CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING

Sri B.Subba Reddy [1], R. Sravani [#2], Sk.Habeebunnisa[3], S.Likitha[4], S.V.S.Sai prasanth[5]
[1,2,3,4,5]*Department of ECE, N.B.K.R. Institute of Science and Technology*
Tirupati District, Andhra Pradesh, India.

Abstract*— *The most frequent issue in the ultramodern world is the discovery of credit card fraud. We used algorithms that used machine literacy approaches to uncover credit card fraud. We're using Decision Trees, Random timbers and Extreme Gradient boosting algorithms. The effectiveness of the model can be decided by using some public data as a sample. Also an factual world credit card data group from a fiscal institution is examined. Along with this, some clatter is supplemented to the data samples to supplementary check the soundness of the systems. The significance of the styles used in the paper is the first system constructs a tree against the conditioning performed by the stoner and using this tree swindles will be suspected. In the alternate system, a stoner exertion- grounded timber will have constructed and used this timber an attempt will be made in relating the suspect. The investigational issues show that the mainstream optional fashion attains decent perfection degrees in seeing fiddle circumstances in credit cards.

Keywords*— *Decision Tree algorithm, the Random Forest algorithm, and Extreme Gradient boosting algorithms

I. INTRODUCTION

Falsification of the credit card can be defined as the unapproved use of a client's card data to produce purchases or to dismiss finances from the cardholder's record. The misconduct highway robbery starts from the credit card when notoriety inaptly acquires the number published on a card or the essential records for the card to be operated. The proprietor of the card, the agent by whom the card is issued and indeed the patron of a card might not be informed of the fraud until the record is used to produce purchases. As shopping through internet-grounded operations and paying bills online has come into practice, there's no longer demand for a physical card to produce purchases. Fraud discovery in online shopping systems is the hottest content currently. Fraud investigators, banking systems, and electronic payment systems similar as PayPal must have an effective and complex fraud discovery system to help fraud conditioning that change fleetly. According to a Cyber Source report from 2017, the present fraud loss by order channel, that is, the chance of fraud loss in their web store was 74 per cent and 49 per cent in their mobile channels. Grounded on this information, the assignment is to determine anomalies across patterns of fraud geste that have changed relative to the history. The rising ofE-commerce business has redounded in a gentle growth within the operation of credit cards for online deals and purchases. With the rise in the operation of credit cards, the number of fraud cases has also doubled. Credit card frauds are which is done to gain plutocrat deceptively without the knowledge of the cardholder.

II. LITERATURE SURVEY

[1] Classifier, Unsupervised, Neural Network Models for Credit Card Fraud Detection, International Journal of Engineering Research and Technology (IJERT), Volume 9, Issue 04, 2020. U. Anjali Mohan, S. Karishma, L. Bhavya, V. Sasidhar Reddy, and (April 2020),

Recent years have seen a sharp rise in online purchases. A substantial number of these are credit card transactions made online. Banks and other financial institutions urgently require applications for identifying credit card fraud. Credit card fraud might involve making unauthorised withdrawals from an account or buying items without paying for them. Incidents of credit card fraud grew as people's thirst for money increased. As a result, the cardholder experiences a huge financial loss.

Renjith, Shini [2]. (2018). Using a support vector machine approach, fraudulent sellers in online marketplaces can be found. 10.14445/22315381/IJETT-V57P210, International Journal of Engineering Trends and Technology, 57, pp. 48–53.

The percentage of worldwide retail spending that goes to e-commerce has been steadily rising over the years, clearly demonstrating a shift in customer focus from bricks and mortar to clicks in the retail industry. Online marketplaces have emerged as one of the major forces driving this expansion in recent years. In-depth research is being done on fraudulent e-commerce buyers and their transactions, and various control and prevention measures are being considered. Marketplace seller fraud, also known as merchant fraud, is another type of fraud

that occurs there. One straightforward example is the sale of cheap goods and services that are never delivered.

[3] Suharjito, Suharjito & Saputra, Adi (2019). Fraud detection in e-commerce using machine learning. 10.14569/IJACSA.2019.0100943.

The number of internet users is growing, and with it, so are e-commerce transactions. We also see an increase in the amount of fraud in online purchases. Machine learning will be utilised to produce fraud protection in e-commerce; this work analyses the best machine learning method, which is Decision Tree, Naive Bayes, Random Forest, and Neural Network. The neural network obtains the maximum accuracy of 96%, followed by random forest (95%), naive bayes (95%), and decision tree (91%), according to the evaluation using the confusion matrix.

[4] A. K. Rai and R. K. Dwivedi, "Fraud Detection in Credit Card Data Using Unsupervised Machine Learning Based Scheme," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, pp. 421-426. In this paper, a method for identifying credit card fraud using unsupervised learning based on neural networks (NN) is proposed. The proposed strategy performs better than the current K-Means clustering, Isolation Forest, Local Outlier Factor, and Auto Encoder (AE) approaches. The proposed NN-based fraud detection approach has a 99.87% accuracy rate, compared to the 97%, 98%, 98%, and 99.75% accuracy rates of the existing methods AE, IF, LOF, and K Means.

III. METHODOLOGY

A. Dataset Description:

The dataset includes credit card transactions performed by European cardholders in September 2013.

We have 492 frauds out of 284,807 transactions in our dataset of transactions that took place over the course of two days. The dataset is very skewed, with frauds making up 0.172% of all transactions in the positive class. It only has numeric input variables that have undergone PCA transformation. Regrettably, we are unable to offer the original characteristics and additional context for the data due to confidentiality concerns. The major components obtained with PCA are features V1, V2,..., V28. The only features that have not been changed with PCA are "Time" and "Amount." The seconds that passed between each transaction and the dataset's first transaction are listed in the feature "Time."

The transaction amount is represented by the feature "Amount," which can be utilised for example-dependent, cost-sensitive learning. The response variable, feature "Class," has a value of 1 in cases of fraud and 0 in all other cases.

B. Dataset pre-processing:

Cleaning and converting raw data into a format that can be easily examined and used in machine learning models is known as data preparation. The preprocess function reads a CSV file in the context of the offered code, validates the dataset for any missing values, and then produces a success message.

The following processes are typically involved in data pre-processing:

1. Data cleaning is the process of deleting duplicate records, dealing with missing or null values, fixing inaccurate or inconsistent data, and handling outliers.

2. Data transformation is the process of transforming the data into a format that will allow for analysis. This can involve managing text data, categorical variable encoding, and scaling the data.

3. The process of feature extraction entails choosing pertinent features from the dataset that are crucial for modelling and analysis.

4. Feature engineering is the process of constructing new features out of pre-existing ones that can add fresh data to analysis and modelling.

5. Data integration is the process of combining various datasets into one dataset for analysis.

6. Data Reduction: In this step, redundant or irrelevant data are removed in order to reduce the size of the dataset.

C. Model selection:

It performs binary classification on the supplied dataset using pre-built classifiers from the Scikit-learn and XGBoost libraries.

The models employed are: RandomForestClassifier: An ensemble technique built on decision trees that employs bootstrap aggregation (bagging) to enhance performance and lessen overfitting. A decision tree method called DecisionTree Classifier builds a tree-like model of decisions and potential outcomes.

XGBClassifier: An improved gradient boosting approach that adds regularisation to lessen overfitting and employs decision trees as base learners. The train test split function from Scikit-learn is used to divide the pre-processed data into training and testing sets before the models are trained on it. In order to solve the issue of class imbalance, the code also employs the SMOTE algorithm from the imblearn library to oversample the minority class (fraudulent transactions).

D. Flask App:

A well-liked web framework for Python developers to create web applications is Flask. The offered credit card fraud detection code uses Flask to build a REST API. Credit card transactions can be sent to the model using this API for prediction. The Flask server receives a transaction through the API, analyses the data, sends it to the machine learning model for prediction, and finally sends the prediction result back to the client who made the request. The Flask app is used to specify the routes, construct the API endpoints, and process incoming requests. It also helps to make the model available via an HTTP endpoint as a web service.

IV. RESULTS

Fig2.graph comparing the accuracy of Models

According to the code, a Flask web application for credit card fraud detection is being used. It has various methods for loading data, pre-processing data, viewing data, training and assessing various models, and making predictions. This online application appears to be made with the help of the scikit-learn, imbalanced-learn, pandas, numpy, pygal, and flask libraries.
Users of the application can upload a CSV file containing credit card transaction data, pre-process the data, see the data, train and test various models, including Random Forest, Decision Tree, and XGBoost, and also utilise the selected model to forecast future data. The accuracy, precision, and recall scores are used to assess the Random Forest, Decision Tree, and XGBoost models.

V. CONCLUSION

In this research, we use three machine-learning techniques to detect credit card fraud: the Decision Tree algorithm, the Random Forest algorithm, and the XGBoost Classifier algorithm. We found the Random Forest algorithm to be the best technique, and it shows whether the credit card is fraudulent or not.

VI. REFERENCES

[1] Taha, Altyeb & Malebary, Sharaf. (2020). An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine. IEEE Access. 8. 25579-25587.

[2] Assaghir, Zainab & Taher, Yehia & Haque, Rafiqul & Hacid, Mohand-Said & Zeineddine, Hassan. (2019). An Experimental Study With Imbalanced Classification Approaches for Credit Card Fraud Detection. IEEE Access.

[3] L. Meneghetti, M. Terzi, S. Del Favero, G. A Susto, C. Cobelli, “DataDriven Anomaly Recognition for Unsupervised Model-Free Fault Detection in Artificial Pancreas”, IEEE Transactions On Control Systems Technology, (2018) pp. 1-15

[4] F. Carcillo, Y.-A. Le Borgne and O. Caelen et al., “Combining unsupervised and supervised learning in credit card fraud detection”, Information Sciences, Elsevier (2019), pp. 1-15.

[5] Ashphak, Mr. & Singh, Tejpal & Sinhal, Dr. Amit. (2012). A Survey of Fraud Detection System using Hidden Markov Model for Credit Card Application Prof. Amit Sinhal. 1.

[6] Renjith, Shini. (2018). Detection of Fraudulent Sellers in Online Marketplaces using Support Vector Machine Approach. International Journal of Engineering Trends and Technology. 57. 48-53. 10.14445/22315381/IJETT-V57P210.

[7] Saputra, Adi & Suharjito, Suharjito. (2019). Fraud Detection using Machine Learning in e-Commerce. 10.14569/IJACSA.2019.0100943.

[8] A. K. Rai and R. K. Dwivedi, "Fraud Detection in Credit Card Data using Unsupervised Machine Learning Based Scheme," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 421-426, doi: 10.1109/ICESC48915.2020.9155615.

[9] John O. Awoyemi, Adebayo O. Adetunmbi, Samuel A. Oluwadare et al., “Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis”, IEEE, 2017.

[10] Rajendra Kumar Dwivedi, Sonali Pandey, Rakesh Kumar “A study on Machine Learning Approaches for Outlier Detection in Wireless Sensor Network” IEEE International Conference Confluence, (2018).