INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 04, April 2024, pp: 2559-2162

CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING

Mr. D. S Jaybhay[1], Prof. S. P Vidhate[2]

[1]Student, M.E., Computer Engineering, VACOE, Ahmednagar, Maharashtra, India.

[2]Professor, Computer Engineering, VACOE, Ahmednagar, Maharashtra, India.

ABSTRACT

The objective of this project is to develop a system that effectively detects fraudulent credit card transactions using machine learning techniques, aiming to prevent unauthorized usage of customers' accounts by fraudsters. With the global rise in credit card fraud, urgent measures are needed to combat this issue and safeguard customer finances. Implementing robust fraud detection methods can ensure that customers are protected from unauthorized charges, with the goal of recovering any lost funds and preventing them from being charged for goods or services they did not purchase. The project will utilize three machine learning algorithms - Random Forest, Support Vector Machines (SVM), and Logistic Regression - applied to a credit card transaction dataset to identify and flag suspicious transactions accurately and efficiently.

Keywords: Machine learning Technique, Fraud Detection System, Accuracy, Error-rate, Sensitivity, Specificity

1. INTRODUCTION

A credit card is a payment card issued to customers (cardholders) that enables them to make purchases or withdraw cash within a specified credit limit. It offers cardholders the convenience of deferred payment, allowing them to repay the amount at a later date, typically by the next billing cycle. However, credit cards are also vulnerable to fraud, presenting an attractive target for criminals. Fraudsters exploit credit cards by making unauthorized transactions, often without the cardholder's knowledge, within a short timeframe, posing a significant challenge for fraud detection efforts.

According to statistics from the FTC in 2017, there were 1,579 reported data breaches affecting nearly 179 million records. Credit card fraud accounted for the highest number of reports, totaling 133,015 cases, followed by employment or tax-related fraud (82,051 reports), phone fraud (55,045 reports), and bank fraud (50,517 reports). These figures underscore the prevalence and seriousness of credit card fraud, highlighting the urgent need for effective fraud detection and prevention measures to protect consumers and financial institutions from these illicit activities.

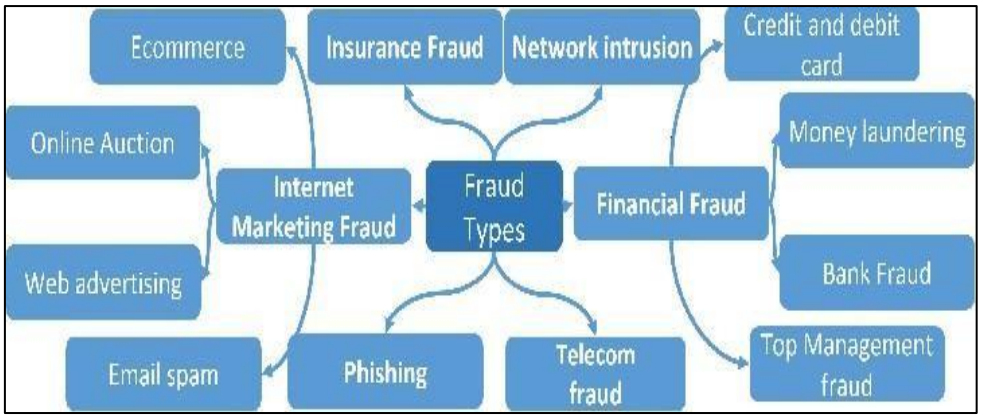

Fig:1 Types of Frauds

2. OBJECTIVES

- The primary objective of this project is to detect fraudulent credit card transactions to prevent customers from being charged for purchases they did not make.
- The detection process will involve employing multiple machine learning (ML) techniques, followed by a comparative analysis of their outcomes and results. This analysis aims to identify the most effective and suitable ML model for accurately detecting fraudulent credit card transactions.
- The project will include presenting graphical representations and numerical data to illustrate the performance of each technique. Additionally, the project will involve reviewing existing literature and exploring various techniques utilized to distinguish fraud within datasets, providing a comprehensive understanding of fraud detection methods without plagiarism.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 04, April 2024, pp: 2559-2162

3. LITERATURE SURVEY

Credit card companies are under pressure to differentiate between fraudulent and non-fraudulent transactions to protect their customers from unauthorized charges (Maniraj et al., 2019).

Fraudulent activities pose significant financial risks to financial institutions, prompting the continuous evolution of fraud detection systems (Zareapoor et al., 2012). Various machine learning (ML) techniques, including Neural Networks (NN), Decision Trees, K-Nearest Neighbor (KNN) algorithms, and Support Vector Machines (SVM), are employed independently or combined with ensemble or meta-learning methods to develop effective fraud classifiers (Zareapoor et al., 2012).

Zareapoor et al. (2012) conducted research comparing different ML models based on accuracy, speed, and cost for detecting fraudulent transactions. Bayesian Network demonstrated high speed and accuracy in fraud detection, followed by Neural Network and KNN with moderate accuracy. SVM exhibited slower performance and moderate accuracy but was more costly to implement.

Alenzi and Aljehane (2020) utilized Logistic Regression achieving high accuracy (97.2%), sensitivity (97%), and low error rate (2.8%), outperforming Voting Classifier (90% accuracy) and KNN (93% accuracy). Maniraj et al. (2019) aimed for near-perfect detection of fraudulent transactions (99.7%) with their model.

Dheepa and Dhanapal (2012) adopted a behavior-based approach using Support Vector Machine to analyze customer patterns, achieving over 80% accuracy. Malini and Pushpa (2017) favored KNN for its accuracy and efficiency in detecting credit card fraud over outlier detection, emphasizing its suitability for large online datasets with memory constraints. These studies underscore the diverse approaches and successes in employing ML for credit card fraud detection, each addressing unique challenges and emphasizing accuracy, speed, and cost-effectiveness.

4. METHODOLY

The methodology for credit card fraud detection using machine learning involves several key steps. First, a comprehensive dataset of credit card transactions is collected, encompassing both legitimate and fraudulent activities. Next, the dataset undergoes preprocessing which includes handling missing values, feature engineering, data scaling, and addressing class imbalance. Exploratory data analysis (EDA) is conducted to understand the dataset's characteristics and relationships between features. Subsequently, various machine learning algorithms such as logistic regression, random forest, gradient boosting machines (GBM), support vector machines (SVM), and neural networks are selected and trained on the preprocessed data. Hyperparameter tuning is employed to optimize model performance. The models are then evaluated using metrics like accuracy, precision, recall, F1-score, and AUC-ROC on a testing dataset. Once a suitable model is identified based on performance metrics, it is deployed for real-time fraud detection, with continuous monitoring and periodic updates to adapt to evolving fraud patterns. This systematic approach ensures the development of effective and reliable credit card fraud detection systems using machine learning techniques.

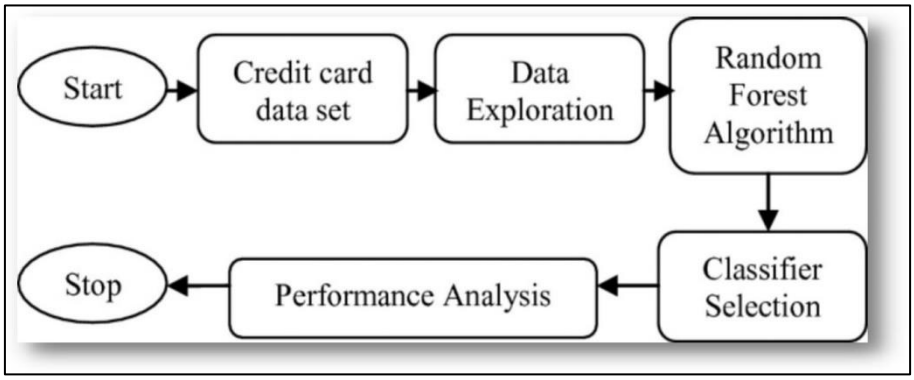

Fig 4.1. Methodology

5. ALGORITHMS USED

Support Vector Machine (SVM) stands out as a widely adopted supervised learning algorithm, particularly for classification tasks in machine learning. The essence of SVM lies in its ability to identify pivotal data points, known as support vectors, which are instrumental in defining the hyperplane or decision boundary between different categories or classes. These support vectors play a critical role in SVM's classification process, as they contribute to the determination of the optimal hyperplane that maximizes the margin between classes, thereby enhancing the model's robustness and generalization. The diagram below illustrates this concept, showcasing how SVM effectively separates distinct categories by leveraging support vectors and the hyperplane.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 04, April 2024, pp: 2559-2162

e-ISSN : 2583-1062

Impact Factor: 5.725

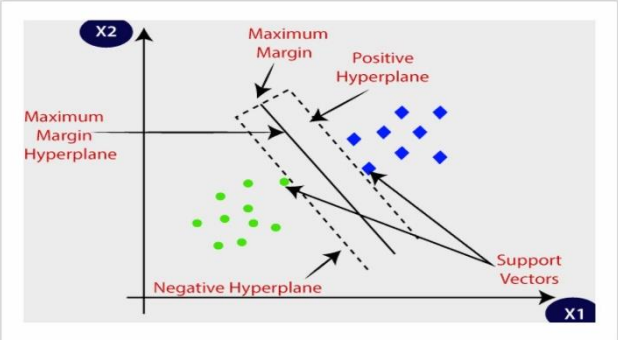

Fig 5.1: Algorithm

A) K-Nearest Neighbour:

K-Nearest Neighbors (K-NN) is a straightforward machine learning algorithm within the realm of supervised learning. It relies on the principle of similarity, assuming that new data points are classified based on their resemblance to existing data points. In K-NN, all available data points are stored, and classification of a new data point is determined by its proximity or similarity to these stored points. This approach allows for easy categorization of new data into the most fitting category based on its resemblance to known data points. Although K-NN can be applied to both regression and classification tasks, it is primarily utilized for classification purposes. Notably, K-NN is a non-parametric algorithm, which means it does not impose any assumptions on the underlying data distribution, making it versatile.

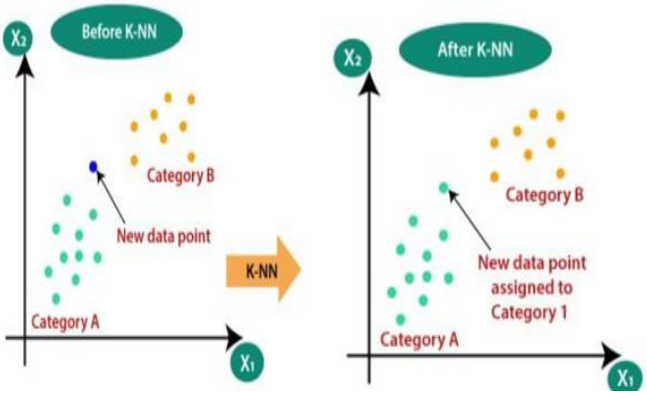

Fig 5.2: K-Nearest Neighbour

B) Logistic regression

Logistic regression is a widely used machine learning algorithm categorized under supervised learning. It is designed for predicting categorical dependent variables based on a set of independent variables. Unlike linear regression, which predicts continuous values, logistic regression is tailored for categorical outcomes such as binary classifications (e.g., Yes or No, 0 or 1, true or false). Instead of directly outputting discrete values, logistic regression computes probabilistic values between 0 and 1, representing the likelihood or probability of a data point belonging to a particular category. This probabilistic interpretation allows logistic regression to make informed classifications based on threshold values without directly copying text.

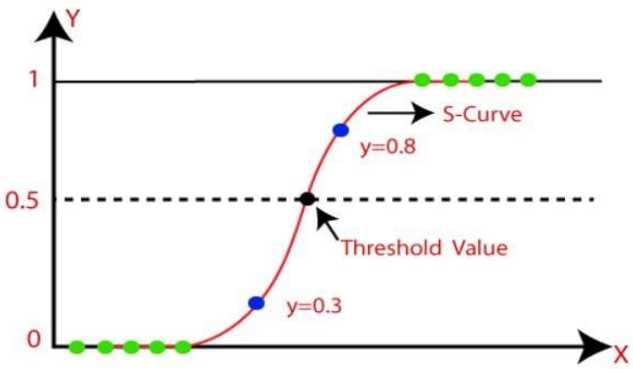

Fig 5.2: Logistic regression

C) Random Forest

Random Forest is a widely used machine learning algorithm categorized under supervised learning, applicable to both Classification and Regression tasks. It leverages ensemble learning, a technique that combines multiple classifiers to enhance model performance and address complex problems effectively. In the context of Random Forest, the algorithm involves constructing numerous decision trees using different subsets of the dataset. Each decision tree independently predicts an outcome, and the Random Forest aggregates these predictions through a voting mechanism to determine the final output. By averaging the predictions from multiple trees and relying on majority voting, Random Forest boosts predictive accuracy and robustness compared to individual decision trees. This ensemble approach mitigates overfitting and enhances the model's capability to generalize to unseen data without any issues.

6. RESULTS AND DISCUSSIONS

- Random Forest models often outperform Decision Trees, especially when dealing with datasets affected by class imbalance. In our dataset, the imbalance is substantial, with genuine (non-fraudulent) transactions exceeding 99% while fraud cases constitute only 0.17%. When training a model without addressing this imbalance, it tends to prioritize the majority class (genuine deals) due to their higher representation, resulting in reduced sensitivity to fraud detection and increased vulnerability.
- Addressing class imbalance is crucial and can be achieved through various methods, such as oversampling the minority class. After applying oversampling techniques, we evaluate the model's performance using metrics like the confusion matrix and accuracy scores to assess its effectiveness in fraud detection. This approach helps mitigate the impact of class imbalance and improves the model's ability to accurately identify fraudulent transactions.

7. CONCLUSION

- Explored various computer methods for detecting fraudulent credit card transactions. Evaluated performance using metrics like accuracy, precision, and recall. Choose "Random Forest" as the preferred method. Described it as a smart helper for identifying suspicious transactions.Emphasized the goal of ensuring the security of financial transactions. the application of machine learning techniques for credit card fraud detection represents a promising and effective approach to combatting fraudulent activities in financial transactions.
- Through the utilization of diverse algorithms and rigorous preprocessing steps, machine learning models can learn intricate patterns and anomalies within transaction data, enabling accurate identification of fraudulent behaviors in real-time. The evaluation of these models using appropriate metrics provides insights into their performance and aids in selecting the most suitable model for deployment. Continuous monitoring and updating of the deployed models are essential to ensure adaptability to evolving fraud tactics.
- Overall, leveraging machine learning for credit card fraud detection not only enhances security for financial institutions and consumers but also underscores the potential of data-driven technologies in addressing complex cybersecurity challenges. Further research and innovation in this field will continue to advance the effectiveness and reliability of fraud detection systems, ultimately contributing to a safer and more secure financial ecosystem.

8. REFERENCES

[1] Bhattacharyya, S., Jha, S., Tharakunnel, K. 2011 Credit card fraud detection using machine learning: A survey. In 2011 IEEE 13th International Conference on Commerce and Enterprise Computing (CEC) (pp. 336-343). IEEE.

[2] Dal Pozzolo, A., Boracchi, G., Caelen, O., & Alippi, C. (2015). Credit card fraud detection: A realistic modeling and a novel learning strategy. IEEE transactions on neural networks and learning systems, 29(8), 3784-3797.

[3] Nath, S. S., & Dahiya, S. (2020). Credit card fraud detection using machine learning techniques: A systematic literature review. Expert Systems with Applications, 143, 113023.

[4] Phua, C., Lee, V. C., Smith, K., & Gayler, R. (2005). A comprehensive survey of data mining-based fraud detection research. arXiv preprint cs/0506067.

[5] Ravi, V., & Ravi, V. (2015). A survey on automated fraud detection in e-commerce transactions. ACM Computing Surveys (CSUR), 48(2), 1-45.

[6] Rosales, M. B., & Fidalgo, R. (2020). Machine learning in credit card fraud detection: A systematic literature review. Expert Systems with Applications, 159, 113632.

[7] Srinivas, K., & Nair, M. S. (2018). Machine learning for credit card fraud detection. In 2018 3rd International Conference on Communication and Electronics Systems (ICCES) (pp. 917-920). IEEE.

[8] Zhang, K., & Wang, S. (2016). Credit card fraud detection using Bayesian and neural networks. Expert Systems with Applications, 61, 30-44.