CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING ALGORITHM

[1]Mr. Ajinkya Atik, UG Student SKN Sinhgad Institute of Technology & Science, Lonavala, Maharashtra, India, ajinkyaatik@gmail.com

[2]Mr. Moiz Ansari, UG Student SKN Sinhgad Institute of Technology & Science, Lonavala, Maharashtra, India, moiz.ans11@gmail.com

[3]Mr. Sushilkumar Dhok, UG Student SKN Sinhgad Institute of Technology & Science, Lonavala, Maharashtra, India, sushilkumardhok8@gmail.com

[4]Mr. Swapnil Barge, UG Student SKN Sinhgad Institute of Technology & Science, Lonavala, Maharashtra, India, swapnilbarge9696@gmail.com

[5]Mrs. Sareeka Deore, Asst. Professor SKN Sinhgad Institute of Technology & Science, Lonavala, Maharashtra, India, stdeore.sknsits@sinhgad.edu

Abstract: The rapid growth of online transactions has led to an increase in credit card fraud, posing significant challenges to financial institutions and consumers. In this study, we propose a robust credit card fraud detection system utilizing machine learning algorithms, specifically Support Vector Machine (SVM) and Random Forest. The dataset used in this research consists of a comprehensive collection of credit card transactions, including both legitimate and fraudulent activities. Feature engineering techniques are employed to extract relevant information from the dataset, such as transaction amount, location, time, and type. we implement SVM and Random Forest classifiers to train and evaluate the model's performance in detecting fraudulent transactions. SVM is known for its effectiveness in handling high-dimensional data and finding optimal decision boundaries, while Random Forest excels in handling large datasets with high variability. the evaluation of the proposed models is conducted using standard metrics such as precision, recall, and F1-score. Additionally, we employ techniques such as cross-validation to ensure the robustness of the models and avoid overfitting. experimental results demonstrate the effectiveness of both SVM and Random Forest in accurately identifying fraudulent credit card transactions. Furthermore, ensemble techniques such as bagging and boosting could potentially enhance the performance of the classifiers further. the proposed credit card fraud detection system provides a practical solution for financial institutions and e-commerce platforms to mitigate the risks associated with fraudulent activities. By leveraging machine learning algorithms such as SVM and Random Forest, it offers an efficient and scalable approach to safeguarding financial transactions in the digital era.

Keywords: ***SVM Algorithm, Kernel, Detection, Fraud, Social Impact, Finances, Random Forest Algorithm.***

I. INTRODUCTION

Credit card fraud scrub [3] as much as 5 of the world's GDP (Gross Domestic Product.) every year. Combating credit card fraud using AI is to detect the suspicious activities. Combating credit card fraud typically requires most entities that complete financial transactions to keep thorough records of their clients' accounts and activities. If they come across any information that appears to be suspicious, they are required to report it to the government for further investigation [5]. In this Transaction records is check to detect credit card fraud activity if the suspicious data is detected. Here we use Artificial Intelligence and Machine Learning Algorithm to detect the suspicious activities and solve it by training the data of that activity [3]. We are going to use supervised and unsupervised algorithm techniques. Credit card fraud detection is a crucial system designed to identify and prevent unauthorized or fraudulent use of credit cards. As digital transactions become more prevalent, the risk of fraudulent activities, such as unauthorized transactions and identity theft, has increased. The primary objective is to distinguish between legitimate and unauthorized transactions in real-time or post-transaction analysis. This involves utilizing technological tools, algorithms, and methodologies to spot suspicious patterns,

 DOI : 10.35291/2454-9150.2024.0195

anomalies, or deviations from typical card usage behaviour. Key components include machine learning, data analytics, pattern recognition, and anomaly detection [2]. These systems analyse historical transaction data to train models, recognize fraudulent patterns, and conduct real-time monitoring. They also consider behavioural analysis, geolocation, device analysis, and collaboration among institutions to collectively combat fraud. The continuous evolution and integration of advanced technologies such as AI and big data analysis are essential in the ongoing battle against credit card fraud. The primary goal is to protect consumers and financial institutions by reducing fraudulent transactions while minimizing disruptions to legitimate cardholders [1].

II. BACKGROUND STUDY

Project Scope: [6] Collect a large dataset of credit card transactions (real or synthetic). Perform data cleaning to handle missing or inconsistent data. Compare and evaluate various machine learning algorithms suitable for fraud detection Split the dataset into training and testing sets for model evaluation. Evaluate models using appropriate metrics Develop a user-friendly interface for the fraud detection system. Implement encryption and secure communication protocols to protect sensitive data. User and Characteristics: Varying demographics and credit behaviours. Ability to interpret data patterns and anomalies. Understanding of financial regulations and laws. Assumptions and Dependencies: User must require the Python. User has to install the Spyder on his pc. User has to login to the system.

III. LITERATURE SURVEY

Paper Name: Dataset shift quantification for credit card fraud detection

Author: Yvan Lucas1,2, Pierre-Edouard Portier1, Lea Laporte ´ 1, Sylvie Calabretto1, Liyun He-Guelton3, Frederic Oble3 and Michael Granitzer2

Abstract: Machine learning and data mining techniques have been used extensively in order to detect credit card frauds. How ever purchase behaviour and fraudster strategies may change over time. This phenomenon is named dataset shift [1] or concept drift in the domain of fraud detection [2]. In this paper, we present a method to quantify day-by-day the dataset shift in our face-to-face credit card transactions dataset (card holder located in the shop). In practice, we classify the days against each other and measure the efficiency of the classification. The more efficient the classification, the more different the buying behaviour between two days, and vice versa. Therefore, we obtain a distance matrix characterizing the dataset shift. After an agglomerative clustering of the distance matrix, we observe that the dataset shift pattern matches the calendar events for this time period (holidays, week-ends, etc). We then incorporate this dataset shift knowledge in the credit card fraud detection task as a new feature. This leads to a small improvement of the detection.

Paper Name: Credit Card Fraud Detection Using Machine Learning

Author: D. Tanouz, R Raja Subramanian, D. Eswar, G V Parameswara Reddy, A. Ranjith kumar, CH V N M praneeth

Abstract: Credit card is the commonly used payment mode in the recent years. As the technology is developing, the number of fraud cases are also increasing and finally poses the need to develop a fraud detection algorithm to accurately find and eradicate the fraudulent activities. This research work proposes different machine learning based classification algorithms such as logistic regression, random forest, and Naive Bayes for handling the heavily imbalanced dataset. Finally, this research work will calculate the accuracy, precision, recall, f1 score, confusion matrix, and Roc-auc score.

Paper Name: Credit Card Fraud Detection Using Deep Learning

Author: Anu Maria Babu, Dr. Anju Pratap

Abstract: This paper discusses a method in the fraud detection interface area. The approach proposed is to use imbalanced highly skewed transactional data and a convolutional network for the detection of frauds. The dataset used here is the machine learning Kaggle dataset for credit card fraud detection that contains highly skewed data. The evaluated features are 1 for fraud and 0 for non-fraud class. The analysis of fraud detection was an important tool in banking sectors. Nowadays, the artificial neural network has become the least successful method for credit card fraud detection. The system currently used to detect fraud is plagued by misclassifications and highly false positives. In such situations here this research paper uses the in cooperation of convolutional neural network layers in an attempt to build a model for detecting credit card fraud that gives us a high level of accuracy. The goal is to predict fraud under 300 epochs, the current approach can be classified accurately at 99.62 [7]

Paper Name: A Novel Approach for Credit Card Fraud Detection using Decision Tree and Random Forest Algorithms

Author: Dileep M R, Navaneeth A V, Abhishek M

Abstract: In the world of finance, as the technology grown, new systems of business making came into picture. Credit card system is one among them. But because of lot of loop holes in this system, lot of problems are aroused in this system in the method of credit card scams. Due to this the industry and customers who are using credit cards are facing a huge loss. There is a deficiency of investigation lessons on examining practical credit card figures in arrears to privacy issues. In the manuscript an attempt has been made for finding the frauds in the credit card business by using the

 DOI : 10.35291/2454-9150.2024.0195

algorithms which adopted machine learning techniques. In this regard, two algorithms are used viz Fraud Detection in credit card using Decision Tree and Fraud Detection using Random Forest. The efficiency of the model can be decided by using some public data as sample. Then, an actual world credit card facts group from a financial institution is examined. Along with this, some clatter is supplemented to the data samples to auxiliary check the sturdiness of the systems. The significance of the methods used in the paper is the first method constructs a tree against the activities performed by the user and using this tree scams will be suspected. In the second method a user activity-based forest will have constructed and using this forest an attempt will be made in identifying the suspect. The investigational outcomes absolutely show that the mainstream elective technique attains decent precision degrees in sensing scam circumstances in credit cards.

IV. METHODOLOGY

Description of the Dataset Used for Training and Testing: The dataset used in this study comprises [4] transactional data collected from credit card transactions. It includes features such as transaction amount, time, type of transaction (e.g., purchase, cash withdrawal), and other relevant attributes. The dataset is labelled, with each transaction classified as either legitimate or fraudulent.

Preprocessing Steps: [6] Data Cleaning: Any missing or inconsistent data points are addressed. This may involve imputation of missing values or removal of incomplete records. Normalization: [4] Features are scaled to a standard range to ensure that they contribute equally to the SVM model. Common normalization techniques include min-max scaling or z-score normalization. Feature Selection: Relevant features are selected to train the SVM model. This step may involve statistical analysis or domain knowledge to identify the most informative attributes for fraud detection. Explanation of SVM Algorithm and its Parameters: Support Vector Machine (SVM) is a supervised learning algorithm used for classification tasks. It works by finding the hyperplane that best separates data points belonging to different classes.

Parameters of SVM include: [5]

Kernel: Determines the type of hyperplane used to separate the data. Common choices include linear, polynomial, and radial basis function (RBF) kernels. Regularization Parameter (C): Controls the trade-off between maximizing the margin and minimizing the classification error. A higher value of C allows for more complex decision boundaries. Kernel Parameters: If using non-linear kernels such as polynomial or RBF, parameters like degree (for polynomial) and gamma (for RBF) need to be tuned. Cross-Validation and Performance Evaluation Metrics: Cross-validation is employed to assess the generalization performance of the SVM model. Common techniques include k-fold cross-validation, where the dataset is divided into k subsets, and the model is trained and tested k times. Performance evaluation metrics include accuracy, precision, recall, F1-score, and area under the receiver operating characteristic (ROC) curve. These metrics provide insights into the model's ability to correctly classify both legitimate and fraudulent transactions.

Implementation [1]

Details of Implementation of SVM for Credit Card Fraud Detection: The SVM model is implemented using a suitable programming language such as Python, along with libraries like scikit-learn for machine learning tasks. The implementation involves loading the dataset, preprocessing the data, training the SVM model, and evaluating its performance. Description of Programming Language and Libraries Used: Python is chosen as the programming language due to its popularity and extensive libraries for data analysis and machine learning. Libraries such as scikit-learn are utilized for implementing SVM and other necessary preprocessing and evaluation tasks. Training Process and Parameter Tuning: The dataset is split into training and testing sets, typically with a ratio such as 70:30 or 80:20. The SVM model is trained on the training set using various combinations of parameters, and its performance is evaluated on the testing set. Techniques like grid search or randomized search are employed for parameter tuning, where different combinations of hyperparameters are tested to find the optimal configuration that maximizes the model's performance. This methodology ensures a systematic approach to building and evaluating an SVM model for credit card fraud detection, aiming to achieve high accuracy and robustness in identifying fraudulent transactions while minimizing false positives.

V. WORKING

Here's a breakdown of how credit card fraud detection works in distinct points: Data Analysis and Machine Learning Models: Utilizes historical transaction data to train algorithms and machine learning models. Algorithms distinguish patterns associated with fraudulent activities through supervised and unsupervised learning techniques. [2] Real-Time Monitoring: Constantly monitors live transactions for any irregularities or suspicious activities. Analyses transaction parameters such as amounts, locations, times, and typical purchasing behaviour. Behavioural Analysis: Establishes a baseline of normal cardholder behaviour to identify deviations or anomalies that might indicate fraudulent transactions. Compares ongoing transactions with a cardholder's typical spending patterns.

Geolocation and Device Verification: Validates the transaction location against the cardholder's usual locations. Examines the device used for the transaction to detect inconsistencies, such as unfamiliar devices or locations. Alerts and Intervention: Triggers alerts or interventions

 DOI : 10.35291/2454-9150.2024.0195

when suspicious transactions are detected. Human intervention or additional verification steps are initiated for potentially fraudulent transactions. Adaptive System Improvement: Constantly evolves and adapts to new fraud tactics by integrating advanced technologies and learning from new data. Regularly updates and improves the fraud detection system to stay ahead of evolving fraudulent behaviours. Credit card fraud detection works by amalgamating these methods and technologies to swiftly and accurately differentiate between legitimate and fraudulent transactions [6].

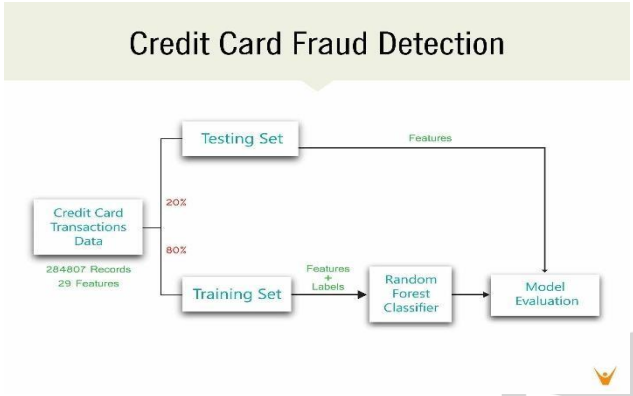

Figure 1: Random Forest Classifier using features

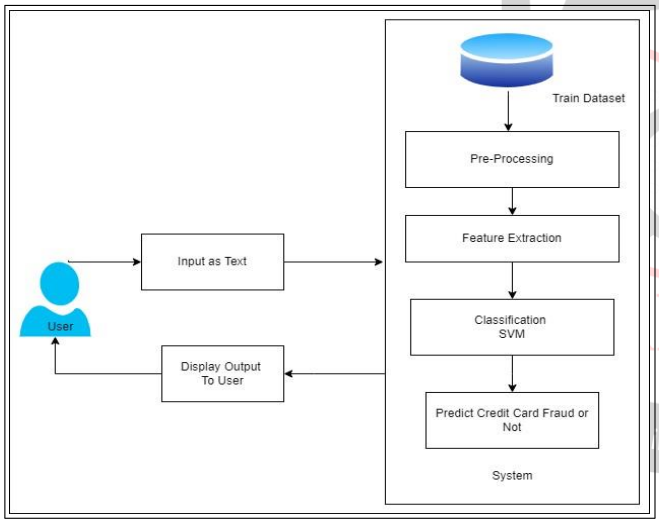

Figure 2:System Architecture

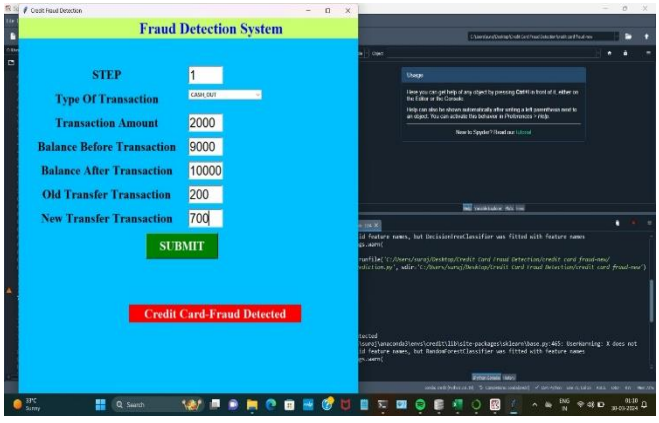

Figure 3: Final Result (Fraud detected)

VI. MATH

The Model which we represented is in two forms mathematical and programmable. The mathematical form contains all the calculations which we performed to give an output. Firstly, we calculated the Confusion Matrix (TP, FP, FN, TN). Secondly, we calculated the Bias and Variance with respect to its High and Low Form. The Calculations is performed below.

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

Table 1: Matrix Table

Accuracy: Accuracy is used to measure the performance of the model. It is the ratio of Total correct instances to the total instances.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Precision: Precision is a measure of how accurate a model's positive predictions are. It is defined as the ratio of true positive predictions to the total number of positive predictions made by the model.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall: Recall measures the effectiveness of a classification model in identifying all relevant instances from a dataset. It is the ratio of the number of true positive (TP) instances to the sum of true positive and false negative (FN) instances.

$$\text{Recall} = \frac{TP}{TP+FN}$$

VII. DISCUSSION

In this study, our objective was to create a machine learning model that could effectively detect fraudulent credit card transactions in a dataset containing both legitimate and fraudulent activities. The significance of this task cannot be emphasized enough, given the ongoing challenges posed by credit card fraud to financial institutions and consumers. Our approach involved several key stages. Initially, we gathered a comprehensive dataset that included a mix of legitimate and fraudulent transactions. This dataset served as the basis for training and assessing our machine learning models. Data preprocessing was crucial in preparing the data for analysis. We conducted data cleaning to address

 DOI : 10.35291/2454-9150.2024.0195

inconsistencies or missing values, and we performed feature engineering to extract pertinent information from the raw transaction data. Furthermore, we standardized the features to ensure equal contribution from all variables to the model's learning process.

Regarding machine learning algorithms, we explored a variety of options such as logistic regression, decision trees, random forests, support vector machines (SVM), and neural networks. Each algorithm has its own strengths and weaknesses, and our goal was to determine the most appropriate approach for our specific fraud detection task. We assessed the performance of each model using a range of metrics, including accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC-ROC). Our findings indicated significant differences in the performance of the various machine learning models. While some algorithms achieved high accuracy rates, others excelled in terms of precision or recall. It was crucial to note the trade-offs between minimizing false positives and false negatives. For example, certain models demonstrated high precision but lower recall, indicating a tendency to correctly classify fraudulent transactions while potentially missing some instances of fraud. Conversely, other models showed higher recall at the expense of precision, resulting in a higher number of false positives.

VIII. CONCLUSION

Finally, it is concluded that The Credit card is an intrinsically secure device. Credit cards have proven to be useful for media Eventually replacing all of the things we carry around in our wallets, including credit cards. The credit card can be an element of solution to a security problem in the modern world. Future research could focus on identifying new features or combinations of features that could provide deeper insights into transaction patterns. Advanced feature engineering techniques like feature selection algorithms and domain-specific feature creation could be explored

REFERENCES

[1] Subramanian R.R., Seshadri K. (2019) Design and Evaluation of a Hybrid Hierarchical Feature Tree Based Authorship Inference Technique. In: Kolhe M., Trivedi M., Tiwari S., Singh V. (eds) Advances in Data and Information Sciences. Lecture Notes in Networks and Systems, vol 39. Springer, Singapore

[2] R. R. Subramanian, R. Ramar, “Design of Offline and Online Writer Inference Technique”, International Journal of Innovative Technology and Exploring Engineering, vol. 9, no. 2S2, Dec. 2019, ISSN: 2278-3075

[3] R. R. Subramanian, B. R. Babu, K. Mamta and K. Manogna, Design and Evaluation of a Hybrid Feature Descriptor based Handwritten Character Inference Technique,” 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Tamilnadu, India, 2019, pp. 1-5

[4] Joshva Devadas T., Raja Subramanian R. (2020) Paradigms for Intelligent IOT Architecture. In: Peng SL., Pal S., Huang L. (eds) Principles of Internet of Things (IoT) Ecosystem: Insight Paradigm. Intelligent Systems Reference Library, vol 174. Springer, Cham

[5] Andrew. Y. Ng, Michael. I. Jordan, On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes”, Advances in neural information processing systems, vol. 2, pp. 841-848,2020

A. Srivastava, A. Kundu, S. Sural, A. Majumdar, “Credit card fraud detection using hidden Markov model,” IEEE Transactions on Dependable and Secure Computing, vol. 5, no. 1, pp. 37–48, 2021K. Elissa, “Title of paper if known,” unpublished.

[6] Srivastava, A. Kundu, S. Sural, A. Majumdar, “Credit card fraud detection using hidden Markov model,” IEEE Transactions on Dependable and Secure Computing, vol. 5, no. 1, pp. 37–48, 2021K. Elissa, “Title of paper if known,” unpublished.

 DOI : 10.35291/2454-9150.2024.0195