JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

DETECTION OF CREDIT CARD FRAUD VIA MACHINE LEARNING

Hanumanthappagari Twinkle[1], Dr. S. Giriprasad[2]

M. Tech Student[1], Associate Professor[2]

Department of Electronics and Communication Engineering, VISWAM Engineering College, Madanapalle, Andhra Pradesh, India

Abstract:

In this study, we use machine learning methods, more specifically Decision Tree, Random Forest, and Extreme Gradient boosting algorithms, to address the crucial problem of credit card fraud. To evaluate the effectiveness of these models, we utilize publicly available data as a sample and assess their performance on real credit card transaction data obtained from a financial institution. To test the resilience of our systems, we introduce synthetic noise into the data samples. Our approach makes two key contributions. First, we construct a decision tree based on user activities, which enables us to identify potential fraud by analyzing user behavior patterns. Second, we create a forest of decision trees using user activity as the basis, allowing us to identify suspicious transactions by aggregating the outputs of multiple trees. The experimental results showcase the high accuracy achieved by our proposed methods in detecting fraudulent activities in credit card transactions. This research offers valuable insights and practical solutions to combat credit card fraud, benefiting both financial institutions and their customers.

Keywords: Decision Tree, Random Forest and Extreme Gradient boosting algorithms.

Introduction:

The way we make purchases and carry out financial transactions has undergone tremendous change as a result of the quick development of technology and the rising popularity of e-commerce. In recent years, credit card fraud has become an urgent issue in tandem with this digital revolution. Credit card fraud refers to the unauthorized use of a customer's card data to make purchases or withdraw funds without the cardholder's consent. This fraudulent activity can occur when individuals obtain the card number or essential card details through illicit means. In the modern era, the traditional notion of a physical card is no longer a prerequisite for fraudulent transactions. With the widespread use of internet-based applications and online payment systems, criminals have found new avenues to exploit unsuspecting individuals. This evolution has necessitated the implementation of robust fraud detection systems to protect consumers, financial institutions, and electronic payment platforms from such illicit activities. The detection of fraud in online shopping systems has become a critical area of focus for fraud investigators, banking institutions, and electronic payment service providers like PayPal. These entities must continuously refine and enhance their fraud detection systems to keep pace with the rapidly changing methods employed by fraudsters. An insightful report by Cyber Source in 2017 highlighted the escalating threat, revealing that 74 percent of fraud losses occurred in web stores, while 49 percent were attributed to mobile channels. This data underscores the need to identify anomalies and adapt to evolving patterns of fraudulent behavior. As the popularity of e-commerce and the use of credit cards for online purchases both continue to climb, the prevalence of credit card fraud has skyrocketed. In order to steal money in secret, fraudsters attempt to deceive cardholders into handing over their money. To combat this pervasive issue, effective methods for detecting and preventing credit card fraud must be employed, safe guarding consumer interests and preserving the reliability of online transactions. This essay will examine the various aspects of credit card fraud, look at the tactics used by fraudsters, and discuss the necessity of developing cutting-edge fraud detection systems in order to minimize this ongoing menace. By understanding the challenges and implementing proactive measures, we can fortify our defenses against credit card fraud and foster a safer environment for online commerce.

1. Related works:

Credit Card Fraud Detection using Classification, Unsupervised, and Neural Networks Models: In recent years, the proliferation of online transactions has witnessed a significant surge, with a notable portion being comprised of credit card transactions. As a result, there is a pressing need within the banking and financial industries for robust applications capable of detecting credit card theft. Credit card fraud serves the purpose of either acquiring goods or services without payment or illicitly withdrawing funds from unauthorized accounts. Sadly, there has been an increase in credit card fraud cases due to the desire for financial gain. Such fraudulent operations have serious consequences, since cardholders suffer significant financial losses. It is imperative to create and put into action efficient

countermeasures to this threat. The field of fraud detection requires continuing study and innovation due to the ongoing development of credit card fraud schemes. Collaborative efforts between financial institutions, technology providers, and regulatory bodies are vital to staying ahead of fraudsters and safeguarding the interests of cardholders. By leveraging the power of technology and implementing stringent security measures, the banking and financial industries can mitigate the risks associated with credit card fraud and ensure a safer online transaction environment.

Detection of Fraudulent Sellers in Online Marketplaces using Support Vector Machine Approach: The amount of money spent on e-commerce globally has steadily increased over the years, reflecting a clear change in customer interest away from brick-and-mortar stores and toward online retailers. 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. Merchant fraud refers to another type of fraud that occurs in marketplaces on the seller side. One straightforward example of this kind of fraud is the sale of goods or services at low prices but with no guarantee of delivery. This study makes an effort to propose a framework using machine learning methods to identify such dishonest merchants.

Fraud Detection using Machine Learning in e-Commerce: The number of internet users is growing, and with it, so are e-commerce transactions. We also see an increase in the amount of online transaction fraud. Machine learning will be utilized to build fraud protection in e-commerce. This research will study the best machine learning method, which will be a combination of Decision Tree, Naive Bayes, Random Forest, and Neural Network. The neural network has a 96 percent accuracy rating in the confusion matrix examination, followed by random forest with a 95 percent rating, Naive Bayes with a 95 percent rating, and decision tree with a 91 percent accuracy rating. The average F1-Score can go from 67.9 to 94.5 percent, and the average G-Mean can go from 73.5 to 84.6 percent, thanks to the Synthetic Minority Over-sampling Technique (SMOTE).

Fraud Detection in Credit Card Data using Unsupervised Machine Learning Based Scheme: The rise of communication and e-commerce technology has made credit cards the most widely used method of payment for both online and off-line purchases. Therefore, it is crucial that this system's security guards against fraudulent transactions. Each year, there are more fraudulent credit card data transactions. Researchers are experimenting with cutting-edge methods to identify and stop such scams in this direction. However, there will always be a need for certain methods that can quickly and accurately identify these frauds. 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, while the existing AE, IF, LOF, and K Means methods have accuracy rates of 97%, 98%, 98%, and 99.75%, respectively.

Fraud Detection in Online Credit Card Transactions: A Comparative Study of Machine Learning Approaches: With the rapid growth of e-commerce and online transactions, the occurrence of fraudulent activities has become a significant concern for financial institutions and customers alike. Online credit card fraud poses a substantial threat to the security and trustworthiness of electronic payment systems. Therefore, there is an urgent need to develop effective fraud detection mechanisms to identify and prevent fraudulent transactions. By examining different models and techniques, this study aims to provide insights into the strengths and limitations of each approach, enabling financial institutions to make informed decisions in selecting the most suitable fraud detection system. The study evaluates a range of machine learning algorithms, including supervised learning methods such as logistic regression, decision trees, and support vector machines, as well as unsupervised learning techniques like clustering and anomaly detection. The evaluation criteria include accuracy, precision, recall, and F1-score, considering both the ability to identify fraud accurately and minimize false positives.

2. Methodology:

Proposed system:

We evaluate the performance of these models using publicly available data as well as actual credit card transaction data obtained from a financial institution. To test the robustness of our systems, we introduce synthetic noise into the data samples. Our approach includes a decision tree based on user activities and a forest of decision trees based on user activity, enabling the identification of potential fraud and suspicious transactions. Experimental results demonstrate the high accuracy of our methods in detecting fraudulent activities in credit card transactions, offering valuable insights and practical solutions to combat credit card fraud for the benefit of both financial institutions and their customers.

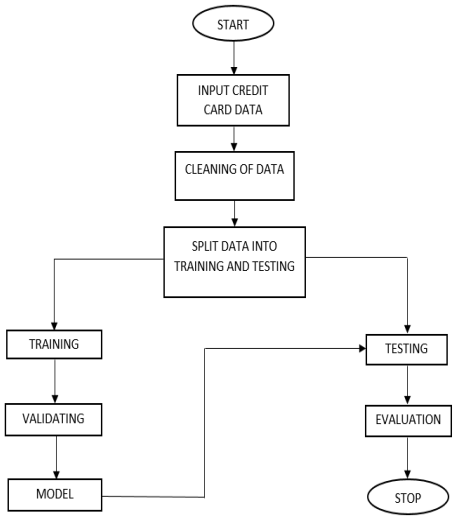


Figure 1: Block diagram

3. Implementation:

The algorithms listed below were used to complete the project.

1. Random Forest:

A well-liked machine learning algorithm called Random Forest is renowned for its adaptability and durability in handling numerous jobs including classification and regression. It is a member of the ensemble teaching family, which integrates various independent models to produce more precise forecasts. The name of the method is derived from the idea of a forest, where each tree is a decision tree. In Random Forest, various subsets of the training data are used to construct a collection of decision trees. The algorithm randomly selects subsets of the data and features, creating diverse trees that learn different aspects of the problem. During training, each decision tree independently makes predictions, and the final prediction is obtained through a voting or averaging mechanism. One of the key advantages of Random Forest is its ability to handle large feature sets without overfitting. By randomly selecting a subset of features for each tree, it reduces the correlation among trees and increases the overall model's generalization capability. Moreover, Random Forest is less prone to overfitting than individual decision trees, making it more resistant to noise and outliers in the data. Another benefit of Random Forest is its interpretability. It can provide insights into feature importance, allowing users to understand which variables contribute the most to the predictions. This information is valuable for feature selection and understanding the underlying relationships in the data.

2. XG Boost:

XG Boost (Extreme Gradient Boosting) is a potent machine learning method that has seen tremendous growth in the fields of predictive modeling and data science. It is well renowned for handling structured and tabular data with remarkable performance and efficiency. The gradient boosting framework is implemented in XG Boost, which combines various weak predictive models to create a strong predictive model. One of the key advantages of XG Boost is its ability to handle a variety of data types, including numerical and categorical features. It employs a gradient boosting approach, where subsequent models are built to correct the errors made by previous models. This iterative process allows XG Boost to optimize the objective function, which can be customized based on the specific problem at hand. XG Boost incorporates several techniques to improve model performance, such as regularization, tree pruning, and parallel processing. It also provides important features like feature importance ranking, which helps in understanding the contribution of each feature towards the model's predictions use to its exceptional accuracy and scalability, XG Boost has been successfully applied in various domains, including finance, healthcare, and online advertising. Its efficiency and ability to handle large datasets make it a popular choice for data scientists and machine learning practitioners.

3. Decision Tree:

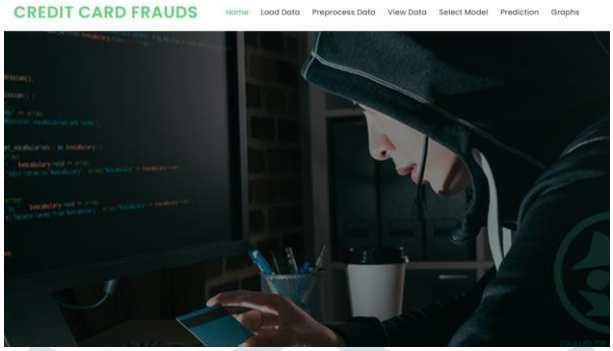
A decision tree is a popular machine learning algorithm used for both classification and regression tasks. It resembles a flowchart-like structure with internal nodes representing features or attributes, and leaf nodes representing the outcome or decision. The method creates a hierarchical structure by iteratively partitioning the data depending on feature values before making a choice. The decision tree algorithm divides the input into the most advantageous features at each node based on factors like Gini impurity or information gain. This process continues until a stopping criterion is met, such as reaching a maximum depth or a minimum number of samples at a node Decision trees provide a number of benefits, one of which is interpretability because the resulting tree is simple to comprehend and depict. Additionally, they are able to manage both categorical and numerical features. However, overfitting is a risk with decision trees

that can be reduced by adopting procedures like pruning or assembling approaches like random forests. In summary, decision trees are versatile and intuitive machine learning models that can be used for various tasks. Their ability to handle different types of features and provide interpretability makes them a valuable tool in the field of data analysis and predictive modelling.

4. Results and Discussion:

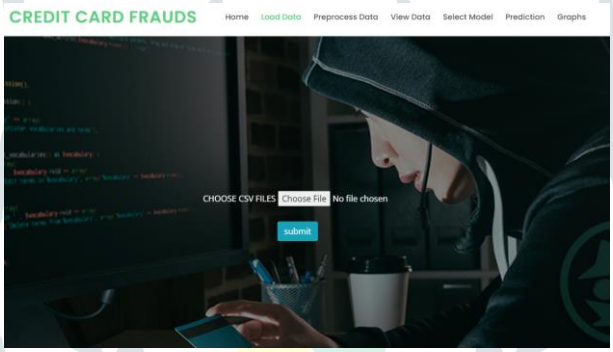
The following screenshots are depicted the flow and working process of project.

Home Page: Here user view the home page for credit card fraud detection appellation.



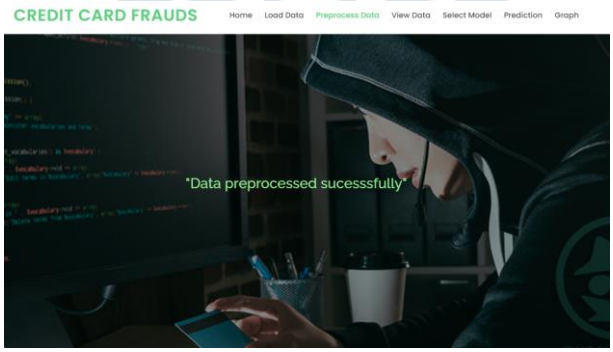
Load Page:

User will Load the Data set



Preprocessing Page:

User preprocess the data

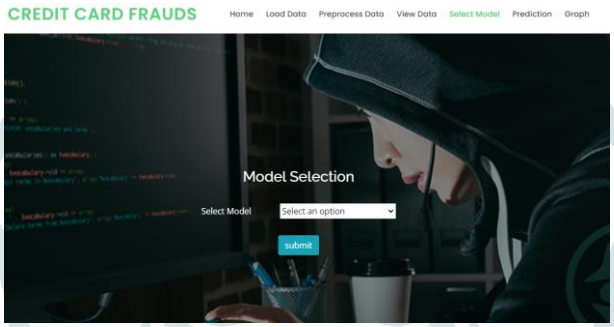


View Page:

User View the Data

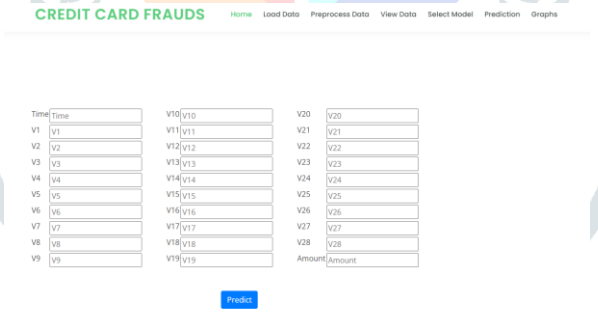
Model:

User will View the accuracy on every algorithm.



Prediction page:

User will give a proper input and view the result



5. Conclusion:

In this study, three machine learning methods are used to identify credit card fraud. Extreme gradient boosting, Decision Tree, Random Forest, and Extreme gradient boosting are the best strategies we could find for determining if a credit card transaction is fraudulent or not.

6. 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).
[11] Dal Pozzolo, A., Caelen, O., Le Borgne, Y. A., Waterschoot, S., Bontempi, G. (2015). Learned lessons in credit card fraud detection from a practitioner perspective. Expert Systems with Applications, 42(10), 5076-5086.
[12] Bhattacharyya, S., Bose, S., Maulik, U. (2011). A novel approach for credit card fraud detection using genetic algorithm. Applied Soft Computing, 11(1), 464-474.
[13] Chen, T., Kao, D. Y., Wei, D., Sanders, W. H., Chiang, M. (2018). Deeppwning: Adversarial deep learning for financial signal representation and trading. IEEE Transactions on Information Forensics and Security, 13(11), 2899-2912.
[14] Bahnsen, A. C., Aouada, D., Ottersten, B. (2014). Feature engineering strategies for credit card fraud detection. Expert Systems with Applications, 41(4), 1855-1864.
[15] Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., Muller, P. A. (2018). Deep learning for time series classification: a review. Data Mining and Knowledge Discovery, 33(4), 917-963.