IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12632

Credit Card Fraud Detection using Machine Learning and Deep Learning

Shrushti Deshmukh[1], Ayodhya Patil[2], Diksha Sonawane[3], Mayuri Hirnawale[4],

Dr.(Mrs) S. S. Raskar[5]

Department of Computer Engineering MESCOE, Pune, India

Abstract :This paper discusses how machine learning techniques can be used to detect credit card fraud. It covers the types of fraud, challenges, and various ML algorithms used. The steps for building an ML-based fraud detection system are explained, and current and future trends are examined. Overall, ML-based fraud detection systems can significantly improve accuracy and efficiency, leading to better customer protection and reduced financial losses.

Keywords :Machine Learning, Deep Learning, Logistic Regression, CNN.

I. INTRODUCTION

'Fraud' in credit card transactions is unauthorized and unwanted usage of an account by someone other than the owner of that account. Necessary prevention measures can be taken to stop this abuse and the behavior of such fraudulent practices can be studied to minimize it and protect against similar occurrences in the future.

In other words, Credit Card Fraud can be defined as a case where a person uses someone else's credit card for personal reasons while the owner and the card issuing authorities are unaware of the fact that the card is being used. Fraud detection involves monitoring the activities of populations of users in order to estimate, perceive or avoid objectionable behavior ,which consist of fraud, intrusion, and defaulting.

This is a very relevant problem that demands the attention of communities such as machine learning and data science where the solution to this problem can be automated. This problem is particularly challenging from the perspective of learning, as it is characterized by various factors such as class imbalance. The number of valid transactions far outnumber fraudulent ones. Also, the transaction patterns often change their statistical properties over the course of time. These are not the only challenges in the implementation of a real-world fraud detection system, however. In real world examples, the massive stream of payment requests is quickly scanned by automatic tools that determine which transactions to authorize. Machine learning algorithms are employed to analyse all the authorized transactions and report the suspicious ones. These reports are investigated by professionals who contact the cardholders to confirm if the transaction was genuine or fraudulent.

II. LITERATURE REVIEW

This section describes the literature review and Credit Card Fraud Detection Using Machine Learning.

In this paper, Authors [1] Kuldeep Randhawa1, Chu Kiong Loo1, Manjeevan Seera, Chee Peng Lim, Asoke K. Nandi proposes a methodology using machine learning algorithms for credit card fraud detection. Two algorithms (Random Forest, Logistic Regression) were tested on a publicly available dataset with data under-sampling. Hybrid methods with AdaBoost and majority voting were also tested. Majority voting generated the best MCC score of 0.942 for 30% noise added to the dataset. The paper concludes that majority voting is a secure technique for detecting credit card fraud in the presence of noise [1].

Article [2] F. K. Alarfaj et al.: CCF discusses credit card fraud (CCF) as a growing threat to financial institutions and the need for a robust classifier to handle the changing nature of fraud. The performance of machine learning (ML) methods varies based on the type of input data, such as the number of features, transactions, and correlation between features [2].

In [3], H. Tingfei et al.: proposed an oversampling method based on VAE for detecting credit card fraud. Although it achieved encouraging results, it cannot be applied to the unsupervised environment and may perform poorly when dealing with completely novel fraud data. The baseline model had the best performance in the open dataset, but future research will focus on improving the recall rate of the model while increasing precision and F-measure to achieve a recall performance comparable to the SMOTE and GAN methods [3].

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12632

Article [4] A. A. Taha, S. J. Malebary: presented performance of the proposed approach was evaluated through experiments with two real-world datasets, and it outperformed other machine learning algorithms in terms of accuracy, AUC, precision, and F1-score. The study highlights the importance of efficient parameter optimization for enhancing the predictive performance of the proposed approach [4].

III. PROPOSED SYSTEM

The above Fig 1 demonstrates mainly working of internal process. In this scenario overall system is designed in n tiers separately. The raw data taken as input, then pre-process the raw data by using data pre-processing tools. Therefore, the structured format of raw data is obtained by removing missing values, null values and outliers. After that, a one hot ending technique is applied to enlarge the dataset during the classification process. Our study includes various machine learning algorithms to build models. The results were obtained and predicted the unknown test sample. The measure the perforce of our proposed method, the accuracy and precision is calculated. The given process is continuous until we will get better results from system.

A. System Architecure

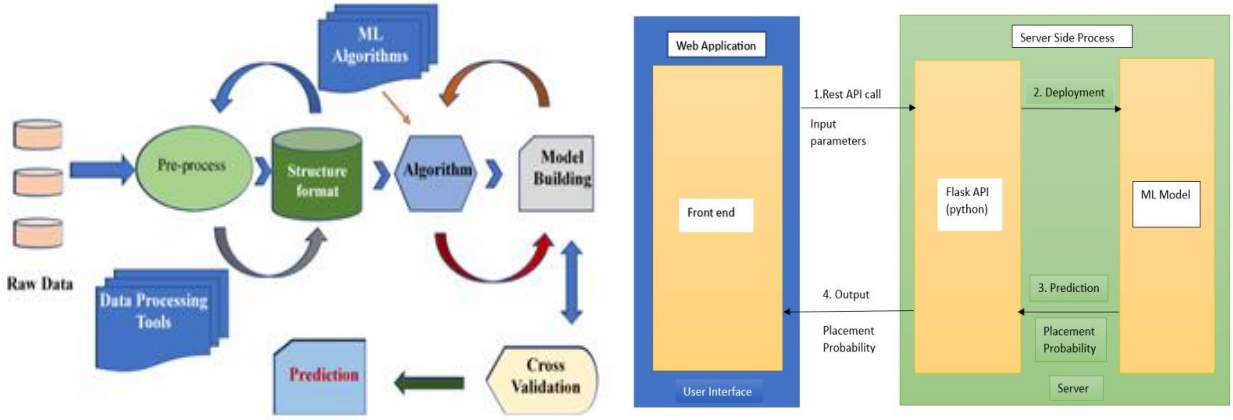

Fig. 1. System Architecture

The code above is a Python Flask application that serves as a web interface for the Credit Card Fraud Detection model. The model is trained and saved using machine learning techniques on a separate script, and the Flask application is used to interact with the model and provide predictions.The ValuePredictor function takes a list of features for a single transaction as input, loads the trained model from the model.pkl file using the pickle module, applies the model to the input data, and returns the predicted result.The home function returns the index.html template file, which contains a form for users to input the transaction features.

The predict function receives the user inputs from the form as a dictionary, converts the dictionary to a list, and converts the list elements to float data type. Then, it calls the ValuePredictor function to get the predicted result, and based on the result, it assigns the prediction message to either "Given transaction is fraudulent" or "Given transaction is NOT fraudulent". Finally, it returns the result.html template file, which displays the prediction message.The __name__ == "__main__" condition is used to check whether the script is being run directly or being imported as a module. If the script is being run directly, the Flask application runs in debug mode by calling the run method with the debug argument set to True. This allows for debugging in case of any errors.

This code is performing credit card fraud detection using machine learning algorithms. The dataset used is a credit card fraud dataset, which has 31 columns and 284,807 rows. The dataset is imbalanced, as it contains only 492 fraud cases.

The code performs the following steps:

1. Imports necessary libraries
2. Reads the dataset
3. Performs data visualization using seaborn and matplotlib libraries
4. Performs SMOTE (Synthetic Minority Over-sampling Technique) to balance the dataset
5. Performs Feature Scaling using Standard Scaler

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12632

6. Performs Dimensionality Reduction using PCA (Principal Component Analysis)
7. Applies machine learning algorithms to classify the data
8. Evaluates the performance of the algorithms using various metrics like accuracy, F1-score, and confusion matrix

B. Data collection

Credit card transaction data is collected from various sources, such as banks, credit card companies, and merchants. The data includes transaction features, such as the transaction amount, the time of day, the location of the transaction, and the type of merchant.

Sr.no	List of attributes/Features used in model		
	Features	**Description**	**Type**
1	Distance from home	This can help to identify transactions that occur far away from cardholder's usual location	nominal
2	Distance from latest transaction	This can help to identify multiple transactions occurring at distant location within period of time	nominal
3	Ratio to median Purchase	This can help identify transactions that device significantly from the cardholders typical spending behavior	ordinal
4	Online order	Online transactions may carry a higher risk of frau due to the remote nature	nominal

C. Data Preprocessing

The transaction data is pre-processed to prepare it for analysis. This may involve cleaning the data, removing duplicates, handling missing values, and transforming the data into a format suitable for analysis.

1. The code imports the necessary libraries for data manipulation and visualization, including pandas, numpy, matplotlib, seaborn, sklearn, and random.
2. It reads in a CSV file called 'creditcard.csv' and assigns it to a variable called 'data'.
3. The code then plots two distribution plots using seaborn for the 'Amount' and 'Time' columns in the data.
4. It creates a variable called 'class0' that filters the data for all rows where 'Class' is equal to 0.
5. It creates a variable called 'class1' that filters the data for all rows where 'Class' is equal to 1.

DOI: 10.17148/IJARCCE.2023.12632

6. It shuffles the 'class0' dataframe and selects the first 2000 rows and assigns it to a new dataframe called 'd1'.
7. It concatenates the 'd1' and 'class1' dataframes into a new dataframe called 'df_temp'.
8. It shuffles the 'df_temp' dataframe and saves it to a CSV file called 'creditcardsampling.csv'.
9. It plots a countplot using seaborn for the 'Class' column in the 'df' dataframe .
10. The code then installs the 'imblearn' library and imports the **SMOTE** oversampling algorithm.

It applies the SMOTE algorithm to the 'X' and 'Y' variables. It creates a new dataframe using the oversampled 'X' and 'Y' variables. It selects the 'Time', 'Amount', and 'Class' columns from the new dataframe and scales them using the StandardScaler from sklearn. It drops the 'Time' and 'Amount' columns from the new dataframe and concatenates the scaled columns with the remaining columns. It performs PCA on the new dataframe to reduce its dimensionality to 7 principal components. It saves the reduced dataframe to a new CSV file called 'finaldata.csv'. It splits the reduced dataframe into training and testing sets using the train_test_split function from sklearn.
It creates an instance of the SVM classifier using the 'rbf' kernel and sets the probability parameter to True. It fits the SVM classifier to the training data. It makes predictions using the SVM classifier on the testing data. It creates a GridSearchCV object to perform a grid search on the SVM classifier to find the best parameters for the model. It fits the GridSearchCV object to the training data. It saves the SVM classifier model to a file called 'model.pkl' . It loads the SVM classifier model from the file 'model.pkl' into a variable called 'model'.

A) Decision Tree :

A Decision Tree is a popular machine learning algorithm used in credit card fraud detection due to its interpretability and ability to capture complex decision-making processes. In the context of credit card fraud detection, a Decision Tree algorithm constructs a tree-like model by recursively partitioning the data based on the most informative features

B) Random Forest Algorithm

Random Forest is a versatile and powerful ensemble learning algorithm used in the above model for classification tasks. It operates by constructing multiple decision trees and combining their predictions to make final classifications.

C) CatBoost

CatBoost is a powerful machine learning algorithm that can be lever-aged for credit card fraud detection. CatBoost is particularly well-suited for this task due to its ability to handle categorical features and imbalanced datasets commonly encountered in fraud detection scenar-ios. With its gradient boosting framework, CatBoost can effectively

IV. CONCLUSION

The use of machine learning for credit card fraud detection has shown significant potential in improving fraud detection rates while reducing false positives. The effectiveness of ML based fraud detection models heavily relies on the quality of the data used to train them, which should be accurate, diverse, and representative of the actual fraud patterns.Some of the commonly used ML techniques for credit card fraud detection include supervised learning, unsupervised learning, and deep learning algorithms. These techniques can identify anomalies in transactional data, flagging suspicious transactions for further investigation.
Furthermore, ML-based fraud detection systems can adapt to evolving fraud patterns and trends, making them more effective than traditional rule-based systems. They also provide real-time detection, enabling quick responses to fraudulent activities.

REFERENCES

[1] K. Randhawa et al.: Credit Card Fraud Detection Using AdaBoost and Majority Voting, date of publication February 15, 2018.
[2] F. K. Alarfaj et al.: CCF Detection Using State-of-the-Art ML and DL Algorithms April 18, 2022.
[3] E. Esenogho et al.: Neural Network Ensemble With Feature Engineering for Improved Credit Card Fraud Detection February 15, 2022.
[4] H. Tingfei et al.: Using Variational Auto Encoding in Credit Card Fraud Detection August 25, 2020.
[5] A. A. Taha, S. J. Malebary: Intelligent Approach to Credit Card Fraud Detection Using an OLightGBM February 11, 2020.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023
DOI: 10.17148/IJARCCE.2023.12632

[6] X. Zhang, Y. Han, W. Xu, and Q. Wang, ‘‘HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture,’’ Inf. Sci., vol. 557, pp. 302–316, May 2021, doi: 10.1016/j.ins.2019.05.023.

[7] M. C. M. Oo and T. Thein, ‘‘An efficient predictive analytics system for high dimensional big data,’’ J. King Saud Univ. Comput. Inf. Sci., vol. 34, no. 1, pp. 1521–1532, Jan. 2022.

[8] B. Lebichot, Y.-A. Le Borgne, L. He-Guelton, F. Oblé, and G. Bontempi, ‘‘Deep-learning domain adaptation techniques for credit cards fraud detection,’’ in Proc. INNS Big Data Deep Learn. Conference, Genoa, Italy, 2019, pp. 78–88.

[9] S. R. Venna, A. Tavanaei, R. N. Gottumukkala, V. V. Raghavan, A. S. Maida, and S. Nichols, ‘‘A novel data-driven model for real-time influenza forecasting,’’ IEEE Access, vol. 7, pp. 7691–7701, 2019.

[10] J. Gao, Z. Zhou, J. Ai, B. Xia, and S. Coggeshall, ‘‘Predicting credit card transaction fraud using machine learning algorithms,’’ J. Intell. Learn. Syst. Appl., vol. 11, no. 3, pp. 33–63, 2019, doi: 10.4236/jilsa.2019.113003.