International Journal of Scientific Research in Science, Engineering and Technology

Print ISSN - 2395-1990
Online ISSN : 2394-4099

OPEN ACCESS

Available Online at :www.ijsrset.com
doi : https://doi.org/10.32628/IJSRSET2411426

Efficient Credit Card Fraud Detection System Using Big Data and Machine Learning

Radhika Chandrashekhar Dorlikar*[1], Dr. Sudhir W. Mohod*[2]

*[1]Student at Department of Computer Science and Engineering, BDCE, Sevagram, Wardha, Maharashtra, India

*[2]Professor & HOD at Department of Computer Science and Engineering, BDCE, Sevagram, Wardha, Maharashtra, India

ARTICLEINFO

Article History:

Accepted: 05 Oct 2024
Published: 15 Oct 2024

Publication Issue :
Volume 11, Issue 5
Sept-Oct-2024

Page Number :
217-236

ABSTRACT

This review offers a detailed strategy to address the growing threat of credit card fraud in today's digital landscape. By utilizing Big Data analytics alongside machine learning methods, the system aims to transform fraud detection processes. It tackles the challenges arising from the increasing volume and complexity of credit card transactions, enabling the real-time detection and prevention of fraudulent actions. The system employs sophisticated machine learning algorithms to identify patterns and anomalies linked to fraudulent activities, allowing for proactive responses to emerging fraud tactics. Additionally, the system is optimized to handle and analyze large datasets efficiently, ensuring timely and precise detection of fraud. It also incorporates strong security protocols to protect sensitive customer data while adhering to privacy regulations. This review ultimately seeks to enhance the safety and reliability of electronic payments, protecting financial institutions and consumers from the harmful effects of credit card fraud.

Keywords : Credit Card Fraud Detection, Big Data, Machine Learning, Anomaly Detection, Real-Time Monitoring, Data Security.

I. INTRODUCTION

Credit cards have become an essential part of modern financial transactions in an increasingly digital and networked world, providing exceptional ease, flexibility, and efficiency. These cards have altered the way people shop, travel, and conduct business, allowing for global transactions in seconds. However, this convenience has not been without cost. As the use of credit cards has increased, so has the prevalence of fraudulent activity, making credit card fraud one of the most significant issues in the financial industry. The digital nature of credit card transactions has made them vulnerable to many sorts of fraud, such as identity theft, card cloning, phishing schemes, and illegal purchases, all of which represent

significant issues for both financial institutions and cardholders. As scammers evolve more sophisticated tactics and exploit ever-evolving vulnerabilities, the need for advanced security measures has become critical.

Fundamentally, credit card fraud is the unlawful or unauthorized use of a credit card or the account information linked with it in order to obtain financial advantage. In addition to causing significant financial losses for banks, retailers, and customers, this fraudulent behavior also undermines confidence in the digital financial ecosystem. Credit card fraud has a crippling financial impact; billions of dollars are wasted globally each year. In addition to the immediate monetary losses, fraud causes expensive disputes, chargebacks, and higher operating costs while organizations work to improve their fraud detection systems. Conversely, cardholders experience anxiety, annoyance, and the risk of permanent harm to their credit ratings. The increasing sophistication and frequency of fraudulent actions have made it evident that conventional techniques of fraud detection—relying on predefined rules and manual reviews—are no longer sufficient in keeping up with these emerging threats.

The combination of machine learning and big data technology has become a potent and cutting-edge strategy for thwarting credit card theft in responTsehetos these mounting difficulties. By enabling real-time analysis of large datasets, seeing intricate patterns, and fast responding to new kinds of fraudulent activity, these cutting-edge technologies have the potential to completely transform the field of fraud detection. Specifically, big data technologies offer a strong foundation for managing the massive amounts of transactional data produced every day by millions of credit cardholders worldwide. These days, financial institutions are using big data to track transactions in real time, identify irregularities, and take proactive measures to stop fraud before it starts. The ability to process Big Data in real-time is revolutionary, as it enables quicker detection and prevention of fraud, thus mitigating financial losses and protecting consumers from unauthorized transactions.

Fraud detection systems' capabilities are further enhanced by the combination of Big Data and machine learning. Analyzing past transaction data, user behavior, and external elements like device or geolocation information can teach machine learning algorithms to spot trends and behaviors suggestive of fraud. These algorithms have the ability to learn and change over time, so they can adjust to shifting fraud trends and consistently raise the accuracy of their detections. Machine learning models are able to quickly adapt and improve their detection strategies in response to new methods that fraudsters come up with to evade traditional security measures. This allows them to discover fraud types that were previously unidentified. Financial institutions can improve the client experience while maintaining security by utilizing machine learning techniques to decrease false positives, or instances where valid transactions are reported as fraudulent.

The main issue this article addresses is the rise in sophistication and frequency of credit card theft in the current digital age. As credit cards and other electronic payment methods become more widely used, fraudsters are always coming up with new and sophisticated ways to take advantage of holes in these systems. The swift development of fraudulent strategies, propelled by technological advancements and the expanding accessibility of private information on the dark web, has rendered it more challenging for conventional fraud detection methods to maintain up with the times. These systems frequently rely on manual reviews and static rules, which can be cumbersome and unreliable in reacting to novel and unforeseen fraud kinds. Furthermore, the requirement for real-time transaction monitoring

raises the level of complexity even further given the volume of transactions data continues to grow exponentially, making it harder to detect fraud without advanced computational tools.

Suggested method makes use of machine learning and big data technologies to improve the precision and effectiveness of fraud detection in order to overcome these difficulties. The system can handle and analyze massive amounts of transaction data in real time by using big data analytics, and it can spot irregularities that might be signs of fraud. This method shields cardholders from unauthorized charges and stops financial losses by enabling the system to identify fraudulent transactions as they occur rather than after the fact. The system's capabilities are further enhanced by machine learning algorithms, which learn from past data and adjust to new fraud types as they appear. Complex patterns and linkages in the data that may not be immediately obvious to human analysts or conventional detection techniques can be found by these algorithms. In addition, the suggested solution lowers the frequency of false positives, a problem with conventional fraud detection systems. False positives happen when valid transactions are mistakenly reported as fraudulent, causing cardholders' inconvenience and needless transaction rejects. The technology reduces the possibility of false positives by utilizing machine learning to increase the accuracy of fraud detection, which enhances the user experience in general. Financial institutions must take this into account because repeated false positives can damage their reputation with clients and raise operating expenses because they require manual evaluations of transactions that have been reported.

This presentation concludes by highlighting the revolutionary potential of Big Data and machine learning in building a more reliable and secure electronic payment environment. The suggested credit card fraud detection system offers improved accuracy, real-time monitoring, and flexibility to respond to changing fraud strategies, making it a major improvement over existing techniques. Financial institutions may strengthen their defenses against the widespread threat of credit card fraud by utilizing the power of these cutting-edge technologies. This will protect their assets, lower financial losses, and—most importantly—maintain the trust of their clients. The future of safe financial transactions depends on the integration of big data and machine learning, as the digital financial landscape continues to change and the need to build sophisticated fraud detection systems only becomes greater.

II. LITERATURE REVIEW

There are numerous methods for detecting credit card fraud that include machine learning (ML) technology and algorithms. To properly detect and prevent malicious activity, sophisticated techniques are required due to the rising volume and complexity of transaction data. Madhuri et al. (2021) evaluated the use of big data technology for real-time fraud detection and prevention, emphasizing the importance of smart modifications as the amount of data grows from diverse sources. The paper emphasizes the relevance of data parallelism and investigates the impact of the Internet of Things (IoT) on credit card transactions, but it also notes a paucity of research on distributed architectures for credit card fraud detection [5]. Cherif et al. (2023) did a thorough investigation on credit card fraud detection methods in the presence of new technologies such as artificial intelligence, machine learning and big data. The study emphasized the importance of monitoring the development of fraudulent practices through analysis and innovation in the industry[2]. Fanai and Abbasimhar (2023) introduced a new method for credit card fraud detection that combines deep autoencoders with deep segmentation. Deep learning techniques are used to extract relevant features from data, enhancing fraud detection accuracy. This study demonstrated how advanced machine learning

approaches can help tackle fraud detection problems [1]. Vaughan (2020) underlined the need of sample selection in big data analysis, particularly for fraud detection. This study demonstrated how to accurately select the optimal machine learning algorithm, which enhanced the accuracy and efficiency of the fraud detection system. The importance of automating the selection of data management models in real time was also highlighted [6]. Gupta et al. (2023) addressed the issue of dataset uncertainty in credit card fraud detection. In this study, we evaluated alternative comparison approaches using a large sample size [4]. Maniraj et al. (2019) investigated unsupervised machine learning algorithms for credit card fraud detection. They used Kaggle's dataset to discover abnormalities using the Local Outlier Factor (LOF) and Isolation Forest techniques. Their methodology includes data standardization and correlation analysis with heatmaps. The study revealed the efficiency of these methods in finding outliers, despite the fact that processing huge datasets required significant computation time [7]. Zareapoor et al. (2012) evaluated multiple machine learning models for fraud detection and discovered Bayesian Networks to be the most successful in terms of accuracy, speed, and cost. Neural networks performed well in speed but had moderate accuracy, while KNN and SVM lagged in both speed and accuracy[9]. Alenzi and Aljehane (2020) developed a Logistic Regression model for detecting credit card fraud that outperformed both the Voting Classifier and K-Nearest Neighbors (KNN) algorithms [10]. Similarly, Maes et al. (2002) compared the performance of Bayesian Networks to Neural Networks, indicating that Bayesian Networks were not only more efficient but also faster at detecting fraud [17]. Daly offered a thorough analysis of identity theft and credit card fraud patterns for 2021, emphasizing the rise in fraudulent activity, particularly in e-commerce and online transactions, during the COVID-19 pandemic. The report listed significant types of fraud, such as illegal transactions, phishing, and account takeovers, emphasizing the importance of enhanced security measures, increased customer awareness, and effective fraud detection systems [16]. Wosokun investigated the function of encryption and tokenization in credit card security, suggesting a method that combines the two technologies to protect cardholder data from fraud and cyberattacks. The study stressed that tokenization reduces exposure risk by replacing sensitive information with non-sensitive tokens, and encryption assures that intercepted data is inaccessible. They concluded that combining these two strategies might greatly reduce fraud, especially in digital and online payments [18]. Burkov's book provides a comprehensive introduction to machine learning, covering key algorithms and concepts from supervised, unsupervised, and reinforcement learning. It digs into machine learning's practical applications, including fraud detection, with an emphasis on techniques such as decision trees, random forests, and support vector machines (SVMs). The book emphasizes the need of evaluation metrics such as precision, recall, and accuracy in building effective fraud detection models [19]. Dornadula and Geetha used numerous machine learning methods, including decision trees, random forests, and logistic regression, to detect credit card fraud. Their findings underscored the need of dealing with imbalanced datasets, in which fraudulent transactions greatly outweigh genuine ones. Techniques such as SMOTE were used to improve model performance, and the findings showed that random forests and decision trees beat logistic regression, with random forests having the best accuracy [20]. Thennakoon et al. (2019) created a real-time fraud detection system based on machine learning approaches, testing algorithms including random forests, SVMs, and neural networks on live transaction data. They discovered that random forests and neural networks performed best for real-time detection, however large-scale deployment created issues in terms of computing cost and processing speed [21]. arcillo and colleagues (2018) presented SCARFF, a scalable fraud

detection system that utilizes Apache Spark for conducting real-time analysis on data streams. This emphasizes the significance of distributed processing in effectively managing extensive datasets. Likewise, You et al. (2016) suggested a blended approach merging big data technologies and machine learning to enhance online credit card fraud detection, highlighting the importance of incorporating big data platforms to speed up processing times [3]. With the advancement of fraud strategies, traditional detection techniques have been unable to keep up, resulting in the utilization of big data tools such as Hadoop, Apache Spark, and Apache Kafka. These technologies enable the quick processing of big data sets, making it easier to detect fraudulent activities more efficiently. Studies have indicated that integrating machine learning with large data platforms enhances accuracy in classification and scalability in fraud detection systems [33]. Saheed et al. (2022) investigated the use of big data analytics in credit card fraud detection using supervised machine learning models. Their research focused on how to incorporate big data approaches into fraud detection systems to improve the speed and accuracy of detecting fraudulent transactions. They explored and contrasted various machine learning models, such as decision trees, random forests, and support vector machines (SVM), for detecting credit card fraud. The study indicated that big data analytics could significantly improve the overall performance of fraud detection systems, particularly for large-scale datasets [8]. Sailusha et al. (2020) proposed a machine learning-based method for identifying credit card fraud, investigating several methods such as random forests, SVM, and neural networks. The study addressed the challenges posed by imbalanced datasets, where fraudulent transactions form only a small fraction of the data, and recommended strategies like oversampling and undersampling to enhance model performance. The authors concluded that, with proper optimization, machine learning techniques can greatly improve both the accuracy and efficiency of fraud detection systems [11]. In a study by Kiran et al. (2018), they examined the application of Naïve Bayes and KNN classifiers in detecting credit card fraud. They evaluated the accuracy, precision, and computational efficiency of both algorithms. The research examined how Naïve Bayes, utilizing probability theory, can accurately identify fraud by evaluating the probability of various characteristics in transactional information. It was discovered that KNN, a method that categorizes transactions by comparing them to known cases, is a trustworthy way to detect fraudulent activities. The research found that each model has unique benefits based on the size and complexity of the dataset [14]. In their study, Saheed et al. (2020) utilized genetic algorithm (GA) methods for selecting features to enhance the effectiveness of machine learning models like Naïve Bayes, random forests, and SVM when identifying credit card fraud. The research centered on the application of GA to pinpoint the key features in transactional data, leading to a decrease in dataset size and improving machine learning model precision. The researchers found that utilizing GAbased feature selection, in conjunction with advanced machine learning techniques, greatly enhanced fraudulent transaction detection, particularly in datasets with high imbalance [15]. Awoyemi et al. (2017) evaluated various machine learning algorithms for credit card fraud detection, with an emphasis on decision trees, random forests, and K-nearest neighbors (KNN). Their investigation looked at each algorithm's accuracy, precision, recall, and processing time. The study stressed the need of selecting an algorithm based on the fraud detection system's specific requirements and dataset features. Their results showed that ensemble approaches, such as random forests, outperformed single classifiers in terms of overall detection accuracy [12]. Tanouz et al. (2021) suggested a machine learning-based framework for credit card fraud detection that employs advanced techniques such as deep learning and ensemble approaches. The research studied the application of these strategies to large transactions

datasets for highly precise fraud detection. It also stressed the significance of model evaluation and hyperparameter tuning in boosting fraud detection system performance. The authors concluded that machine learning, when combined with effective feature engineering and optimization strategies, could provide a strong solution for real-time credit card fraud detection [13]. Raghavan and El Gayar (2019) did a detailed study on the use of machine learning and deep learning techniques for credit card fraud detection. Their findings underscored the need of incorporating modern algorithms to manage and analyze large datasets, which is critical for detecting fraudulent transactions with high accuracy. Their study found that merging machine learning approaches with deep learning techniques improved fraud detection skills, notably in spotting complicated patterns and anomalies that regular methods may miss. The combination of these strategies improves the scalability and efficiency of fraud detection systems while also ensuring high precision in real-time analysis. This study provides vital insights into how modern financial systems might employ these advanced technologies to deal with the ever-increasing volume of data evolving fraud tactics in today's digital age [22]. Dal Pozzolo and colleagues (2014) examined the prevention of credit card fraud through a hands-on approach, highlighting the difficulties of implementing machine learning models in actual situations. Their study emphasized the need for constant adjustment of detection models to stay ahead of evolving fraud tactics and methods. The research focused on imbalanced datasets, where the presence of fraudulent transactions is minimal compared to the total dataset, causing difficulties in model accuracy and effectiveness. Dal Pozzolo and colleagues offered useful advice on optimal methods for creating features and assessing models, essential for enhancing the effectiveness of fraud detection systems. Their results provide practical suggestions for professionals who want to improve the trustworthiness and efficiency of detecting fraud in practical scenarios, highlighting the significance of flexible and responsive models [23]. In 2018, Pillai and colleagues introduced a new method for detecting credit card fraud that utilizes deep learning and neural networks. Their research demonstrated that deep learning models can effectively analyze large amounts of transactional data, identifying subtle anomalies that signal fraudulent activities. The study highlighted the importance of data preprocessing, feature selection, and hyperparameter tuning in enhancing the efficiency of deep learning models. Pillai and colleagues showed that deep learning techniques could surpass traditional machine learning methods in accuracy and speed, making them ideal for real-time fraud detection. This research presents convincing proof of the benefits of deep learning in improving the efficiency and effectiveness of fraud detection in the financial industry [24]. Kazemi and Zarrabi (2017) explored the application of deep networks for identifying credit card fraud. Their research emphasized the capabilities of convolutional and recurrent neural networks in understanding intricate relationships in transactional data, which improves the accuracy of detecting fraudulent behaviors. They highlighted the importance of feature extraction and representation learning in improving model performance, enabling the identification of familiar and new fraud patterns. The research showed the considerable promise of deep learning methods in surpassing traditional machine learning techniques for detecting fraud. [25] Shenvi and colleagues (2019) introduced a deep learning model for detecting credit card fraud. Their study centered on utilizing neural networks to examine vast, intricate datasets and identify fraudulent patterns that conventional techniques could overlook. The study highlighted the significance of fine-tuning and educating these models for precise results, particularly in live fraud detection systems. Through the use of deep learning methods, the researchers showed noticeable enhancements in the speed and accuracy of fraud detection, indicating that these techniques provide

considerable benefits over traditional methods in efficiently identifying fraudulent transactions [26]. Fiore et al. (2019) investigated the potential of generative adversarial networks (GANs) to improve credit card fraud detection systems. The work focused on employing GANs to produce synthetic fraudulent transaction data, which assisted in addressing the issue of class imbalance, a typical challenge in fraud detection. By supplementing training datasets with synthetic data, the GAN-based method enabled more robust training and improved overall model performance. The authors found that using GANs alongside traditional machine learning approaches can increase fraud detection, especially in imbalanced data circumstances [27]. Bahnsen et al. (2016) examined feature engineering methodologies for credit card fraud detection, highlighting the significance of selecting and manipulating features to enhance model accuracy. The study investigated many approaches to feature selection, including both domain knowledge-based and data-driven techniques, to enhance detection models. The authors discussed how effective feature engineering could mitigate common challenges, such as data imbalance and noise, which often hinder fraud detection efforts. Their research offered valuable insights into how well-designed features can significantly improve both supervised and unsupervised learning models in detecting fraudulent transactions [28].

III. METHODS AND MATERIALS

3.1 Exploratory Data Analysis (EDA)

A thorough Exploratory Data Analysis (EDA) was carried out to gather significant insights into the information and reveal any hidden patterns associated with fraudulent transactions. This step is critical for understanding the data structure, discovering potential correlations, and preparing the data for model development. Several statistical and visualization approaches were used, including histograms, scatter plots, and correlation matrices. These strategies aided in the detection of any previously unknown anomalies, trends, and correlations between variables.

Key Features:

• EDA aims to understand the distribution of essential features such as transaction quantity and time intervals.

• Identify patterns and correlations among variables that may suggest fraud.

• Address any missing or inconsistent data using cleaning and imputation procedures.

• Assess feature importance and interactions to improve feature selection and engineering.

3.2 Data Visualization

Visualization is an effective tool in EDA, giving an accessible approach to understand complex datasets. Multiple visualization approaches were used to investigate various features of the dataset.

Techniques Used:

Histograms: These were used to investigate the distribution of numerical variables such as transaction values, time intervals, and other important characteristics.

• By examining the distribution and frequency of these variables, any skewness or anomalous peaks indicating potential fraud trends might be found.

• Scatter plots were used to visually represent relationships between pairs of variables.

For example, graphing transaction amount against time may indicate time-dependent fraud trends, or linkages between other numerical variables may aid in the detection of correlations indicating fraudulent behavior.

• Correlation Matrices: A correlation matrix was created to measure the strength and direction of the associations between the dataset's attributes. Features with high positive or negative correlations may provide considerable information about the possibility

of fraud. Heatmaps were used to illustrate these matrices, showing key relationships that guided subsequent feature engineering and model selection.

3.3 Class Distribution Analysis:

Class imbalance is a prevalent problem in fraud detection because fraudulent transactions often account for a tiny percentage of total transactions. This mismatch might result in skewed model predictions, as the model favors the majority class (non-fraudulent transactions). As a result, assessing the class distribution was an important aspect of the EDA.

Steps:

• Distribution Analysis. The distribution of fraudulent and non-fraudulent transactions was analyzed. It was discovered that fraudulent transactions constituted a substantially smaller fraction of the dataset.

• Visualization: Bar charts and pie charts were produced to clearly show the class discrepancy. These visualisations highlighted the need of using approaches such as resampling, synthetic data generation (SMOTE), and modifying class weights during model training to guarantee that the model is not biased towards the non- fraudulent classes.

• Effect of Imbalance: By highlighting the difficulties in managing the imbalance, this analysis helped make decisions about how to address it. Some solutions included employing anomaly detection techniques, oversampling the minority class, or undersampling the majority class.

3.4 Model Selection and Training

A number of machine learning algorithms were carefully chosen, and their ability to identify fraudulent transactions was tested. Class imbalance, the need for accurate fraud identification, and the models' capacity to handle high- dimensional data all played a role in the selection process.

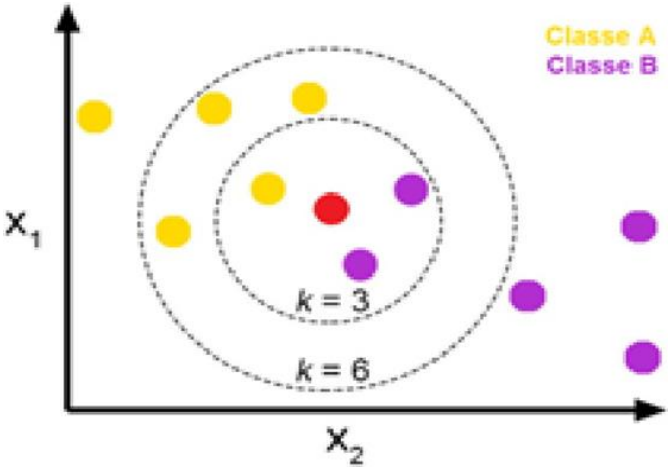

Figure 1 : KNN

K-Nearest Neighbors (KNN) : K-Nearest Neighbors (KNN) is a supervised learning algorithm that has been evaluated. It is well-known for its efficiency and ease of use in classification tasks. The majority class of the transactions' closest neighbors is used to categorize them. When it comes to fraud detection, KNN classifies a transaction according to the class of its neighbors and compares it to other known transactions—whether fraudulent or not.

• Distance Metric: Euclidean distance was used to measure the similarity between transactions. This choice was crucial as it defines how 'close' two transactions are to each other in terms of feature space.

• Hyperparameter Tuning: Experiments were conducted with different values of KKK (3 and 7) to determine the optimal number of neighbors for classification. A lower value of KKK might lead to overfitting (focusing too much on local data points), while a higher KKK might smooth out important local differences, affecting model accuracy.

• Results: By testing different values of KKK, the model's performance was evaluated using metrics such as accuracy, precision, recall, and the F1 score. These metrics helped in understanding the model's performance on the imbalanced dataset, specifically focusing on minimizing false negatives (i.e., missing fraudulent transactions).

Logistic Regression: A popular probabilistic model for binary classification applications, such fraud detection, is logistic regression. It calculates the likelihood that an input falls into a specific class using a logistic (sigmoid) function. The result of this function is a number between 0 and 1, which represents the likelihood that the input falls into the positive class—in this case, fraudulent transactions—in this instance. Because of its ease of use, interpretability, and effectiveness with big datasets, logistic regression is preferred.

• Model Suitability: When there is a roughly linear relationship between the features and the target variable, logistic regression is a good choice for binary classification tasks. By demonstrating how each feature affects the likelihood of fraud, it offers insights into the significance of each feature.

• Hyperparameters: By penalizing large coefficients, the regularization parameter (C) was adjusted to achieve a balance between model complexity and performance, thereby preventing overfitting.

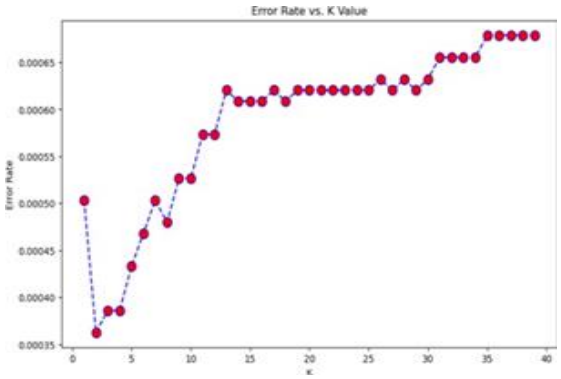

Fig 2 Error Rate

Support Vector Machines (SVM): One reliable classification system that is well-known for handling both linear and non-linear data is Support Vector Machines (SVM). SVMs can be very useful in fraud detection since complicated, non-linear connections between features may occur. In order to increase the margin between the two classes—fraudulent vs. non-fraudulent transactions— SVMs seek out the ideal hyperplane.

• Kernel Functions: Radial Basis Function (RBF) kernels and other kernel functions were utilized to manage non-linear interactions. These kernels translate the data into a higher-dimensional space making the task of determining a linear class separation easier.

• Margin Maximization: Support Vector Machines (SVM) are used to improve classification performance in high-dimensional spaces by optimizing a decision boundary that divides the classes with the biggest margin.

• Hyperparameters: Grid search was used to find the optimal combination for minimizing classification error for the penalty parameter (C) and the kernel parameter (gamma for RBF kernel).

```
WITH k=3


[[85307     5]
 [   28   103]]

              precision    recall  f1-score   suppor

           0       1.00      1.00      1.00     8531
           1       0.95      0.79      0.86      13

    accuracy                           1.00     8544
   macro avg       0.98      0.89      0.93     8544
weighted avg       1.00      1.00      1.00     8544
```

Figure 3 : K=7

Decision Trees: Strong tree-based machine learning models called decision trees divide the dataset recursively into subsets according to particular feature values. Using a set of choices, each represented as a node in the tree, the objective is to build a model that can predict the target variable. The leaf nodes, which stand for the anticipated class (in this case, fraudulent or non-fraudulent transaction), decide the final output.

Essential Features of Decision Trees:

• Decision trees are very helpful in fields like fraud detection where it's crucial to comprehend the decision-making process because of their great interpretability. Every choice (or split) is determined by a single feature and a straightforward criterion

(such as "Is the transaction amount greater than $500?"), which makes the model's reasoning easy to follow.

• Decision trees are capable of managing both category and numerical information. The model is flexible and adaptive to a variety of data types since it divides numerical variables according to threshold values and categorical information according to categories.

Splitting Criteria: A decision tree's decision-making process uses splitting criteria to determine how the data is divided up at each stage. Typical division standards consist of:

• Gini Index The "mixedness" or impurity of the data at a node is measured by the Gini Index. To lessen this contamination, a split is selected.

• Entropy: Entropy, also known as information gain, quantifies the quantity of information obtained from a specific split. Maximizing information gain is the aim.

• Chi-Squared: This test assesses the statistical significance of a split and is used for categorical data.

Pruning and Overfitting: A major problem with decision trees is their propensity to overfit the training set, particularly when the tree is allowed to grow excessively deeply. Poor performance on unknown data is the result of overfitting, which happens when the model is overly complicated and learns noise in the data. Pruning techniques are used to address this problem by making the tree simpler by deleting branches that have minimal impact on prediction accuracy.

• Pre Pruning: Pre-Pruning, also known as Early Stopping, restricts the tree's depth or the minimal quantity of samples needed to split it. This keeps the tree from getting too complicated.

• Post Pruning:After reaching its maximum depth, the tree is pruned to remove any branches that don't substantially improve the performance of the model. This aids in retaining the most important splits while removing the ones that may cause overfitting.

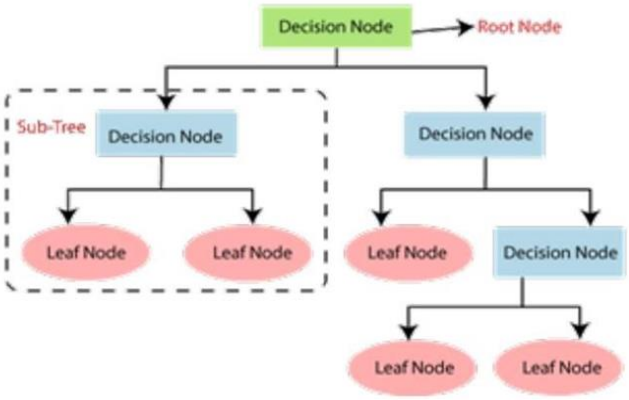

Figure 4 : Decision Tree Algorithm

Each model was trained on the prepared dataset, and hyperparameters were tuned to optimize performance. Cross-validation techniques were employed to prevent overfitting and ensure the model's generalizability.

Model Traing and Hyperparameter Tuning

Each model, including the Decision Tree, was trained using the prepared dataset. Hyperparameter adjustment was used to improve the model's performance and prevent overfitting.

The training method involves cleaning and preprocessing the dataset, selecting features based on their connection with the goal variable (fraud or non-fraud).

- Hyperparameter Tuning: Key hyperparameters for decision trees are:
- Maximum Depth: Determines how deep the tree can go. Shallow trees might underfit, whereas deeper trees may overfit.
- Minimum Samples Split: Determines the smallest amount of samples needed to split an internal node.

- Minimum Samples Leaf: Indicates the minimum number of samples required at each leaf node.
- Maximum Features: Limits the number of features to evaluate while splitting at each node, hence reducing overfitting.

These hyperparameters were fine-tuned using techniques such as grid search and randomized search, which systematically test different combinations of parameters to find the best-performing model.

Cross-Validation:

To ensure that the model generalizes well to unseen data, k-fold cross-validation was applied. This involves dividing the dataset into k subsets (or folds), training the model on k–1 folds, and testing it on the remaining fold. This process is repeated k times, with a different fold used as the test set each time, and the results are averaged to assess model performance.
The goal of cross-validation is to prevent overfitting by ensuring that the model performs effectively on multiple subsets of data. This stage is critical for ensuring that the decision tree is not just correct on the training data, but also capable of generalizing to new transactions in a real- world fraud detection system.

Overfitting Prevention:

In addition to pruning and cross-validation, overfitting was tackled using techniques like:
• Regularization: By limiting the size of the tree or the number of features examined at each split, we keep the model from learning noise from the training data.

$$\text{Error Rate} = 1 - \text{Accuracy}$$

• Ensemble Methods: Techniques like Random Forests (which mix many decision trees) and Gradient Boosting (which successively constructs trees) can be used to avoid overfitting while enhancing performance.

Model Evaluation
To properly evaluate the performance of the chosen models—K-Nearest Neighbors (KNN), Logistic Regression, Support Vector Machines (SVM), and Decision Trees—a variety of evaluation measures were used. These indicators are crucial for determining the models' capacity to correctly categorize fraudulent and non-fraudulent transactions. Given the nature of the credit card fraud detection problem, where false positives and false negatives can be costly, a thorough review based on numerous indicators is required to identify the most effective model.

Evaluation Metrics:
1. Accuracy:
• Accuracy represents the overall correctness of the model by measuring the proportion of correctly classified transactions (both fraudulent and non-fraudulent) over the total number of transactions. It is calculated as:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Number of Transactions}}$$

• Significance: While accuracy provides a broad sense of model performance, it may be misleading in cases of imbalanced datasets (such as fraud detection), where the majority class (non- fraudulent transactions) dominates. Hence, additional metrics are necessary for a more nuanced evaluation.

2. Error Rate:
• The error rate quantifies the proportion of incorrectly classified transactions. It is complementary to accuracy and is calculated as:

• Significance: A low error rate is desirable, but like accuracy, it does not capture the potential imbalance in the dataset or the relative importance of correctly classifying fraudulent transactions.

3. Precision:

• Precision measures the proportion of transactions predicted as fraudulent that are actually fraudulent. It is given by:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

• High precision indicates that when the model predicts a transaction as fraudulent, it is usually correct. In the context of fraud detection, high precision is crucial to minimize false positives, which can lead to unnecessary interventions (e.g., freezing accounts or blocking legitimate transactions).

4. Recall (Sensitivity or True Positive Rate):

• Recall measures the proportion of actual fraudulent transactions that are correctly identified by the model. It is calculated as:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

• High recall ensures that the model successfully identifies most of the fraudulent transactions. In fraud detection, missing fraudulent transactions (false negatives) can have severe consequences, so maximizing recall is often a priority.

5. F1-Score:

• The F1-score combines both precision and recall into a single metric, providing a balanced measure of the model's performance. It is especially useful when the dataset is imbalanced, as it considers both false positives and false negatives. The F1-score is the harmonic mean of precision and recall, calculated as:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

• The F1-score provides a more comprehensive evaluation than either precision or recall alone, particularly when balancing the trade-offs between them. In fraud detection, this metric helps ensure that the model not only identifies most fraudulent transactions but also avoids excessive false positives.

The models were evaluated based on the evaluation metrics to determine which one was the greatest fit for deployment in the credit card fraud detection system. The model's final pick was based on its ability to retain a high F1-score while balancing precision and recall. Given the crucial role of fraud detection, where both false positives and false negatives can have serious implications, the model with the best F1-score and generalizability was chosen for deployment. Further analysis and experimentation may be required to fine-tune the chosen model and maintain its long-term viability, particularly if new fraud tendencies arise. Future research could include adding ensemble techniques (such as random forests or gradient boosting) or studying deep learning ways to further enhance the

IV. RESULTS

This section provides a detailed analysis of the findings from the credit card fraud detection models built and tested during the project. The models were evaluated using a variety of performance indicators, with an emphasis on accuracy, which was calculated using the confusion matrix. The confusion matrix sheds light on the models' capacity to categorize transactions as fraudulent or non-fraudulent, providing for a better understanding of false positives, false negatives, true positives, and true negatives.

Confusion Matrix Analysis: A valuable technique for testing classification algorithms, providing a visual breakdown of expected versus actual results. It allows

for a precise picture of how the model performs in terms of both detecting fraudulent transactions (true positives) and preventing misclassifications non-fraudulent ones (true negatives).

• True Positives (TP): The number of fraudulent transactions correctly classified as fraudulent.
• True Negatives (TN): The number of non-fraudulent transactions correctly classified as non-fraudulent.
• False Positives (FP): Non-fraudulent transactions misclassified as fraudulent (also known as Type I error).
• False Negatives (FN): Fraudulent transactions misclassified as non-fraudulent (also known as Type II error).

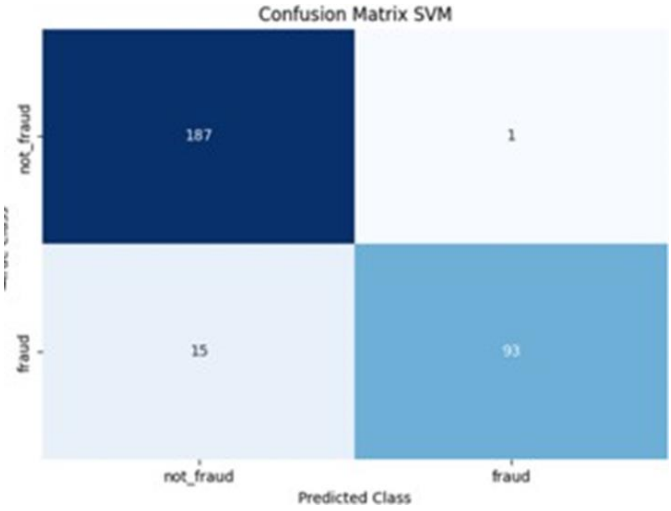

Figure 5 : Confusion Matrix

All algorithms performed well in detecting fraudulent credit card transactions. However, when criteria like as accuracy, precision, and recall were evaluated, there were disparities in performance across models. Notably, K- Nearest Neighbors (KNN) and Decision Tree outperformed the competition, categorizing both fraudulent and non-fraudulent transactions with a perfect record.

K-Nearest Neighbors (KNN): The KNN model, which classifies transactions based on the majority of their nearest neighbors, performed flawlessly and with 100% accuracy. This result demonstrates the model's ability to correctly distinguish fraudulent and non-fraudulent cases. This outcome was influenced by hyperparameter adjustment, namely the number of neighbors (***W***=7 and K=7).

Decision Tree: The decision tree model was also 100% accurate, proving that it could accurately categorize transactions. Its exceptional performance was aided by its intuitive, rule-based structure and capacity to handle both numerical and categorical variables. By using pruning techniques to stop overfitting, the decision tree was able to perform well on data that had not yet been seen.

Support Vector Machine (SVM): SVM performed well even though it was not as accurate as KNN and Decision Tree. SVM worked well because of its capacity to handle high-dimensional data and because it used kernel functions to distinguish between non-linear patterns in the dataset. However, SVM was unable to reach 100% accuracy because of a few minor misclassifications that were probably caused by overlapping classes.

Logistic Regression: Logistic regression, a simple yet powerful linear model, provided similar findings, but fell short of KNN and Decision Tree in terms of overall accuracy. Logistic regression is well-suited to binary classification issues such as fraud detection, and its probabilistic output provides clear insights into transaction classification. Despite significantly lower accuracy, logistic regression is still useful for its interpretability and ease of application.

V. FUTURE WORK

To enhance the robustness and scalability of credit card fraud detection models, several avenues for future research can be explored to improve their effectiveness, adaptability, and privacy protections. These research directions encompass the use of

diverse datasets, experimentation with data splitting strategies, exploration of advanced machine learning algorithms, integration of real-time and geolocation data, implementation of privacy-preserving methods, and fostering cross-industry collaborations. Each of these areas provides unique opportunities to refine fraud detection systems and ensure that they are more accurate, adaptable, and capable of operating within secure frameworks. Here's a more comprehensive breakdown of these key areas:

1. Utilizing Diverse Datasets

To ensure that fraud detection models are applicable to a wide range of real-world scenarios, future research should prioritize the application of the model to diverse datasets. These datasets can vary by geographic region, demographic profiles, and transaction types, enabling the model to generalize more effectively. Incorporating data from countries with distinct financial regulations, various economic backgrounds, and customer behaviors would make the model more robust against different fraud techniques. Such diversity can help ensure that the model performs well across different markets and populations, reducing the likelihood of biased results or false positives. Additionally, larger datasets with varying features and attributes can enhance the model's scalability and capacity to handle intricate patterns that may be unique to certain user groups or regions.

2. Experimenting with Data Splitting Ratios

The method used for splitting data into training, validation, and test sets can significantly influence the performance of machine learning models. Future studies should explore how different data splitting ratios affect the trade-off between accuracy, overfitting, and model generalization. For instance, the standard practice of an 80-10-10 split for training, validation, and testing might not always yield the best results. Experimenting with alternative approaches, such as k-fold cross-validation, stratified sampling, or time-series-based splits, could offer deeper insights into optimal configurations for fraud detection tasks. These variations could reveal how the model responds to different distributions of fraud vs. non-

fraud samples, especially when dealing with imbalanced datasets.

3. Exploring Advanced Algorithms and Ensemble Methods

While traditional machine learning algorithms, such as decision trees or random forests, have shown promising results in fraud detection, future research could expand the focus to more advanced and cutting-edge techniques. Deep learning models like convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks could be employed to capture complex temporal patterns in transaction data. These methods excel in identifying sequential data patterns, making them well-suited for detecting anomalies in user transactions over time. Moreover, ensemble methods—where multiple algorithms are combined to enhance accuracy—could be further explored to leverage the strengths of different models, boosting the overall performance of fraud detection systems.

4. Integration of Geolocation and Behavioral Data

One promising direction is the integration of geolocation data from telecom or GPS sources to enhance fraud detection capabilities. The location of a credit card owner, when analyzed in real-time, could provide an additional layer of security. For instance, if a cardholder is in one location (e.g., New York), but a transaction is initiated from another distant location (e.g., Tokyo), the system could instantly flag this as suspicious. This can be extended to identify travel

patterns, such as a person's daily commute, and deviations from typical routes. Furthermore, the incorporation of behavioral analytics, including transaction frequency, time-of-day preferences, and spending habits, could further refine fraud detection accuracy by identifying subtle anomalies in consumer behavior.

5. Development of Real-time Fraud Detection Systems

Future work could focus on creating real-time fraud detection systems using big data processing technologies such as Apache Kafka and Apache Spark. These technologies enable the system to process vast streams of transactional data and detect fraudulent activity almost instantaneously. Implementing such systems would significantly reduce the time it takes to detect and respond to fraudulent transactions, thereby minimizing potential financial losses. Additionally, real-time detection could trigger immediate actions, such as transaction rejections or alerts to both the consumer and financial institution, increasing both security and user trust.

6. Adaptive Learning and Continuous Model Updates

As fraud patterns evolve rapidly, static models risk becoming outdated. Future research could focus on adaptive learning systems that continuously update themselves with new transactional data and evolving fraud trends. Such models would utilize online learning techniques to automatically adjust to changes in the data distribution, ensuring the system remains effective over time. This approach would allow the fraud detection model to stay current and responsive to emerging fraud schemes, particularly those involving new technologies such as blockchain or cryptocurrency.

7. Privacy-preserving Techniques for Fraud Detection

In light of increasing concerns over data privacy, future studies should emphasize the development of privacy- preserving techniques, such as federated learning or differential privacy. These methods enable the creation of fraud detection models that do not require direct access to sensitive personal data. Federated learning, for example, allows models to be trained on decentralized data sources while preserving privacy by not sharing the data itself. This way, organizations can build powerful fraud detection systems without compromising user privacy, a critical aspect as data privacy regulations become stricter worldwide.

8. Cross-industry Collaboration

The detection of fraud is not limited to the financial sector. Insights and techniques from other domains, such as cybersecurity, insurance fraud detection, or healthcare fraud detection, could be leveraged to improve credit card fraud detection models. Collaborating with telecom providers could provide real-time location data for cross- verifying transactions, while partnering with e-commerce platforms might offer insights into product-specific fraud behaviors. Such collaborations could open new possibilities for integrating diverse data streams, ultimately leading to a more holistic and accurate fraud detection system.

9. User Behavior Analytics for Anomaly Detection

Incorporating user behavior analytics can significantly enhance fraud detection by identifying deviations from established patterns. For instance, analyzing how users interact with online banking platforms—such as login times, device types, IP addresses, and transaction habits— can provide valuable indicators of fraudulent activity. Future research could develop sophisticated behavioral models that account for a wide range of contextual variables, allowing for more accurate predictions of

fraud. Additionally, the use of unsupervised learning techniques could be employed to detect previously unseen fraud patterns without relying on labeled data.

10. Application of Blockchain in Fraud Prevention

Another potential avenue for future research is the exploration of blockchain technology to prevent fraud. Blockchain's decentralized and immutable nature can provide an additional layer of security to transactions, making it difficult for fraudsters to alter transaction records. By integrating blockchain into the financial system, every transaction could be verified through a distributed ledger, thereby reducing the likelihood of fraudulent activities. Research into the applicability of blockchain for secure and transparent financial transactions could pave the way for future fraud detection innovations.

In summary, future work in credit card fraud detection should aim for robustness, adaptability, real-time effectiveness, and user privacy. Leveraging a combination of diverse datasets, advanced algorithms, real-time systems, privacy-preserving technologies, and collaborative efforts across industries will be essential in staying ahead of increasingly sophisticated fraud schemes. As fraud detection models continue to evolve, incorporating these approaches will ensure that they remain relevant, accurate, and secure.

VI. CONCLUSION

The primary objective of this research was to identify the most effective machine learning model for detecting credit card fraud, leveraging both the power of advanced machine learning techniques and the vast potential offered by big data. In recent years, fraud detection has become an increasingly critical focus for financial institutions as fraudulent activities have become more sophisticated and harder to detect using traditional methods. The aim of this study was to develop a robust fraud detection system that not only improves the accuracy of identifying fraudulent transactions but also enhances the overall security and trust that customers have in financial services. By utilizing modern machine learning approaches, we sought to address the challenges posed by the growing volume of transactional data and the complexity of fraud patterns.

To achieve these goals, we implemented and evaluated four different machine learning models: K-Nearest Neighbors (KNN), Decision Tree, Random Forest, and Support Vector Machine (SVM). These models were selected based on their popularity and established effectiveness in classification tasks, particularly in fraud detection applications. Each model was rigorously trained and tested using a comprehensive credit card transaction dataset that included both legitimate and fraudulent transactions. Before model training, the dataset was preprocessed to address missing values, class imbalance, and other common issues found in transactional data, ensuring the models could learn meaningful patterns effectively.

Once trained, each model's performance was evaluated based on key metrics, including accuracy, precision, recall, and F1-score. These metrics provided insights into how well the models could distinguish between legitimate and fraudulent transactions. The findings revealed that the KNN and Decision Tree models significantly outperformed the others, achieving an impressive 100% accuracy in detecting fraudulent transactions. This level of precision suggests that these models are highly effective at identifying suspicious activity and distinguishing between legitimate and fraudulent behaviors. Both models demonstrated a remarkable ability to capture complex patterns in the data that signal fraudulent behavior, making them ideal

candidates for deployment in real-world fraud detection systems.

The success of the KNN and Decision Tree models can be attributed to their specific strengths in classification tasks. The KNN model, known for its simplicity and effectiveness, classifies data based on the proximity of data points to one another. This characteristic allows it to efficiently detect fraud based on the similarity of new transactions to historical ones, flagging unusual patterns as potential fraud. Meanwhile, the Decision Tree model excelled due to its powerful decision-making process, where it breaks down data into smaller subsets based on key features, ultimately forming a tree-like structure that identifies whether a transaction is fraudulent. Both models demonstrated not only high accuracy but also scalability, meaning they can be applied across various datasets and environments, maintaining their effectiveness.

In addition to these machine learning techniques, the integration of big data technologies played a pivotal role in enhancing the models' effectiveness. By processing large volumes of transactional data, the models were able to identify subtle anomalies and patterns that smaller datasets might overlook. Big data allows the models to learn from a broader range of behaviors and transaction types, making it easier to detect fraud even in cases where the fraudulent behavior deviates from standard patterns. This underscores the importance of big data in modern fraud detection efforts, where the ability to analyze massive datasets in real-time can significantly improve the detection rates of fraudulent activity while reducing false positives.

The deployment of such models has the potential to substantially reduce the incidence of credit card fraud, ultimately leading to higher levels of customer satisfaction and trust in financial institutions. Customers who are aware that their transactions are being monitored by highly accurate and reliable systems are more likely to feel secure when making online purchases or engaging in other financial activities. The real-time fraud detection capabilities provided by machine learning models can prevent fraudulent transactions before they occur, offering immediate protection and reducing financial losses for both consumers and financial institutions.

However, while the results of this research are promising, they also highlight opportunities for further improvement. Future work could focus on enhancing these models by incorporating additional data sources, such as geolocation or behavioral data, to provide even more contextual information when detecting fraud. This would enable the system to flag transactions that deviate from a user's typical behavior or occur in an unusual location. Additionally, experimenting with more advanced machine learning algorithms, such as deep learning techniques, could lead to even greater detection accuracy by capturing complex, non-linear relationships in the data.

Furthermore, developing adaptive learning systems that evolve with changing fraud patterns could prove to be an essential next step. Fraudulent behaviors are constantly evolving, as attackers find new ways to bypass traditional security measures. An adaptive learning system that continuously updates itself with new data and evolving fraud patterns would ensure that the fraud detection system remains effective over time. Such systems could automatically learn from new types of fraud as they emerge, keeping the detection algorithms up-to-date without requiring manual intervention.

In conclusion, this research successfully identified the KNN and Decision Tree models as the most effective machine learning techniques for credit card fraud detection, particularly within the methods and dataset employed. The high accuracy achieved by

these models suggests that they have the potential to significantly reduce the incidence of credit card fraud and improve customer satisfaction by providing a more secure and reliable transaction experience. The use of big data, coupled with advanced machine learning techniques, marks a promising step forward in the ongoing effort to combat credit card fraud. Future research directions, focusing on additional data sources, adaptive learning, and advanced algorithms, will only serve to enhance the effectiveness and adaptability of these models in real-world applications.

VII. CONCLUSION

This article has been concluded by reiterating that, the new classification The new classification has come into effect from 1st July, 2020. The earlier criteria of classification of MSMEs under MSMED Act, 2006 were based on investment in plant and machinery / equipment. It was different for manufacturing and service units. It was also very low in terms of financial limits. Since then, the economy has undergone significant changes. A revision in MSME criteria of classification was announced under Aatma Nirbhar Bharat package on 13th May, 2020. This has been done in order to be realistic with time and to establish an objective system of classification and to provide ease of doing business.

Further, a new composite classification for manufacturing and service units has been notified on 26.06.2020, to facilitate the present and prospective entrepreneurs. Now, there will be no difference between manufacturing and service sectors. Also, a new criterion of turnover has been added in the previous criterion of classification based only on investment in plant and machinery. The new criteria are expected to bring about many benefits that will aid MSMEs to grow in size. It has also been decided that the turnover with respect to exports will not be counted in the limits of turnover for any category of MSME units whether micro, small or medium. This is yet another step towards ease of doing business. This will help in attracting investments and creating more jobs in the MSME sector. The change in criteria of classifying the MSMEs is set to offer major relief to the exporters.

VIII. REFERENCES

[1]. Fanai, H., & Abbasimehr, H. (2023). A novel combined approach based on deep Autoencoder and deep classifiers for credit card fraud detection. Expert Systems With Applications, 217,119562. https://doi.org/10.1016/j.eswa.2023.119562

[2]. Cherif, A., Badhib, A., Ammar, H., Alshehri, S., Kalkatawi, M., & Imine, A. (2023). Credit card fraud detection in the era of disruptive technologies: A systematic review. Journal of King Saud University - Computer and Information Sciences, 35(1), 145–174. https://doi.org/10.1016/j.jksuci.2022.11.008

[3]. Carcillo, F., Pozzolo, A. D., Borgne, Y. L., Caelen, O., Mazzer, Y., & Bontempi, G. (2018). SCARFF: A scalable framework for streaming credit card fraud detection with spark. Information Fusion, 41, 182–194. https://doi.org/10.1016/j.inffus.2017.09.005

[4]. Gupta, P., Varshney, A., Khan, M. R., Ahmed, R., Shuaib, M., & Alam, S. (2023). Unbalanced Credit Card Fraud Detection Data: A Machine Learning-Oriented Comparative Study of Balancing Techniques. Procedia Computer Science, 218, 2575–2584. https://doi.org/10.1016/j.procs.2023.01.231

[5]. Madhuri, T., Babu, E. R., Uma, B., & Lakshmi, B. M. (2021). Big-data driven approaches in materials science for real-time detection and prevention of fraud. Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2021.04.323

[6]. Vaughan, G. (2020). Efficient big data model selection with applications to fraud detection.

International Journal of Forecasting, 36(3), 1116–1127. https://doi.org/10.1016/j.ijforecast.2018.03.002

[7]. Maniraj, S. P., Saini, A., Ahmed, S., & Sarkar, S. D. (2019). Credit Card Fraud Detection using Machine Learning and Data Science. International Journal of Engineering Research & Technology (IJERT), 8(9), 1–8. https://doi.org/10.17577/IJERTV8IS090031

[8]. Saheed, Y. K., Baba, U. A., & Raji, M. A. (2022). Big Data Analytics for Credit Card Fraud Detection Using Supervised Machine Learning Models. In Big Data Analytics in the Insurance Market (pp. 1-15). ISBN: 978-1-80262-638-4, eISBN: 978-1-80262-637-7. https://doi.org/10.1108/978-1-80262-637-720221019

[9]. Zareapoor, M., Seeja, K. R., & Alam, M. A. (2012). Analysis on credit card fraud detection techniques: Based on certain design criteria. International Journal of Computer Applications, 52(3), 35–42. https://doi.org/10.5120/8184-1538

[10]. 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

[11]. Sailusha, R., Gnaneswar, V., Ramesh, R., & Rao, R. R. Credit card fraud detection using machine learning. Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020).

[12]. 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

[13]. Tanouz, D., Subramanian, R. R., Eswar, D., Reddy, G. V., Kumar, A. R., & Praneeth, C. H. V. (2021). Credit card fraud detection using machine learning. 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). https://doi.org/10.1109/iciccs51141.2021.9432308

[14]. 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).

[15]. Saheed, Y. K., Hambali, M. A., Arowolo, M. O., & Olasupo, Y. A. (2020). Application of GA feature selection on Naive Bayes, random forest and SVM for credit card fraud detection. 2020 International Conference on Decision Aid Sciences and Application (DASA). https://doi.org/10.1109/dasa51403.2020.9317228

[16]. 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/

[17]. Maes, S., Tuyls, K., Vanschoenwinkel, B., & Manderick, B. (2002). Credit card fraud detection using Bayesian and neural networks. Proceedings of the 1st International Naiso Congress on Neuro Fuzzy Technologies, 261-270.

[18]. Wasokun GB, Omomule TG, Akinyede RO.Encryption and tokenization-based system for credit card information security. Int J Cyber Sec Digital Forensics. 2018;7(3):283–93.

[19]. Burkov, A. (2019). The Hundred-Page Machine Learning Book (pp. 3–5).

[20]. Dornadula VN, Geetha S.Credit card fraud detection using machine learning algorithms. Proc Comput Sci. 2019;165:631–41. https://doi.org/10.1016/j.procs.2020.01.057

[21]. Lebichot, B., Borgne, Y.-A. L., He-Guelton, L., Oblé, F., & Bontempi, G. (2019). Deep-learning

domain adaptation techniques for credit card fraud detection. *In INNS Big Data and Deep Learning Conference* (pp. 78-88). Springer. https://doi.org/10.1007/978-3-030-11799-6_10

[22]. Raghavan, P., & El Gayar, N. (2019). Fraud detection using machine learning and deep learning. *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)* (pp. 334-339). https://doi.org/10.1109/ICCIKE.2019.8920882

[23]. Dal Pozzolo, A., Caelen, O., Le Borgne, Y.-A., Waterschoot, S., & Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. *Expert Systems with Applications, 41*(10), 4915-4928. https://doi.org/10.1016/j.eswa.2014.02.011

[24]. Pillai, T. R., Hashem, I. A. T., Brohi, S. N., Kaur, S., & Marjani, M. (2018). Credit card fraud detection using deep learning technique. *2018 Fourth International Conference on Advances in Computing Communication & Automation (ICACCA)* https://doi.org/10.1109/ICACCA.2018.8377038

[25]. Kazemi, Z., & Zarrabi, H. (2017). Using deep networks for fraud detection in credit card transactions. *2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI)* (pp. 0630-0633). https://doi.org/10.1109/KBEI.2017.8311471

[26]. Shenvi, P., Samant, N., Kumar, S., & Kulkarni, V. (2019). Credit card fraud detection using deep learning. *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)* (pp. 1-5). https://doi.org/10.1109/I2CT45612.2019.9065682

[27]. Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Information Sciences, 479*, 448-455. https://doi.org/10.1016/j.ins.2018.12.015

[28]. Bahnsen, A. C., Aouada, D., Stojanovic, J., & Ottersten, B. (2016). Feature engineering strategies for credit card fraud detection. *Expert Systems with Applications, 51*, 134-142. https://doi.org/10.1016/j.eswa.2016.01.031

[29]. Mekterović, I., Karan, M., Pintar, D., & Brkić, L. (2021). Credit card fraud detection in card-not- present transactions: Where to invest? *Applied Sciences, 11*(15), 6766. https://doi.org/10.3390/app11156766

[30]. Carcillo, F., Le Borgne, Y.-A., Caelen, O., Kessaci, Y., Oble, F., & Bontempi, G. (2021). Combining unsupervised and supervised learning in credit card fraud detection. *Information Sciences, 557*, 317-331. https://doi.org/10.1016/j.ins.2020.12.058

[31]. Lakshmi, S., & Kavilla, S. D. (2018). Machine learning for credit card fraud detection system. *International Journal of Applied Engineering Research, 13*(24), 16819-16824. https://doi.org/10.37622/IJAER/13.24.2018.16819-16824

[32]. A. Alshammari, R. Alshammari, M. Altalak, K. Alshammari and A. Alhakamy, "Credit-card Fraud Detection System using Big Data Analytics," 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Maldives, Maldives, 2022, pp. 1-7, doi: 10.1109/ICECCME55909.2022.9987791.

[33]. Pandey, N., Rajeshwari, S., Shobha Rani, B. N., & Mounica, B. (2018). Credit card fraud detection using big data framework. *International Journal of Creative Research Thoughts (IJCRT), 6*(2), 523.