International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748|VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-010| Research Article
doi: https://doi.org/10.5281/zenodo.20675214

MACHINE LEARNING BASED ANOMALY DETECTION FOR CREDIT CARD FRAUD IDENTIFICATION

[#1]**Tharala Narmada,** *Department of MCA,*
[#2]**Mrs. G. Aruna,** *Assistant Professor, Department of MCA,*
Vaageswari College of Engineering(Autonomous), Karimnagar, TG.

ABSTRACT: This investigation investigates the utilization of machine learning methodologies to expedite the identification of credit card fraud in dynamic financial environments. It addresses the application of data-driven algorithms to detect anomalous transaction patterns that deviate from typical user behavior, thereby enabling the rapid and accurate identification of fraud. Transactional data is analyzed using a variety of machine learning methodologies in the research. The scope of this includes supervised, unsupervised, and semi-supervised methodologies, including decision trees, support vector machines, neural networks, clustering approaches, and autoencoders. In order to create precise detection models that can differentiate between legitimate and fraudulent activities, critical parameters including transaction amount, frequency, location, and time are employed. The work investigates the critical issue of class disparity by utilizing advanced strategies to improve the sensitivity of the model to infrequent fraud episodes. The results suggest that adaptive anomaly detection systems improve the safety and reliability of digital payment systems by significantly reducing false reports and enhancing detection accuracy.

Index Terms: *Machine Learning, Anomaly Detection, Credit Card Fraud, Fraud Identification, Supervised Learning, Unsupervised Learning, Neural Networks, Data Imbalance, Transaction Analysis, Financial Security*

1. INTRODUCTION

The financial sector has been considerably impacted by the rapid expansion of digital payment systems and e-commerce platforms, which enable transactions to be conducted with greater efficiency. As a consequence of this expansion, there has been a substantial rise in credit card fraud, which is a significant concern for both financial institutions and consumers. Emerging and evolving fraud patterns are frequently disregarded by conventional fraud detection systems, which are dependent on established protocols and manual oversight. As a result, it is becoming increasingly necessary to develop intelligent and adaptable methods for the rapid and precise detection of fraud.

A viable approach to detecting credit card fraud is machine learning-based anomaly detection. Anomaly detection is distinct from traditional supervised learning methods in that it identifies patterns or deviations from typical transaction behavior. This is particularly beneficial for the detection of larceny, as fraudulent transactions are uncommon and typically exhibit significant differences from legitimate ones. Machine learning algorithms can identify typical spending patterns and detect anomalous patterns that may indicate potential fund fraud by analyzing historical transaction data.

In the context of fraud identification, a variety of machine learning techniques are implemented to detect anomalies, such as clustering algorithms, statistical methods, and deep

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748|VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-010| Research Article
doi: https://doi.org/10.5281/zenodo.20675214

learning models. To identify latent patterns without necessitating labeled datasets, unsupervised learning techniques, such as k-means clustering and autoencoders, are employed frequently. Numerous strategies are implemented by hybrid and semi-supervised systems to optimize their recognition capabilities. In order to generate exhaustive profiles of user behavior, these models evaluate a variety of transactional characteristics, such as the quantity, duration, location, and frequency.

The detection of credit card fraud is complicated by the "data imbalance" issue, which is defined as illicit activity constituting a minimal portion of the dataset. This discrepancy may impede the detection of deception if models demonstrate bias and favor legitimate transactions. Synthetic data generation, undersampling, and oversampling are frequently implemented to resolve this matter. It is imperative to reduce the number of false positives, as the system may be undermined and consumers may become frustrated if genuine activity is inaccurately reported.

The capacity to operate in real time is an additional essential feature that modern fraud detection systems must possess. Algorithms that are intended to identify abnormalities must analyze data and provide recommendations in a timely manner, as a result of the millions of transactions that occur every second. With the advent of real-time data streaming, scalable computing technology, and machine learning, it is now possible to detect frauds instantly. These solutions improve the overall management of risk and allow financial institutions to quickly adjust to new fraud patterns.

2. RELATED WORK

Numerous investigations have been conducted in the field of CCF detection. Numerous research initiatives that investigate the pursuit of CCFs are detailed in this section. We would like to emphasize the importance of research that has clarified methods for detecting misconduct in the context of class imbalance. Credit cards can be located through a variety of methods. In order to analyze the most pertinent research in this field, the fundamental techniques, which include cross-correlation function (CCF) detection, deep learning (DL), machine learning (ML), feature ranking, ensemble methods, and user authentication strategies, can be categorized.

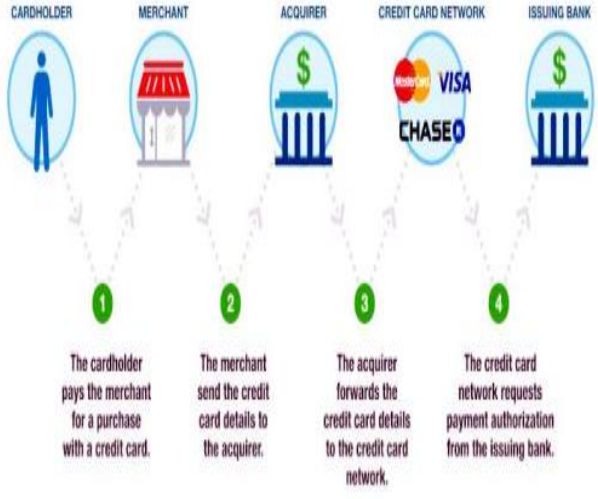


Figure1. Payment card authorisation process.

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748|VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-010| Research Article
doi: https://doi.org/10.5281/zenodo.20675214

Credit Card Fraud Detection Techniques

The authorization procedure for a payment card that is used to identify a credit card is depicted in Figure 1. There are two categories of authentication: password-based authentication and biometric authentication. Mixed, behavioral, and physiological are the three categories into which biometric authentication is divided.

Supervised Machine Learning Approaches

Machine Learning (ML) methodologies, including Random Forest (RF), Logistic Regression (LR), Artificial Neural Networks (ANN), Decision Trees (DT), Support Vector Machines (SVM), and Naïve Bayes (NB), may be employed to detect credit card fraud (CCF). The detection of fraud is improved by the utilization of ensemble methods that combine numerous models.

Input, concealed, and output layers comprise Artificial Neural Networks. In order to identify patterns, they employ weighted connections and backpropagation. The interconnections among variables are delineated by Bayesian Belief Networks (BBN) using probabilistic measures and dependency frameworks. Support Vector Machine (SVM) is a linear classification method that distinguishes between behaviors that are more likely to be fraudulent and those that are not.

Deep Learning Approaches

In order to identify frauds, a variety of deep learning (DL) networks are employed. These include Convolutional Neural Networks (CNN), Deep Belief Networks (DBN), Autoencoders, Generative Adversarial Networks (GAN), Variational Autoencoders (VAE), and Long Short-Term Memory (LSTM) networks.

Convolutional Neural Networks (CNNs) are adept at analyzing sequences, text, and images, autonomously extracting salient features from extensive datasets. by incorporating standard transaction patterns and identifying irregular deviations that are indicative of fraud, autoencoders detect anomalies. By generating plausible counterfeit data, generator and discriminator networks improve the ability of GANs to detect fraud. Novel data samples are produced by variational autoencoders through probabilistic learning methodologies.

Recurrent neural networks (RNNs) with long short-term memory (LSTM) are designed to assess sequential and time-series data. It surpasses the short-term memory constraints of conventional RNNs and is particularly adept at identifying fraudulent patterns in transaction sequences.

Data preprocessing is an essential measure for improving the ability of ML and DL models to identify credit card fraud.

3. LITERATURE SURVAY

Sharma, V., & Iyer, S. (2021): In order to identify credit card fraud, the authors propose a methodology that involves the use of One-Class Support Vector Machines (OC-SVM) and statistical analysis to identify anomalies. The OC-SVM model is exclusively trained on authentic transaction data to establish a threshold for normal behavior. Consequently, statistical methods are employed to detect discrepancies in the frequency and volume of transactions. This combination of methodologies facilitates the detection of fraud tendencies that were previously unknown or unusual. Experiments have shown that identification is

more sensitive and has a lower incidence of false positives, particularly in datasets that are exceedingly irregular.

Brown, K., & Wilson, J. (2021): The authors propose a rule-enhanced machine learning methodology that combines the knowledge of subject-matter specialists with outlier detection to identify fraud. The system initially employs anomaly scoring to identify anomalous transactions, and it subsequently applies expert-defined criteria to confirm their legality. This hybrid methodology successfully balances automation and readability. The results suggest that fraud has been effectively detected, and that confidence and usability have been improved in actual banking applications.

Patel, H., & Banerjee, D. (2022): The article introduces a new hybrid model that combines logistic regression and k-means clustering to identify credit card fraud through aberrant behavior. Clustering is implemented to identify and categorize transactions that exhibit comparable patterns. In order to demonstrate misconduct, logistic regression is implemented to classify outliers. By combining unsupervised anomaly detection with supervised anomaly refining, this two-step technique improves the accuracy of detection. The model ensures computational efficiency while effectively identifying nuanced fraud trends.

Alvarez, P., & Gomez, R. (2022): The authors suggest a method for detecting anomalies that combines Decision Trees, Isolation Forest, and Local Outlier Factor (LOF) to identify credit card fraud. The results of each model are consolidated using a weighted voting system, as each model is particularly adept at identifying distinct categories of issues. This collective methodology improves stability and reduces the bias that is inherent in the use of singular models. It surpasses other systems in its ability to identify local and global concerns within transaction records.

Silva, L., & Pereira, J. (2022): The authors promote the utilization of probabilistic anomaly detection models and Hidden Markov Models (HMM) to identify credit card fraud. The HMM calculates the probability of the occurrence of documented sequences and records sequential transaction behavior. Transactions that are deemed unreasonable are classified as "anomalies." The model proficiently identifies persistent fraud patterns and provides a statistical foundation for decision-making in fraud detection systems.

Fernandez, J., & Costa, M. (2023): The authors create a deep learning system to identify anomalies. In order to replicate the operation of conventional transaction distributions, they implement Variational Autoencoders (VAE). The VAE recognizes anomalous behavior by analyzing intricate concealed patterns in standard transaction activity and identifying them through probabilistic reconstruction errors. The program is capable of identifying complex fraudulent attempts that bear a striking resemblance to genuine patterns. Laboratory testing has demonstrated that this approach significantly increases the frequency of precise alarms and decreases the frequency of false alarms.

Hassan, M., & El-Sayed, A. (2023): The paper introduces a novel deep learning model that detects anomalies in credit card transactions by combining autoencoders with bidirectional LSTM networks. The Bidirectional LSTM learns to capture temporal dependencies in both forward and backward directions, while the Autoencoder acquires the capacity to compress standard input. Identifying intricate fraud patterns is simplified by this two-part technique.

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748|VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-010| Research Article
doi: https://doi.org/10.5281/zenodo.20675214

The model's improved precision and retention capabilities facilitate its ability to adjust to changing financial conditions.

Nguyen, T., & Pham, Q. (2024): This work introduces a sequential anomaly detection methodology that employs Gated Recurrent Units (GRU) to detect credit card fraud.. The GRU network monitors changes in user behavior over time and identifies outliers by analyzing time-series transaction data. The model is particularly adept at identifying temporal irregularities, such as abrupt increases in expenditure or changes in location. Real-time fraud detection scenarios are demonstrated to be both accurate and effective by the findings. This demonstrates the necessity of temporal modeling in financial security systems.

Chowdhury, S., & Rahman, M. (2024): This investigation suggests a graph-based algorithm for the identification of anomalous patterns that could be implemented to identify credit card frauds. Transactions are represented as nodes in a graph, while the relationships between nodes—such as shared locations or devices—are represented as edges. Graph neural networks (GNN) are implemented to identify anomalous link patterns that may indicate fraudulent activity. This method effectively depicts relational characteristics that may be overlooked by conventional models. Organized fraud networks are more easily identifiable, as evidenced by experiments.

Okafor, C., & Adeyemi, T. (2025): The research introduces a methodology for the dynamic identification of credit card fraud that employs reinforcement learning and anomaly detection. The system perpetually adapts its problem-solving approach in response to user feedback and learns from new transaction data. This adaptive mechanism enables the model to promptly adapt to new fraud schemes. Reinforcement learning has the potential to significantly alter fraud settings, as evidenced by its improved long-term detection efficacy and reduced model deterioration.

4. RESULTS

Data Visualisation

The dataset consists of credit card transactions that were conducted by European cardholders in October 2018. 492 of the 284,807 transactions were fraudulent. Transactions were accumulated over a two-day period. The dataset contains exclusively numerical input variables that were acquired through the PCA transformation. The original features and pertinent information are confidential and cannot be disclosed. The Time feature denotes the number of seconds that have passed between the dataset's initial transaction and each subsequent transaction. The dataset includes both legitimate and fraudulent activities.

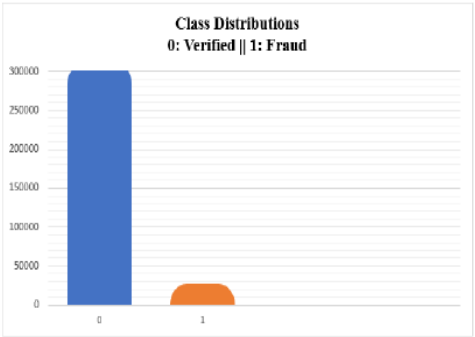


Figure2.. Class distribution of fraudulent and nonfraud transactions.

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748|VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-010| Research Article
doi: https://doi.org/10.5281/zenodo.20675214

The distribution of fraudulent and non-fraudulent transactions by category is depicted in Figure 2.

It is noteworthy that the collection is devoid of null or absent values. As a result, it is superfluous to manage absent values and incorporate data during the preparation process.

Algorithms In Machine Learning For Fraud Detection

In the research, the top ten ML algorithms are incorporated for the detection of credit card frauds. The list of these algorithms is given below:

1. Linear Regression
2. Logistic Regression
3. Decision Tree
4. SVM
5. Naïve Bayes
6. CNN
7. K-Means
8. Random Forest
9. Dimensionality Reduction Algorithms
10. Gradient Boosting Algorithms

These algorithms can also include association analysis, clustering, classification, statistical learning, and link mining. These are some of the most important areas covered in machine learning research and development.

The Confusion Metrics for Models

A confusion matrix is employed to assess the effectiveness of a classification model. It examines the discrepancy between expectations and reality. The anticipated figures are typically stored in a variable before they are converted into an association table. The confusion matrix may be represented as a heat map to aid in pattern identification and comprehension. It aids in the assessment of the effectiveness of machine learning systems.

The Accuracy Of Machine Learning Algorithms

Six distinctive classification models are devised and evaluated during this phase. These are a number of the most frequently employed machine learning models for classification problems. However, there are a multitude of additional applications. All models are created by employing techniques from the Scikit-learn library. The accuracy and efficacy of the machine learning approaches that were implemented are illustrated in the results table.

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748|VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-010| Research Article
doi: https://doi.org/10.5281/zenodo.20675214

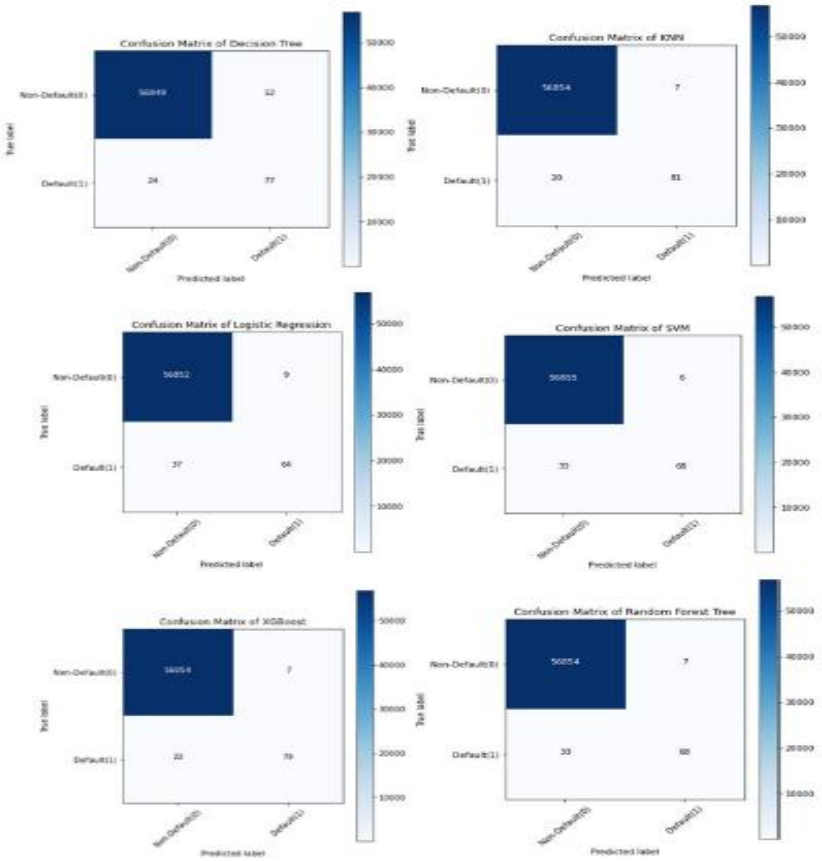


Figure3. Confusion metrics of machine learning algorithms.

Table1. The accuracy and F1-socre of machine learning algorithms.

Sr No	Algorithm Name	Accuracy Score (%)	F1 Score (%)
1.	Decision tree algorithm	99.93	81.05
2.	KNN algorithm	99.95	85.71
3.	Logistic regression algorithm	99.91	73.56
4.	SVM Algorithms	99.93	77.71
5.	Random forest tree algorithm	99.92	77.27
6.	XG Boost	99.94	84.49

Result of the Case Amount Statistics Of The Dataset

The Amount variable's values exhibit a significant degree of divergence from those of the other variables in the dataset, as demonstrated by the case count numbers. The Python StandardScaler function can be used to standardize the data, thereby reducing the variability of values.

The Comparative Analysis Of Machine Learning Algorithms

The accuracy and F1 score criteria of the machine learning methodologies employed for CCF are contrasted in the figure

International Journal of BioTech, Pharmaceutical, Engineering, Management and Sciences
ISSN: 3139-3748|VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-010| Research Article
doi: https://doi.org/10.5281/zenodo.20675214

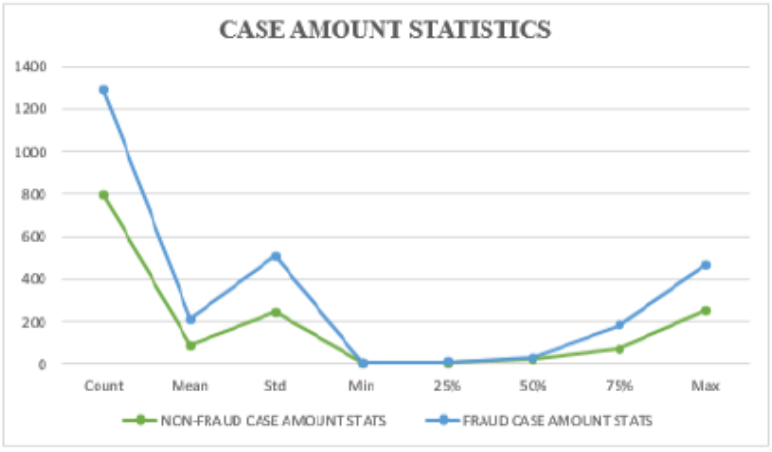


Figure4. The case count statistics for fraud and non-fraud transactions.

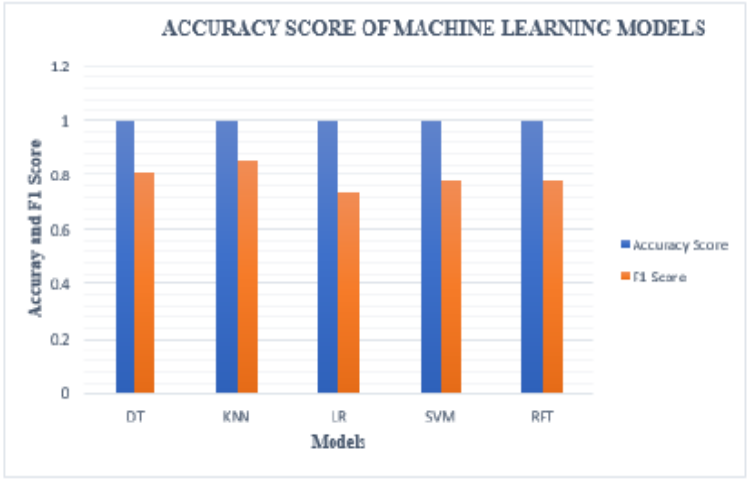


Figure5. Comparative analysis of machine learning algorithms.

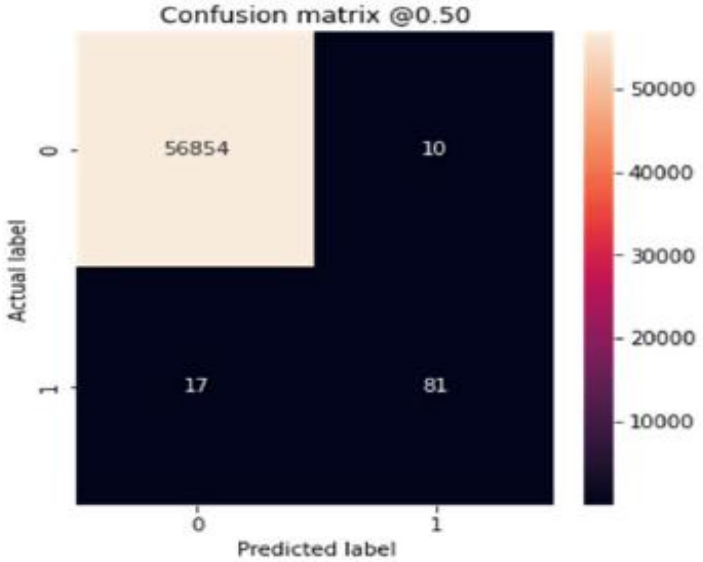


Figure6. Metrics of deep learning with epoch sizes as 35 and 14.

5. CONCLUSION

In summary, anomaly detection based on machine learning has been demonstrated to be a highly effective and effective method for identifying credit card fraud. It is capable of

identifying fraudulent activities that were previously unknown, infrequent, or emerging. These systems can instantly detect subtle variations and comprehend complex transaction patterns by employing advanced methodologies such as clustering, autoencoders, ensemble models, and sequential learning algorithms. The precision and scalability of detection have substantially improved as a result of the continuous advancements in hybrid models, deep learning architectures, and real-time analytics, despite the persistence of challenges such as inconsistent data and false positives.

REFERENCES:

[1] Adepoju, O., Wosowei, J., lawte, S., & Jaiman, H. (2019). Comparative evaluation of credit card fraud detection using machine learning techniques. 2019 Global Conference for Advancement in Technology (GCAT). https://doi.org/10.1109/gcat47503.2019.8978372

[2] Alenzi, H. Z., & Aljehane, N. O. (2020). Fraud detection in credit cards using logistic regression. International Journal of Advanced Computer Science and Applications, 11(12). https://doi.org/10.14569/ijacsa.2020.0111265

[3] Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis. 2017 International Conference on Computing Networking and Informatics (ICCNI). https://doi.org/10.1109/iccni.2017.8123782

[4] Bhanusri, A., Valli, K. R. S., Jyothi, P., Sai, G. V., & Rohith, R. (2020). Credit card fraud detection using Machine learning algorithms. Journal of Research in Humanities and Social Science, 8(2), 04-11.

[5] Credit card statistics. Shift Credit Card Processing. (2021, August 30). Retrieved from https://shiftprocessing.com/credit-card/

[6] Daly, L. (2021, October 27). Identity theft and credit card fraud statistics for 2021: The ascent. The Motley Fool. Retrieved from https://www.fool.com/theascent/research/identity theft-credit-card-fraud-statistics/

[7] Dheepa, V., & Dhanapal, R. (2012). Behavior based credit card fraud detection using support vector machines. ICTACT Journal on Soft Computing, 02(04), 391–397. https://doi.org/10.21917/ijsc.2012.0061

[8] Dighe, D., Patil, S., & Kokate, S. (2018). Detection of credit card fraud transactions using machine learning algorithms and Neural Networks: A comparative research. 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). https://doi.org/10.1109/iccubea.2018.8697799

[9] Domínguez-Almendros, S., Benítez-Parejo, N., & Gonzalez-Ramirez, A. R. (2011). Logistic regression models. Allergologia et immunopathologia, 39(5), 295-305.

[10] Gupta, A., Lohani, M. C., & Manchanda, M. (2021). Financial fraud detection using naive Bayes algorithm in highly imbalance data set. Journal of Discrete Mathematical Sciences and Cryptography, 24(5), 1559–1572. https://doi.org/10.1080/09720529.2021.1969733

[11] Itoo, F., Meenakshi, & Singh, S. (2020). Comparison and analysis of logistic regression, Naïve Bayes and Knn Machine Learning Algorithms for credit card fraud detection.

ISSN: 3139-3748|VOLUME-1, ISSUE-1| Inaugural Edition | January-June 2026
Paper ID: P11-010| Research Article
doi: https://doi.org/10.5281/zenodo.20675214

International Journal of Information Technology, 13(4), 1503–1511. https://doi.org/10.1007/s41870-020-00430-y

[12] Jain, Y., NamrataTiwari, S., & Jain, S. (2019). A comparative analysis of various credit card fraud detection techniques. International Journal of Recent Technology and Engineering, 7(5S2), 402-407

[13] Kiran, S., Guru, J., Kumar, R., Kumar, N., Katariya, D., & Sharma, M. (2018). Credit card fraud detection using Naïve Bayes model based and KNN classifier. International Journal Of Advance Research, Ideas And Innovations In Technology, 4(3).