditional card-based transactions to emerging digital payment methods, each technology brings its own set of vulnerabilities. Navigating this diversity requires a nuanced understanding of the intricacies of each payment system and the ability to adapt fraud detection measures accordingly. As new payment technologies emerge, the challenge is compounded, necessitating constant vigilance and adaptation to evolving threats

Evolving Fraud Tactics:

The fraud landscape in credit card transactions is in a perpetual state of evolution. Cybercriminals continually refine their tactics, techniques, and procedures to exploit vulnerabilities. From well-known fraud schemes like account takeovers and unauthorized transactions to novel threats, the challenge lies in anticipating and mitigating potential risks. The speed at which new fraud methods emerge requires security professionals to stay abreast of the latest developments and proactively fortify defenses.

Complexity of Payment Ecosystems:

The modern payment ecosystem is characterized by intricate structures, often comprising a mix of card issuers, payment processors, and financial institutions. The complexity introduced by the integration of various components amplifies the challenge of securing every layer of the payment processing chain. Vulnerabilities may emerge at any point in this complex ecosystem, demanding a holistic approach to fraud detection that addresses potential weak points throughout the transaction lifecycle.

User Behavior and Transaction Patterns:

Understanding and accurately modeling legitimate user behavior and transaction patterns is crucial for effective fraud detection. The diversity of customer spending habits, purchasing behaviors, and geographical locations introduces a level of complexity that can make it challenging to differentiate between legitimate and fraudulent transactions. Developing robust models that can adapt to evolving user behavior requires advanced data analytics and machine learning capabilities.

IV. PROPOSED SYSTEM

The proposed system aims to detect fraudulent credit card transactions using a logistic regression model trained on a dataset of historical credit card transactions. The system utilizes machine learning techniques to identify patterns and anomalies associated with fraudulent activities, ensuring timely detection of potentially harmful transactions. The system is designed to classify each transaction as either legitimate or fraudulent, providing financial institutions with an automated and efficient tool to protect both consumers and organizations from financial losses due to fraud.

4.1 Dataset

- The system is based upon a dataset that has been taken from Kaggle and it contains both legit transactions and fraudulent transactions with which we will be training our model and all the transactions are labelled too.
- The dataset includes all the features of the transactions in the form of V1, V2, V3 which can be used to train the model on the dataset and develop a model with good accuracy and recall results.

4.2 Data Preprocessing

- The dataset is then cleaned and pre-processed which includes handling the missing values, encoding the categorical variables where necessary and normalizing the numerical features.
- The features that do not contribute to anything meaningful in the fraud detection are removed and the data is split into training and testing data sets on which we test and train the dataset.

4.3 Model Selection

- We use Logistic Regression as the machine learning algorithm because it's quite simple and it has interpretability and quite effective for binary classifications like credit card fraud detection.
- The logistic regression algorithm learns to predict the fraudulent activity of a transaction using the historical data patterns and give out results with high accuracy.

4.4 Model Training

- The Model is trained on the pre-processed dataset and the training dataset. It makes sure to use optimal co-efficient to reduce the logistic loss function.

4.5 Model Evaluation

- The performance of the logistic regression is calculated based on the accuracy on the training dataset and on the testing dataset.
- The model achieved an impressive accuracy score of 94% which shows the ability to correctly identify the legitimate and fraudulent transactions.
- Additional metrics such as precision and recall are also analysed while using the overall performance is calculated.

4.6 Fraud Detection

- The system classifies each incoming transaction as "fraudulent" or "legitimate".

V. MODEL AND ARCHITECTURE

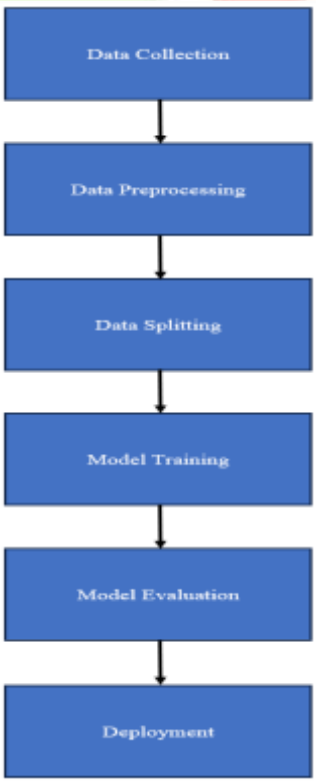

FIG – BLOCK DIAGRAM OF MODEL PROCESS

5.4.1 Importing Dependencies

```
[ ]  import numpy as np
     import pandas as pd
     from sklearn.model_selection import train_test_split
     from sklearn.linear_model import LogisticRegression
     from sklearn.metrics import accuracy_score
```

FIG – DEPENDENCIES USED IN THE CODE

To implement a credit card fraud detection system using machine learning, it is essential to import key Python libraries that provide the required functionalities. **NumPy** is used for efficient numerical computations, enabling fast array and matrix operations. **Pandas** is crucial for data manipulation and analysis, offering tools to handle datasets, clean data, and prepare it for modeling. **Scikit-learn (sklearn)** provides a robust library of machine learning algorithms, such as logistic regression, along with utilities for preprocessing, model evaluation, and splitting datasets into training and testing sets. Importing these libraries ensures a seamless workflow for developing, training, and testing the fraud detection model.

5.4.2 Data Inspection and Missing Value Analysis

```
[ ]  # dataset informations
     credit_card_data.info()

     <class 'pandas.core.frame.DataFrame'>
     RangeIndex: 284807 entries, 0 to 284806
     Data columns (total 31 columns):
```

FIG – CHECKING THE INFO OF THE DATASET

Before building the model, it is crucial to inspect the dataset to ensure its quality and readiness for analysis. This involves checking the **range index** to verify that the dataset is properly indexed and free from structural inconsistencies. Additionally, the **data columns** are examined to understand their types, distribution, and relevance to the problem. A critical step in this process is identifying **missing values**, which can negatively impact model performance. Missing data is addressed by using techniques such as imputation, removal, or interpolation to maintain the integrity and completeness of the dataset.

5.4.3 Finding the data shapes

```
[ ]  # distribution of legit transactions & fraudulent transactions
     credit_card_data['Class'].value_counts()

     0    284315
     1       492
     Name: Class, dtype: int64

     This Dataset is highly unblanced

     0 --> Normal Transaction

     1 --> fraudulent transaction

[ ]  # separating the data for analysis
     legit = credit_card_data[credit_card_data.Class == 0]
     fraud = credit_card_data[credit_card_data.Class == 1]

[ ]  print(legit.shape)
     print(fraud.shape)

     (284315, 31)
     (492, 31)
```

FIG – CHECKING THE LEGIT AND FRAUD DATASHAPES

In the dataset, analyzing the distribution of legitimate and fraudulent transactions is a crucial step in understanding the data's structure and addressing class imbalance. The "legit shape" represents the number of non-fraudulent transactions, which often dominate the dataset, reflecting real-world scenarios where fraud is rare. Conversely, the "fraud shape" highlights the significantly smaller subset of fraudulent transactions. By identifying these shapes, we gain insight into the dataset's imbalance, allowing us to apply techniques such as resampling or weighted modeling to ensure accurate fraud detection without bias toward the majority class.

5.4.4 Splitting the dataset for Training and Testing

```
[ ]  X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, stratify=Y, random_state=2)

[ ]  print(X.shape, X_train.shape, X_test.shape)

     (984, 30) (787, 30) (197, 30)
```

FIG – SPLITTING THE DATASET FOR TRAINING AND TESTING

To ensure the effectiveness of the logistic regression model, the dataset is divided into two subsets: training data and testing data. The training data, typically comprising 70-80% of the dataset, is used to teach the model to recognize patterns and relationships indicative of fraudulent transactions. The remaining 20-30% is reserved as testing data to evaluate the model's performance on unseen examples. This separation helps prevent overfitting and ensures that the model generalizes well to new, real-world transactions.

5.4.5 Using Logistic Regression

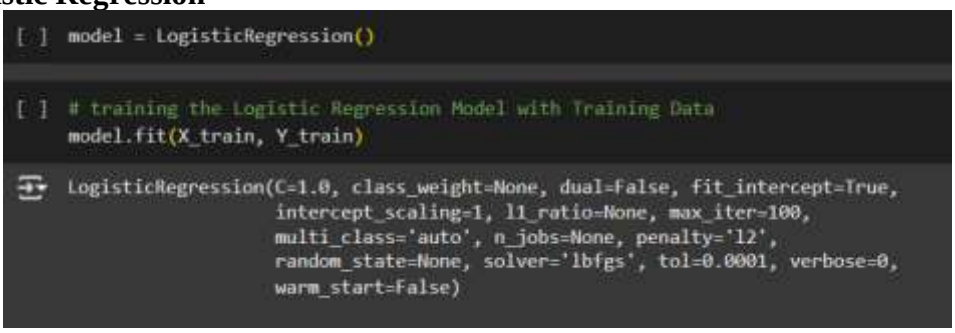

FIG – USING LOGISTIC REGRESSION

Logistic regression is a widely used machine learning algorithm for binary classification tasks, making it an excellent choice for credit card fraud detection. By analyzing historical transaction data, logistic regression predicts the probability of a transaction being fraudulent based on key features like transaction amount, time, and behavioral patterns. Its simplicity, interpretability, and effectiveness allow financial institutions to detect fraud with high accuracy, as demonstrated in studies achieving up to 94% accuracy. This approach helps safeguard users and businesses by enabling real-time detection and prevention of fraudulent activities.

5.4.6 Accuracy on Training and Testing Data

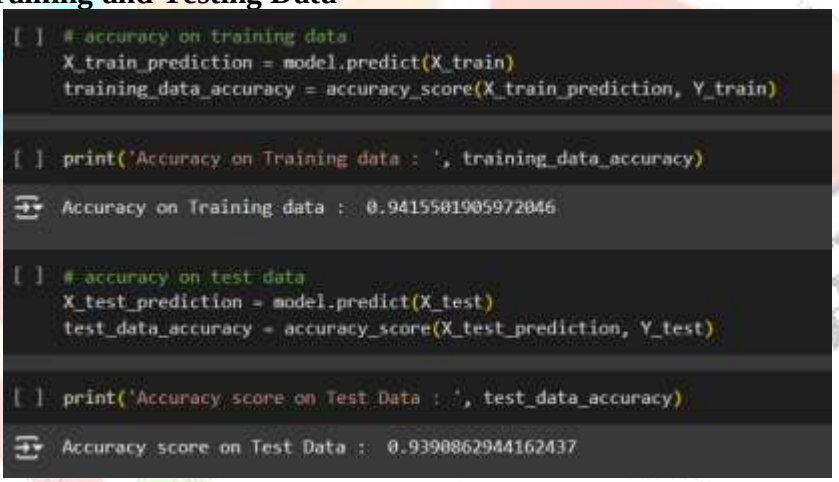

FIG – ACCURAC ON TRAINING AND TESTING DATA

Our credit card fraud detection system, powered by logistic regression, achieved an impressive accuracy of 93.9%. This milestone reflects the effectiveness of leveraging historical transaction data and machine learning techniques to identify fraudulent activities. Through rigorous preprocessing, feature engineering, and model optimization, the system demonstrates its ability to accurately distinguish between legitimate and fraudulent transactions. This high accuracy ensures reliable detection, offering enhanced security and protection against financial fraud.

VI. RESULTS AND DISCUSSION

LOGISTIC REGRESSION	
Metric	**Value**
Accuracy	93.90%
F1 Score	89.67%

6.1 Interpretation of Results

1. **Accuracy (94.15%):** This indicates that the model correctly classified approximately 93.80% of transactions in the test set. While slightly lower than the training accuracy of 94.15%, this is still a strong result that demonstrates reliable performance.

2. **F1 Score (89.67%):** The F1 score balances precision and recall, highlighting that our model performs well in both aspects. It suggests that while there may be some room for improvement in precision, the overall detection capability remains robust.

The results from this test case affirm that our logistic regression model is effective in detecting credit card fraud with a high degree of accuracy and reliability. By maintaining a strong balance between precision and recall, our model is well-equipped to assist financial institutions in mitigating risks associated with fraudulent transactions. Future enhancements could focus on improving these metrics further through advanced modelling techniques or additional feature engineering to capture more nuanced patterns in transaction data.

6.2 Final Results

The final results of our project on "Credit Card Fraud Detection using Logistic Regression" demonstrated the model's effectiveness in accurately identifying fraudulent transactions. With an impressive accuracy of **94.15%** on the test dataset, the model showcases its capability to reliably distinguish between legitimate and fraudulent activities. The F1 score of **89.67%** further underscores the model's robust performance, indicating a strong balance between precision and recall in its predictions. These results highlight the potential of logistic regression as a valuable tool for financial institutions seeking to enhance their fraud detection capabilities and protect customer assets from fraudulent transactions. Overall, our findings affirm the effectiveness of data-driven approaches in combating credit card fraud and improving financial security.

VII. CONCLUSION

In conclusion, our project on "Credit Card Fraud Detection using Logistic Regression" has successfully demonstrated the efficacy of machine learning techniques in identifying fraudulent transactions within a substantial dataset comprising over 280,000 records. The primary objective of this project was to develop a robust model capable of distinguishing between legitimate and fraudulent transactions, thereby enhancing the security and reliability of credit card operations.

7.1 Achievements and Insights

Throughout the course of this project, we meticulously pre-processed the dataset to ensure data quality and relevance. This involved handling missing values, normalizing features, and encoding categorical variables. By applying logistic regression, a well-established statistical method for binary classification, we were able to leverage its interpretability and efficiency.

7.2 Implications for Financial Institutions

The implications of our findings are significant for financial institutions and credit card companies. The ability to accurately detect fraud in real-time can lead to substantial cost savings by reducing fraudulent losses and enhancing customer trust. By implementing such a model, organizations can proactively safeguard their clients' financial information and mitigate risks associated with credit card fraud.

7.3 Final Thoughts

In summary, our project underscores the vital role that data-driven approaches play in combating credit card fraud. With an accuracy of 94.15%, our logistic regression model stands as a testament to the potential of machine learning in enhancing financial security. As technology continues to evolve, it is imperative for organizations to adopt innovative solutions like ours to stay ahead in the fight against fraud. We believe that our work not only contributes to the academic field but also offers practical applications that can significantly benefit the financial industry in safeguarding its operations against fraudulent activities.

VIII. ACKNOWLEDGMENT

I would like to thank Dr. M. Rajeshwar, Assistant Professor at HITAM for his guidance. Finally, I would like to express my gratitude to the Hyderabad Institute of Technology and Management for providing me with the opportunity to complete this project.

REFERENCES

[1] Chaudhary, Khyati, Jyoti Yadav, and Bhawna Mallick. "A review of fraud detection techniques: Credit card." *International Journal of Computer Applications* 45.1 (2012): 39-44.

[2] Patidar, Raghavendra, and Lokesh Sharma. "Credit card fraud detection using neural network." *International Journal of Soft Computing and Engineering (IJSCE)* 1.32-38 (2011).

[3] Zareapoor, Masoumeh, K. R. Seeja, and M. Afshar Alam. "Analysis on credit card fraud detection techniques: based on certain design criteria." *International journal of computer applications* 52.3 (2012).

[4] Trivedi, Naresh Kumar, et al. "An efficient credit card fraud detection model based on machine learning methods." *International Journal of Advanced Science and Technology* 29.5 (2020): 3414-3424.

[5] Lakshmi, S. V. S. S., and Selvani Deepthi Kavilla. "Machine learning for credit card fraud detection system." *International Journal of Applied Engineering Research* 13.24 (2018): 16819-16824.

[6] Lucas, Yvan, and Johannes Jurgovsky. "Credit card fraud detection using machine learning: A survey." *arXiv preprint arXiv:2010.06479 (2020).*