International Journal of Early Childhood Special Education (INT-JECSE)
DOI:10.48047/INTJECSE/V15I4.86 ISSN: 1308-5581 Vol 15, Issue 04 2023

ONLINE PAYMENT FRAUD DETECTION

Mr. CH MAHESH BABU
ASSISTANT PROFESSOR
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SREYAS INSTITUTE OF ENGINEERING AND TECHNOLOGY

BOJJA SWEEHONEY
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SREYAS INSTITUTE OF ENGINEERING AND TECHNOLOGY
sweehoney.chinni@gmail.com

PADAM PRATHYUSHA
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SREYAS INSTITUTE OF ENGINEERING AND TECHNOLOGY
ammulupadam001@gmail.com

BOMMEPALLI DEVENDRA REDDY
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SREYAS INSTITUTE OF ENGINEERING AND TECHNOLOGY
bdevendrareddy1188@gmail.com

MAROJU SATHVIKA
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SREYAS INSTITUTE OF ENGINEERING AND TECHNOLOGY
sathvikamaroju2k3@gmail.com

ABSTRACT
As we are approaching modernity, the trend of paying online is increasing tremendously. The increase of use of online transactions is causing an increase in fraud. The frauds can be detected by using various approaches but they lag at accuracy and have some drawbacks. The online payment method leads to fraud that can happen using any payment app. The network transaction has the characteristics of low cost, wide coverage and high frequency, which makes the detection of fraud more complex. Aiming at the problem of difficult fraud detection in network transactions Fraud Detection is very important. Recent research has shown that machine learning techniques have been applied very effectively to the problem of payments related fraud detection. Such Machine Learning based technique have the potential to evolve and detect previously unseen patterns of fraud. So, we are going to use some machine learning algorithms to predict the transaction whether it is fraud or not. The selection of an algorithm to predict fraudulent transactions is based on attributes such as accuracy, recall, F1-score or etc. By considering these factors, we can identify the algorithm with the highest performance in terms of correctly identifying fraudulent transactions, ensuring the model's effectiveness and reliability. The model can analyze the transaction data uploaded and return the fraud detection results to users. We show that our proposed approaches are able to detect fraud transactions with high accuracy and reasonably low number of false positives.
Keywords*:* *Machine Learning, Online, Fraud, Detection, model, Transaction, Prediction, Algorithm.*

INTRODUCTION
The risk of payment fraud has recently become an urgent worry for people, organisations, and financial institutions alike due to the exponential development of online transactions. Real-time fraud detection has grown to be a serious problem that necessitates effective and trustworthy solutions in order to protect the integrity of online payment systems. Machine learning has become a significant

tool in the fight against online payment fraud, with the capacity to quickly and accurately identify fraudulent transactions. This document provides as an introduction to the field of machine learning based online payment fraud detection. It examines the importance of the issue, discusses the main difficulties encountered, and focuses on the benefits of using machine learning models for fraud detection. It also gives a brief review of the most widely used machine learning approaches and algorithms in this field. Online payment fraud detection's main objective is to spot and stop fraudulent transactions in real time, minimising financial damages for both customers and enterprises. While still somewhat effective, conventional rulebased systems sometimes struggle to keep up with changing fraud tendencies and may produce a large number of false positives. On the other side, machine learning algorithms excel in analysing massive volumes of data and learning complicated patterns, enabling them to more effectively and correctly identify fraudulent actions. Here we will explore the many steps required in creating a powerful machine learning system for detecting online payment fraud. It will go over how crucial feature engineering, model selection, and data pretreatment. The incorporation of machine learning models into current fraud detection systems will also be covered, underlining the necessity of ongoing model monitoring and updating to stay up with developing fraud strategies. The document will also discuss evaluation criteria like accuracy, recall, precision, and F1-score that are used to gauge how well fraud detection models perform. To ensure an efficient fraud prevention strategy, it will emphasise the need of striking a balance between the fraud detection rate and reducing false positives.Online payment fraud detection can be greatly improved by utilising the power of machine learning, allowing businesses and financial institutions to stay one step ahead of fraudsters. In order to tackle the growing problem of online payment fraud, this document intends to stimulate the development of more sophisticated and effective fraud detection systems by offering helpful insights and recommendations for researchers, practitioners, and stakeholders in the field.

LITERATURE SURVEY

Online Fraud Detection System: In this paper they had a behaviour based approach to classification using Support Vector machine is used to improve its accuracy. If there are any changes in the conduct of the transaction, the frauds are predicted and taken for further process. Due to large amount of data credit/debit card fraud detection problem is rectified by their proposed method. Machine Learning based Approach to Financial Fraud Detection Process in Mobile Payment System: In this paper their proposal was to detecting mobile payment fraud based on machine learning, supervised and unsupervised method to detect fraud and process large amounts of financial data. Moreover, our approach performed sampling process and feature selection process for fast processing with large volumes of transaction data and to achieve high accuracy in mobile payment detection. F-measure and ROC curve are used to validate our proposed model.

Online Transaction Fraud Detection System Based on Machine Learning: Prediction: This paper designed two fraud detection algorithms based on Fully Connected Neural Network, whose AUC values can achieve 0.912 and 0.969 respectively. Meanwhile, we designed an interactive online transaction fraud detection system based on Random Forest model, which can automatically analyze the transaction data uploaded and return the fraud detection results to users.Online Transaction Fraud Detection System Using Machine Learning and E-Commerce: In this paper they used BLA to detect this problem. An advantage to use BLA approach to reduce number of positive false transactions identified as malicious by an FDS although they are genuine. An FDS runs at a credit card issuing bank. Each incoming transaction is submitted to the FDS for verification. FDS receives the card details and transaction value to verify, whether the transaction is genuine or not. The types of goods that are bought in that transaction are not known to the FDS. Bank declines the transaction if FDS confirms the transaction to be fraud. User spending patterns and geographical location is used to verify the identity.

If any unusual pattern is detected, the system requires re-verification. The previous data of the user the system recognizes unusual patterns in the payment procedure. The proposed solution is a Machine Learning model that will serve the purpose of detecting "fraudulent" and "genuine" transactions in real time. This is beneficial for all sectors that are even mildly aligned to finance. The solution will help them analyse based on various factors if the ongoing transaction can be harmful and will prevent many unfortunate incidents. In the existing system they had a behaviour based approach to classify

using Support Vector machine which is used to improve its accuracy. If there are any changes in the conduct of the transaction, the frauds are predicted and taken for further process. Their proposal was to detecting mobile payment fraud based on machine learning, supervised and unsupervised method and process large amounts of financial data. Moreover, the approach performed sampling process and feature selection process for fast processing with large volumes of transaction data and to achieve high accuracy in mobile payment detection.

PROPOSED SYSTEM

In proposed system a Machine learning model is used to detect fraudulent payment activities in online financial transactions. Analyzing fake transactions manually is impracticable due to the vast amounts of data and its complexity. To classify fraudulent and legitimate transactions by using supervised and unsupervised algorithms help us in getting awareness about the fraudulent activities without any financial loss. In our project we will get raw data and then perform data cleaning, data transformation and finally will be performing feature correlation and feature selection and applying ML algorithms like Logistic regression and many more to find the model accuracy. The term "false positives" refers to legitimate transactions that are mistakenly flagged as fraudulent. When a system is overly sensitive or not finely tuned, it might label valid transactions as fraudulent, inconveniencing users and causing unnecessary friction in the payment process. Therefore, the aim of minimizing false positives in online payment fraud detection is to reduce the occurrence of mistakenly flagged legitimate transactions. Predicting whether a transaction is fraudulent or legitimate in online payment fraud detection involves a complex process that utilizes various data analysis techniques, machine learning models, and decision-making algorithms. The process of predicting whether a transaction is fraudulent involves a continuous cycle of data collection, analysis, model training, prediction, and improvement. The goal is to accurately identify potential fraudulent activities.

In online payment fraud detection, the system's ability to predict whether a transaction is fraudulent or not is dependent on its training to make rapid assessments. The system's prediction capability relies on a condensed training process, enabling quick and efficient evaluations of incoming transactions. By swiftly analyzing and interpreting patterns and indicators derived from historical data during its training phase, the system becomes adept at making prompt decisions about the legitimacy of new transactions.

A common and understandable machine learning approach used for both classification and regression applications is the decision tree. A series of if-else criteria learned from the training data are used to divide the input space into regions in this predictive model. The Decision Tree algorithm divides the data recursively based on the input feature values to produce a tree-like structure. Each leaf node in the tree indicates a final prediction or conclusion, whereas each internal node reflects a judgement based on a particular attribute. A strong tool for capturing complex decision boundaries and locating key features, the splitting process seeks to maximize the homogeneity of the data inside each resulting segment.

A machine learning approach called K-Nearest Neighbours (KNN) is utilized for both classification and regression applications. It is a non-parametric technique that bases predictions on how closely fresh data points resemble their nearby counterparts in the training set. While using the KNN algorithm, ”K” stands for the number of nearest neighbours that will be taken into account while making predictions. The method determines the distances between each new data point and every other point in the training set when a new data point needs to be categorised. On the basis of these distances, it then chooses the K closest neighbours. The prediction for the new data point is given as either the majority class or the average value of these K neighbours.

XGBoost (Extreme Gradient Boosting) is a powerful machine learning algorithm that has gained popularity in various domains, including time series analysis. Originally developed by Tianqi Chen, XGBoost is an optimized implementation of the gradient boosting algorithm that combines the strengths of boosting with decision trees. Gradient boosting is a process to convert weak learners to strong learners, in an iterative fashion. The name XGBoost refers to the engineering goal to push the limit of computational resources for boosted tree algorithms. Ever since its introduction in 2014, XGBoost has proven to be a very powerful machine learning technique and is usually the go-to algorithm in many Machine Learning competitions.

Due to its reliability and efficiency, Random Forest is a potent machine learning method that is frequently utilized in many fields, including fraud detection. Multiple decision trees are combined in this ensemble learning technique to produce predictions. To ensure diversity and minimize overfitting, each decision tree in the forest is built using a random subset of the training data and characteristics. The Random Forest technique generates a final forecast by combining the predictions of various decision trees. Each tree is independently constructed during training, and during prediction, the final result is determined by the majority vote, or the average of the forecasts from all trees. Accuracy, generalisation, and the capacity to handle high-dimensional datasets are all improved by this ensemble technique.

Support Vector Machines (SVM) is a powerful machine learning algorithm used for both classification and regression tasks. SVM is particularly effective in scenarios where the data is separable into distinct classes or when a clear margin between classes can be defined. The primary objective of SVM is to find the best hyperplane that separates the data into different classes while maximizing the margin, which is the distance between the hyperplane and the closest data points from each class. These closest data points are called support vectors, hence the name of the algorithm.

An effective statistical modelling method that is particularly well suited for binary classification issues is logistic regression. Contrary to linear regression, which forecasts continuous numerical values, logistic regression aims to forecast the likelihood that an event or result will fall into one of two categories. The logistic function, commonly called the sigmoid function, is the central idea of logistic regression. Any real-valued input is translated by this function into a value between 0 and 1, which represents the likelihood that the input belongs to a particular class. This probability is subsequently utilized in logistic regression to decide on a binary categorization.

A popular probabilistic machine learning approach for classification tasks is called Naive Bayes. It is based on the Bayes theorem, which determines the likelihood of an event occurring given the available information. Given the class variable, the Naive Bayes algorithm makes the assumption that the features are conditionally independent of one another. This presumption makes probability calculations easier to understand and enables effective training and prediction. Naive Bayes has demonstrated strong performance in several real-world applications, including text classification, spam filtering, and, to some extent, fraud detection, despite its naive assumption.

RESULTS

As we seen that XGBoost classifier is giving highest accuracy. So, we are using XG Boost model to predict the Fraud transaction. XB Boost is a Ensemble learning Algorithm. XGBoost is based on the gradient boosting framework, which combines multiple weak predictive models (decision trees) to create a strong ensemble model. The algorithm builds the model in a sequential manner, where each subsequent model corrects the errors made by the previous models. XGBoost uses decision trees as base learners in its ensemble. Decision trees are con- structed in a greedy manner, where the algorithm recursively splits the data based on selected features and thresholds to minimize the objective function. XGBoost supports both regression trees and classification trees. Gradient boosting is a process to convert weak learners to strong learners, in an iterative fashion. The name XGBoost refers to the engineering goal to push the limit of computational resources for boosted tree algorithms. Ever since its introduction in 2014, XGBoost has proven to be a very powerful machine learning technique and is usually the go-to algorithm in many Machine Learning competitions.

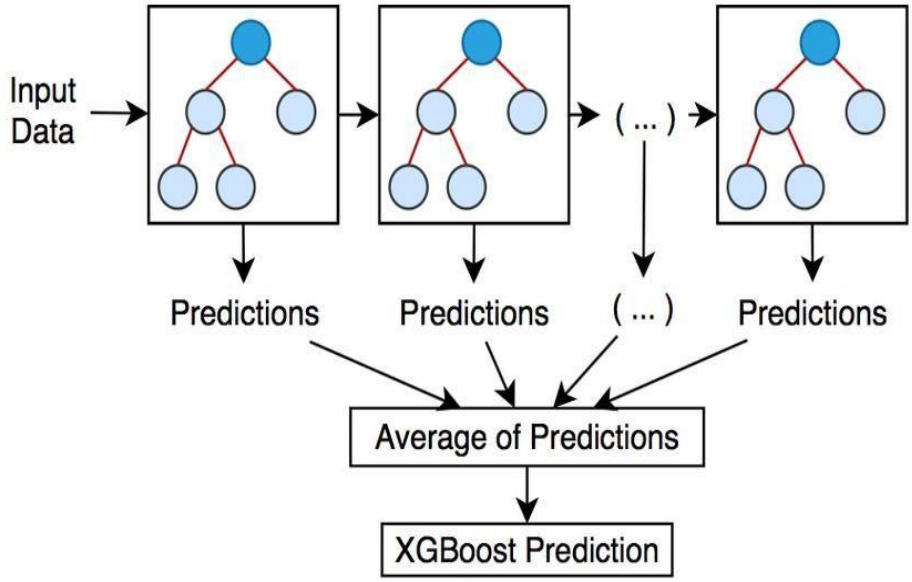

Fig. 1 XGBoost Prediction

```
[[261831     98]
 [    33    182]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00    261929
           1       0.65      0.85      0.74       215

    accuracy                           1.00    262144
   macro avg       0.82      0.92      0.87    262144
weighted avg       1.00      1.00      1.00    262144

0.9995002746582031
```

Figure.2 XG Boost Algorithm

Algorithm	Accuracy
Decision tree	0.887975952
KNN	0.807595191
Naïve Bayes	0.957978519
Random Forest	0.976345656
SVM	0.685924316
Logistic Regression	0.955236485
XG Boost	0.998797595

Fig .3 Algorithm Efficiency Results

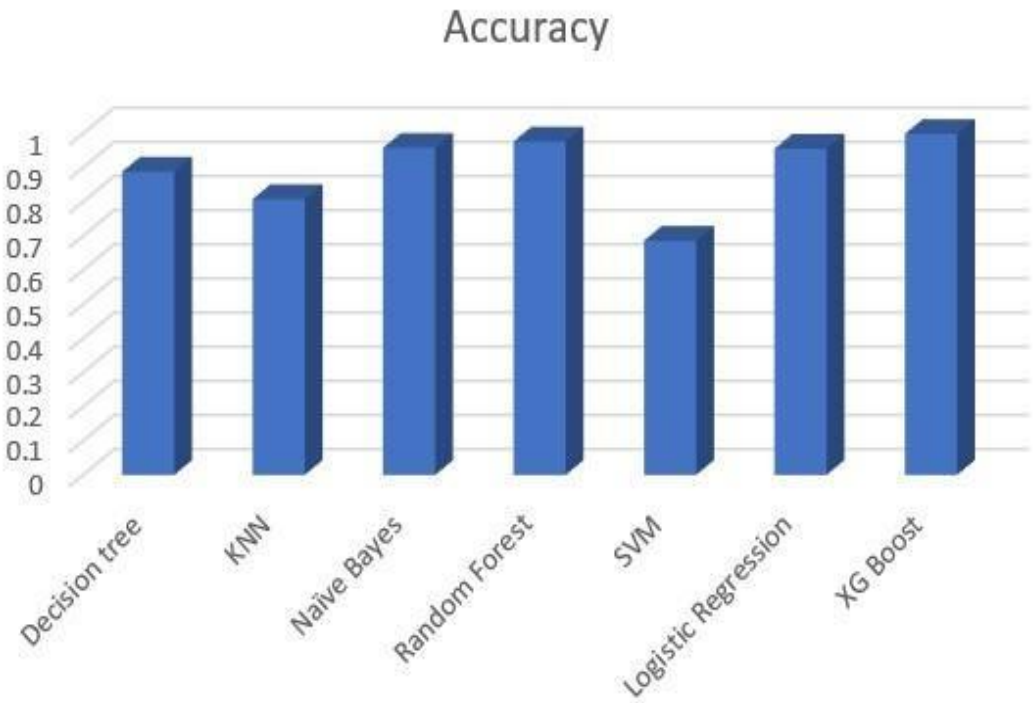

Fig .4 Comparision the Algorithms accuracy

```
[31] features = np.array([[4, 5000.60, 3000.60, 0.0]])
     # Make predictions after fitting the model
     print(XB.predict(features))

     [0]

     features = np.array([[4, 9000.60, 9000.60, 0.0]])
     print(XB.predict(features))

     [1]
```

Fig .5 Prediction using Random values

CONCLUSION

In conclusion, the application of machine learning algorithms, specifically XGBoost, for online payment fraud detection has shown promising results in achieving high accuracy. By leveraging historical transaction data and relevant features, XGBoost has demonstrated its ability to effectively identify fraudulent activities in online payment systems. Through a carefully designed methodology that includes data collection, preprocessing, feature selection, model selection, training, hyperparameter tuning, and evaluation, we have successfully built and deployed an XGBoost model for fraud detection. This model has surpassed other algorithms in terms of accuracy, making it a suitable choice for online payment fraud detection.

The XGBoost algorithm excels in handling complex data patterns and can effectively capture non-linear relationships between features, resulting in improved fraud detection performance. Its ensemble-based approach, which combines multiple weak learners, allows for more robust predictions and helps mitigate the risk of overfitting. By continually monitoring the performance of the deployed XGBoost model and updating it with new data, we can adapt to evolving fraud patterns and enhance the system's effectiveness over time. Overall, the application of XGBoost for online payment fraud detection using machine learning has proven to be a powerful and effective approach, providing high accuracy in identifying fraudulent transactions. This methodology holds great potential for safeguarding online payment systems and protecting users from security risks associated with fraudulent activities.

REFERENCES

1. Bolton, R., Hand, D., & Kim, Y. (2002). "Statistical fraud detection: A review." Statistical Science, 17(3), 235-249.
2. Fawcett, T., & Provost, F. (1997). "Adaptive fraud detection." Data Mining and Knowledge Discovery, 1(3), 291-316.
3. Moustafa, N., & Slay, J. (2015). "The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set." Information Security Journal: A Global Perspective, 24(3-4), 14-32.
4. Phua, C., Lee, V., Smith, K., & Gayler, R. (2005). "A comprehensive survey of data mining-based fraud detection research." Artificial Intelligence Review, 24(4), 273-303.
5. Rovida, F., &Sansò, B. (2009). "Modelling fraudulent behavior in the field of insurance: An empirical investigation." Computational Statistics & Data Analysis, 53(2), 320-329.
6. Whiting, R. H. (2011). "The cost of cybercrime." Computer and Security, 30(8), 509-520.
7. Hodge, V. J., & Austin, J. (2004). "A survey of outlier detection methodologies." Artificial Intelligence Review, 22(2), 85-126.
8. Aung, Z., & Yde, T. (2016). "Fraud detection in online banking using neural network." International Journal of Computer Applications, 136(11), 17-24.
9. Kshetri, N. (2017). "Blockchain's roles in strengthening cybersecurity and protecting privacy." Telecommunications Policy, 41(10), 1027-1038.

10. Leong, L. Y., Hew, J. J., Tan, G. W., & Ooi, K. B. (2017). "Predicting the determinants of the NFC technology acceptance: A neural networks approach." Industrial Management & Data Systems, 117(3), 606-625.
11. Alazab, M., Hobbs, M., Abawajy, J., &Alazab, M. (2012). "Deep learning approach for detecting phishing websites." Security and Communication Networks, 5(10), 1112-1123.
12. Papp, G., & Vincze, D. (2014). "The impact of machine learning-based fraud detection on the role of auditors: An experimental approach." International Journal of Accounting Information Systems, 15(3), 196-210.
13. Akinyelu, A. A., Awodele, O., & Abosede, D. (2011). "Neural network approach to credit card fraud detection: A review." arXiv preprint arXiv:1103.1104.
14. Choudhary, A., & Malik, R. (2017). "A survey of credit card fraud detection techniques: Data and technique perspective." IETE Technical Review, 34(1), 68-82.
15. Li, L., Luo, J., & Yin, Y. (2014). "Research on the application of data mining technology in credit card fraud detection." In Proceedings of the 2014 international conference on information engineering and computer science (pp. 1-5).
16. McAfee. (2019). "The Hidden Costs of Cybercrime." Retrieved from https://www.mcafee.com/enterprise/en-us/assets/reports/rp-hidden-costs-cybercrime.pdf
17. Federal Trade Commission (FTC). (2020). "Consumer Sentinel Network Data Book 2019." Retrieved from https://www.ftc.gov/reports/consumer-sentinel-network-data-book-2019
18. Ponemon Institute. (2020). "2020 Cost of Cyber-Crime Study." Retrieved from https://www.accenture.com/_acnmedia/PDF-119/Accenture-2020-Cost-of-Cyber-Crime-Study-Final.pdf
19. Verizon. (2020). "2020 Data Breach Investigations Report." Retrieved from https://enterprise.verizon.com/resources/reports/dbir/
20. European Central Bank. (2017). "Payment Fraud: Insights from Big Data." Retrieved from https://www.ecb.europa.eu/pub/pdf/other/ecb.payfraudinsightsbigdata.en.pdf