IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Flask-Based Credit card Fraud Detection System with Machine Learning

Asst. Prof.V.Rohini[1] , M.Pavan Teja[2] , M.Balaji[3] , M.Bhavani Shankar[4] , N.V.Sathvick[5] , N.Sumanth[6]

[1,2,3,4,5,6]*Department of ECE , N.B.K.R. Institute of Science and Technology*

Tirupati District, Andhra Pradesh, India.

ABSTRACT:

Both the use of credit cards for ordinary purchases and online purchases is skyrocketing, as is the amount of credit card fraud. Every day, a sizable amount of transactions are fraudulent. a number of contemporary methods, such as artificial neural networks. In order to identify these fraudulent transactions, various machine learning methods, such as Logistic Regression, Decision Trees, Random Forest, Artificial Neural Networks, Logistic Regression, K-Nearest Neighbors, and K-means clustering, among others, are compared. In order to identify the best answer to the problem and subtly produce the outcome of the fraudulent transaction, this paper employs evolutionary algorithms and neural networks. The key goals are to identify the fraudulent transaction and create a strategy for producing test data. This algorithm uses a heuristic method to solve problems of great complexity.

In this project, we suggest a system for detecting credit card fraud that makes use of machine learning to spot fraudulent transactions. In order to accurately identify fraudulent transactions, our system incorporates a range of machine learning methods, such as decision trees, logistic regression, and boosting methods. The Flask web framework is used to create the system, which is intended to be easily deploy-able and flexible in a range of financial situations.

OBJECTIVE:

The Main Objective is to detect online fraud that occurred in credit card transactions when using online financial transaction system.

Currently, the risk of network information insecurity is increasing rapidly in number and level of danger. The methods mostly used by hackers today is to attack end-to end technology and exploit human vulnerabilities.

I. INTRODUCTION:

Today, automated teller machines (ATMs), store readers, banks, and online internet banking systems all read the information from credit and debit cards. They have a special card number, which is crucial. Both the physical security of the plastic card and the confidentiality of the credit card number are essential to its security.

The quantity of credit card transactions is increasing quickly, which has caused a significant increase in fraudulent activity**.** A significant issue that impacts both financial institutions and their clients is credit card theft. Fraudulent transactions have the potential to cause large financial losses, ruin the reputation of a financial institution, and erode client confidence. Thus, it is crucial to find fraudulent transactions. Machine learning algorithms have been a potential strategy for identifying credit card fraud in recent years. This study offers a thorough analysis of various machine learning methods for detecting credit card fraud.

II. LITERATURE SURVEY:

One popular approach is to use supervised machine learning algorithms to identify fraudulent transactions. These algorithms are trained on the

historical transaction data that is labeled as either fraudulent or legitimate. Some popular supervised learning algorithms used for credit card fraud detection include decision trees, random forests, support vector machines, and logistic regression.

[1]Vimala Devi, J and others proposed that Three machine-learning techniques were described and put into use to find fraudulent transactions.

The performance of classifiers or predictors is assessed using a variety of metrics, including the Vector Machine, Random Forest, and Decision Tree. Either these measures depend on or don't depend on prevalence. Also, similar methods are employed in mechanisms that detect credit card fraud, and the outcomes of these algorithms have been contrasted.

[2] Popat and Chaudhary. presented supervised algorithms. Some of the methods utilized include Deep Learning, Logistic Regression, Nave Bayesian, Support Vector Machine (SVM), Neural Network, Artificial Immune System, K Nearest Neighbor, Decision Tree, Fuzzy Logic Based System, and Genetic Algorithm. To perform prediction, grouping, and outlier detection, we compared machine learning techniques.

[3] Sevkli and Kibri. Build a deep learning model utilising the grid search method. The effectiveness of the developed model is contrasted with that of two other well-known machine-learning techniques, logistic regression (LR) and support vector machine (SVM). In order to compare the constructed model's performance to logistic regression and support vector machine models, it is applied to the credit card data set.

[4] A deep learning-based technique for identifying fraud in credit card transactions has been put out by Asha R. B. et al. predicting the occurrence of fraud using machine-learning methods like support vector machines, k-nearest neighbors, and artificial neural networks.

[5] Suhas and Dhotre Borse proposed, The Naive Bayes classification of machine learning was used to foretell legitimate and fraudulent transactions.

III. PROBLEM DEFINITION:

There are several challenges that make this process difficult to accomplish, but one of the main issues with fraud detection is the lack of real-world data for academic researchers to conduct experiments on, as well as the literature that provides experimental results. The sensitive financial information related to the fraud that must be kept secret in order to protect the privacy of the consumer is the cause of this. Here, we list the several qualities that a fraud detection system should possess in order to produce accurate findings. As only a very small portion of all credit card transactions are fraudulent, the system ought to be able to manage skewed distributions. There must to be a suitable way to deal with the noise. Inaccurate dates are just one example of the noise that can be found in data. Regardless of how big the training set is, the noise in the real data restricts how accurate generalization can be. Overlapping data is another issue in this subject. Many transactions could appear to be fraudulent when they're actually legitimate

IV. EXISTING SYSTEM:

Much research has been done on studying credit card fraud detection. In past people manually detect fraud transactions. But, the entire problem of credit card fraud detection suffers from a problem of Imbalanced data (a very highly imbalanced data). This problem requires us to heavily process the data before training any machine learning model like Random Forest etc.

In existing system there are some disadvantages like accuracy low, requires more time, difficult to handle.

V. PROPOSED SYSTEM:

Proposed several machine learning models which are used to classify Fraud are accurate than the models in existing method. Therefore, we propose a machine learning-based approach which combines a new technique of pre-processing the data for features transformation, DecisionTreeClassifier, Logistic Regression and XGBoost ML algorithm give the best accuracy techniques to eliminate the bias and the deviation of instability and performing classifier tests based.

The main advantages of our proposed system are requires less time, good accuracy, easy to handle

VI. METHODOLOGY:

1. Dataset Collection:

The data can be collected from various sources, including financial institutions, credit card companies, and online platforms that provide public datasets for research purposes.

It is crucial to take into account the following criteria during data gathering in order to assure the dataset's quality and relevance:

A. Data balance: To prevent model bias towards either class, the dataset should have an even proportion of both fraudulent and lawful transactions.
B. Data quality: The obtained dataset must be accurate, without missing values or inconsistent entries, and of high quality.
C. Data coverage: The dataset has to include both online and offline transactions, as well as domestic and foreign, low- and high-value transactions.
D. Data privacy: Privacy concerns should be taken into account while collecting and handling credit card transaction data because it contains sensitive information. It is crucial to abide by the rules and legislation governing data protection.

2. Data Pre-processing:

The collected data set needs to be pre-processed to ensure that it is in a format suitable for machine learning models.

A. Data cleaning: The data set must first be cleared of any redundant and extraneous information. This entails getting rid of duplicates, fixing mistakes, and taking care of missing data.
B. Feature selection: The acquired dataset might include a number of features that are unrelated to the fraud detection issue. The process of feature selection entails locating and choosing the most crucial features that will be employed in the machine learning models. The accuracy and effectiveness of the model are enhanced by this phase.
C. Data normalization: The scales and units of the collected data may differ, which may have an impact on how well a machine learning model performs. To enhance the performance of the model, data normalization entails converting the data to a common scale and unit.
D. Data encoding: The project's machine learning techniques call for numeric input data. As a result, utilizing methods like one-hot encoding, categorical data like transaction type and merchant category must be converted to numerical values.
E. Data splitting: Training, validation, and testing sets are created from the pre-processed dataset. The validation set is used to fine-tune the model's parameters, the testing set is used to assess the model's performance, and the training set is used to train the machine learning model.

3. Feature engineering:

Feature engineering is an important step in the process of credit card fraud detection. This step involves selecting the most relevant features that will be used to train the machine learning model. Some of the features commonly used for credit card fraud detection include the transaction amount, location, time of day, and type of transaction.

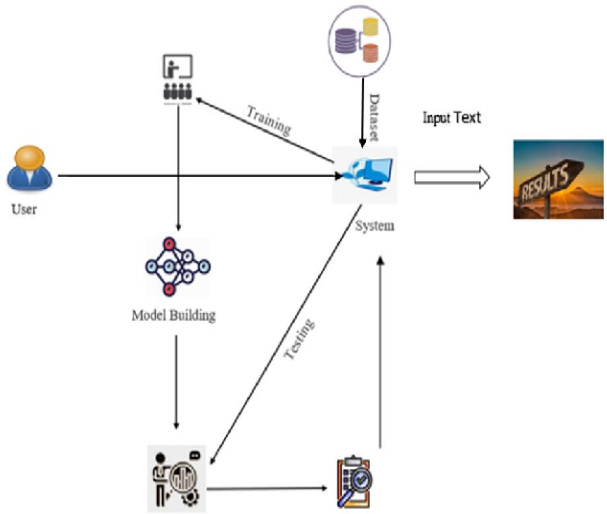
4. Model Architecture:

The model architecture for Flask-Based Fraud Detection System with Machine Learning typically involves a combination of several machine learning techniques like supervised learning and deep learning.

A. Pre-processing layer: This layer processes and prepares the data for use in the model. It includes data cleaning, normalization, and encoding.

B. Supervised learning layer: This layer uses a supervised learning algorithm, such as logistic regression, decision trees, or support vector machines, to classify transactions as fraudulent or legitimate based on the labeled data.

C. Ensemble layer: This layer combines the outputs of the previous layers to generate a final prediction. This can be done using techniques such as voting, stacking, or weighted averaging. These algorithms are explained in the Algorithm section.



5. Flask Framework:

The Flask web framework was used to build a web application that allowed users to interact with the credit card fraud detection model. The web application was designed to be simple and user-friendly, allowing users to submit transactions and receive predictions about whether transaction is fraudulent or legitimate.

Flask was chosen as the web framework due to its simplicity and flexibility, making it easy to build a lightweight and scalable application. Flask provides a number of features that were useful for this project, including:

A. Routing: Flask's routing capabilities allowed developers to define URL routes and corresponding view functions that were executed when a request was made to that URL. This made it easy to create different pages and functionality within the web application.

B. Templates: Flask's template engine allowed developers to easily render HTML templates and pass data to the templates. This made it easy to create a consistent user interface and display the results of the credit card fraud detection model.

C. Forms: Flask's form handling library simplified the process of validating and processing form data submitted by users. This made it easy to collect transaction data from users and pass it to the credit card fraud detection model.

VII. RESULTS:

Table.1:

Model	Decision Tree	Random Forest	XGBoost
Accuracy	91.95	94.02	97.71

Table.1 compares the evaluation parameter named accuracy of two existing methods like Random Forest, Decision Tree with the proposed XGBoost method.The accuracy of XGBoost is higher than the other two methods.

VIII. CONCLUSION:

The project examined the Flask based credit card fraud detection system that uses machine learning to identify fraudulent transactions. Our system utilizes a variety of machine learning algorithms and is designed to operate in real-time, making it an effective solution for detecting credit card fraud in today's fast-paced financial industry. The system achieved high accuracy in identifying fraudulent transactions and has the potential to significantly reduce financial losses caused by credit card fraud. The system achieved high accuracy and low false positive rates, making it an effective solution for financial institutions and their customers.

IX. 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).