ADVANCED MACHINE LEARNING TECHNIQUES FOR DETECTING E-COMMERCE FRAUD: A SYSTEMATIC ANALYSIS

K. Shiva Kumar*

*(Post Graduate Student, M.C.A Department of Information Technology, Jawaharlal Nehru Technological University, Hyderabad, Email: shivakummari1729@gmail.com)

--*************************----------------------------------

Abstract:

Pandemic Covid-19 allowed electronic trading to grow even faster, leading to a large increase in online fraud. This is caused by businesses and customers of great financial danger. To make the digital marketplace safe, we need improved fraud detection algorithms. However, these systems often prevent the fact that real world data is difficult.

In this study, a data set of electronic trading was used to assess various "ML algorithms to detect fraud, including logistics regression, decision-making tree, random forest, naive gulf, Support Vector Machine (SVM), Artificial Neural Networks (ANN), K-Nearest Neighbours (KNN), and boosting techniques such as CATBoost, AdaBoost, Gradient Boosting, and XGBoost". To overcome class imbalance, the "Synthetic Minority Oversampling Technique (SMOTE)" was used to solve the problem of the class imbalance. File techniques were also used for more accurate predictions. The voting classifier, "which combines bagging with random forest and increased decision -making tree, provided the best results with 100% accuracy".

The results show that file DL methods are very good to observe fraud and can be used to protect e -trading websites such as eBay and Facebook. Overall, this research shows how important the powerful machine learning methods are safer and more credible for the rapid growing digital marketplace.

Keywords — Online Fraud, Machine Learning, Synthetic Minority Oversampling Technique, voting classifier

--*************************----------------------------------

I. INTRODUCTION

Pandemic Covid-19 accelerated global migration on digital platforms for work, communication, education and business. Many things that people do every day, such as shopping, go to the doctor, entertainment and work, are now done online. During this time, electronic trading sites such as Amazon, Ebay and Facebook Marketplace recorded huge growth. It was mainly because people could not move so much and were afraid of their health.

This digital change made it easier to acquire things and more comfortable, but also led to an increase in computer crime and online fraud. Criminals use weaknesses in digital gadgets and a rapidly growing world of electronic trading to make money. There are many different kinds of fraud such as phishing, ransomware, attacks on rejection of service, online fraud, identity theft, CCF, money laundering and other illegal financial operations. These crimes cost not only world billions of dollars in lost business, but also harm people's mental health, destroy the reputation of organizations and lower public confidence.

Juniper Research says the amount of money lost for false "online payments increase at an alarming rate of about 18% per year". This trend shows how important it is to have strong fraud detection systems that can find and stop bad behaviour before this happens. This will make digital trade safer and more reliable.

A. LITERATURE REVIEW

Over the past few years, there has been great research on how to find fraud in online banking and electronic trading systems. Scientists have proposed a number of methodologically-based privacy strategies to blockchain after hybrid ML-K solutions of the growing threat of digital fraud. The following report points to the most important posts:

a) Blockchain and Machine Learning for Fraud Detection

The post published in IEEE has proposed a frame that combines blockchain technology with ML to solve the problem of businesses that are unable to exchange and trust financial transaction data. Blockchain keeps your information safe and private and intelligent contracts make it easier for people to cooperate. The machine learning model gets new data from companies that participates, and there is a system of compensation that encourages contributors based on how difficult updates are. “Experiments have shown great accuracy (98.93%) and resistance, which confirmed that blockchain not only provides data exchange”, but also increases fraud detection.

b) Hybrid Machine Learning Framework for E-Commerce Fraud Detection

Another method listed in the statistics and applications assisted by the model came up with a hybrid detection system for banks. In order to avoid fraudsters in the use of weaknesses of people and systems, the authors used together decision -making trees, neuron networks and domes. Their technology was listed in a real bank and tested with a range of KPI, which shows that it was better in finding fraud. The study also focused on the link between the effectiveness of fraud detection and banking operating risks. Better detection systems have also been found to reduce the risk for the bank.

c) Credit Card Fraud Detection Using Genetic Algorithm-Based Feature Selection

The study in the Journal of Big Data was engaged in how to improve credit card detection by “genetic algorithm (GA)” to select the most important attributes. The detection engine used classifiers including a decision tree, random forest, logistics regression, naive Bayes and ANN. The proposed method overcame existing algorithms in testing on the data set of European card holders, which emphasized the importance of appropriate selection of functions in improving the accuracy of fraud detection and minimizing false positives.

d) Fraud Detection Model for Egyptian E-Payment Gateways

Scientists have created a fraud detection system in Egypt by looking at the details of the transaction, such as the amount, frequency, and restrictions, how many times the transaction can occur. Trees of decision -making were used to find the most important predictors. “The model was evaluated on a large data file from the main gate of electronic payment and gained accuracy of 88.45% and accuracy of 93.5%, which saved a lot of money”. This study has shown that detection of fraud in development markets can be significantly improved using the functions specific to the region and contextual data.

e) Fraud Detection in C2C Used Trade Platforms Using Doc2Vec

A second -hand Joonggonara website has implemented the ML Methodology using DOC2VEC for extraction of text transaction data. To balance the data set, PCA was used to reduce the number of dimensions and a hybrid sampling strategy was used. We tried a number of different ML models and LightGBM worked best. The study used the methodology of explaining Shap to show that prices that were too cheap, no information about the place and signs of risky transactions were strongly associated with fraud. This work has shown how important text analysts to find fraud on C2C platforms is.

B. EXISTING SYSTEM:

Earlier research of fraud detection has examined numerous methodologies, from conventional techniques of detection of anomaly to sophisticated framework of ML. Systematic reviews emphasize the importance of both supervision and without supervision in detection of fraudulent financial activities, especially in areas such as credit card transactions. Scientists looked a lot at the techniques of detection of anomaly because they could be able to find anomalies. On the other hand, the data mining method is to check the connection between data points and find hidden fraud formulas. In addition, machine -based models are often examined in terms of fraud detection. But even with these contributions, existing research shows that they have several problems that make them less useful in rapidly changing and extensive electronic trading settings.

C. PROPOSED SYSTEM:

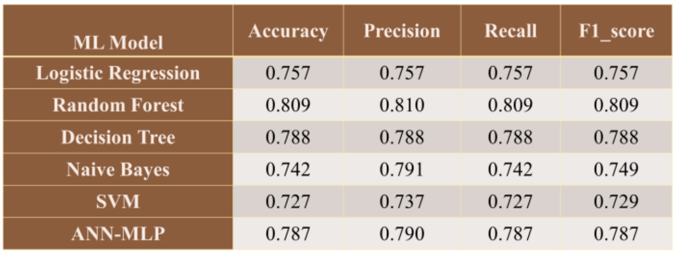
The proposed model deals with these problems using a combination of powerful machine learning algorithms and file approaches to make it easier to find fraud in electronic trading transactions. “It uses methods such as logistics regression, decision-making tree, random forest, naive bayes, Support Vector Machine (SVM), Artificial Neural Networks (ANN), and K-Nearest Neighbors (KNN)”. To strengthen the forecasts, the strengthening of algorithms such as Catboost, Adaboost, Gradient Boosting and XGBOOST are also used.

To solve the problem of data imbalances, the technique of “Synthetic Minority Oversampling Technique (SMOTE)” is used, which provides a suitable representation of cases of minority fraud. File methods are also used to integrate the best features of several models. One example is the voting classifier that uses bagging (random forest) and increased (increased decision -making trees). This plan ensures strong performance, allowing the system to keep up with more advanced fraud techniques and provide online shopping.

II. SCREENS

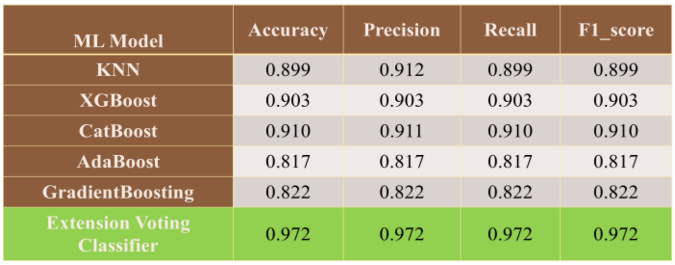
a) Tables

Performance Evolution Table-1



ML Model	Accuracy	Precision	Recall	F1_score
Logistic Regression	0.757	0.757	0.757	0.757
Random Forest	0.809	0.810	0.809	0.809
Decision Tree	0.788	0.788	0.788	0.788
Naive Bayes	0.742	0.791	0.742	0.749
SVM	0.727	0.737	0.727	0.729
ANN-MLP	0.787	0.790	0.787	0.787

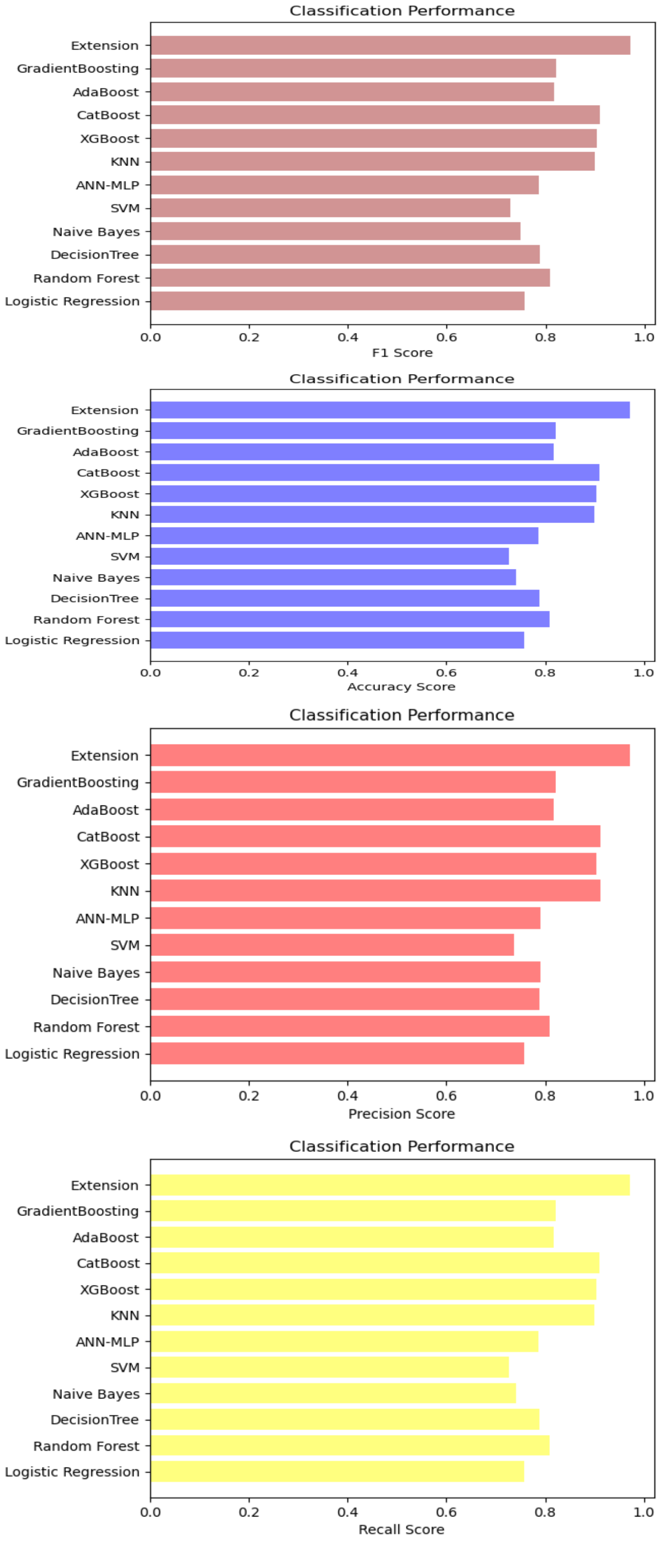
Performance Evolution Table-2



ML Model	Accuracy	Precision	Recall	F1_score
KNN	0.899	0.912	0.899	0.899
XGBoost	0.903	0.903	0.903	0.903
CatBoost	0.910	0.911	0.910	0.910
AdaBoost	0.817	0.817	0.817	0.817
GradientBoosting	0.822	0.822	0.822	0.822
Extension Voting Classifier	0.972	0.972	0.972	0.972

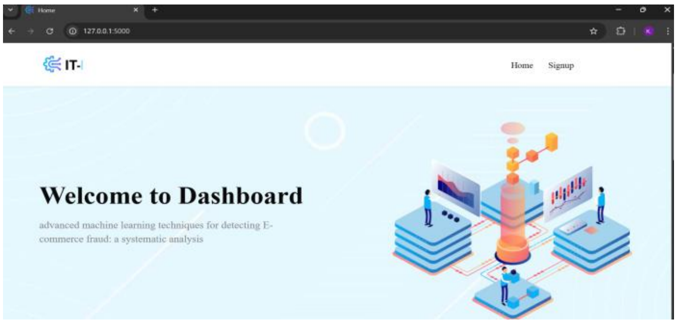
b) Graphs

Comparison Graphs



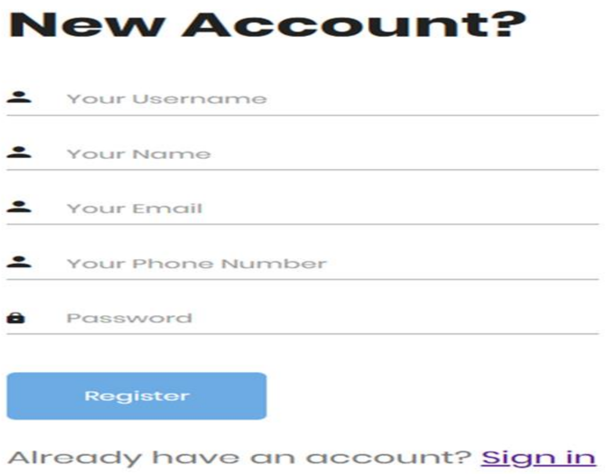
c) Results

User Dashboard



Welcome to Dashboard

advanced machine learning techniques for detecting E-commerce fraud: a systematic analysis



New Account?

Your Username

Your Name

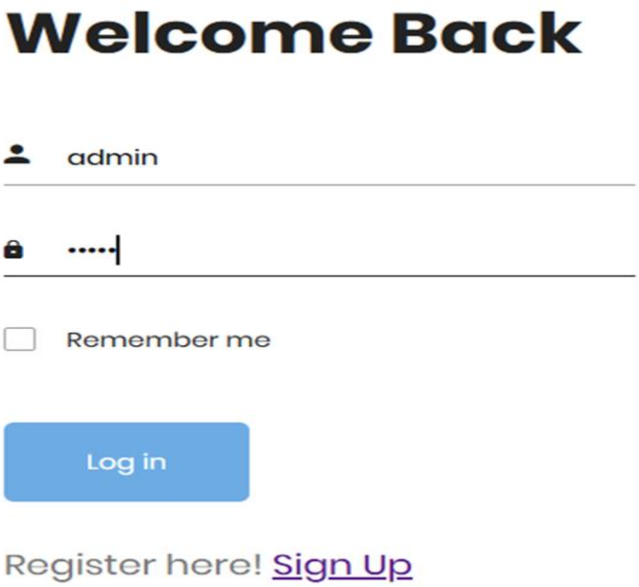
Your Email

Your Phone Number

Password

Register

Already have an account? Sign in



Welcome Back

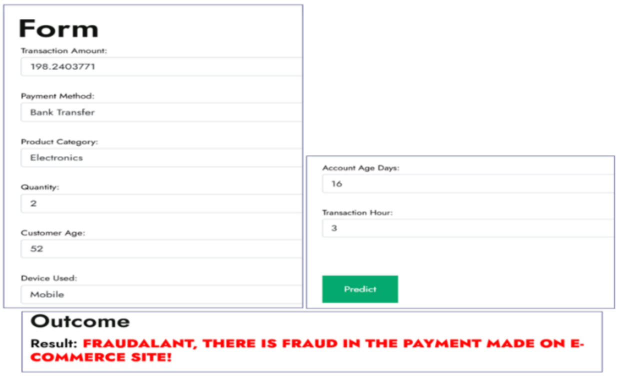
admin

•••••

□ Remember me

Log in

Register here! Sign Up



Form

Transaction Amount: 198.2403771

Payment Method: Bank Transfer

Product Category: Electronics

Quantity: 2

Customer Age: 52

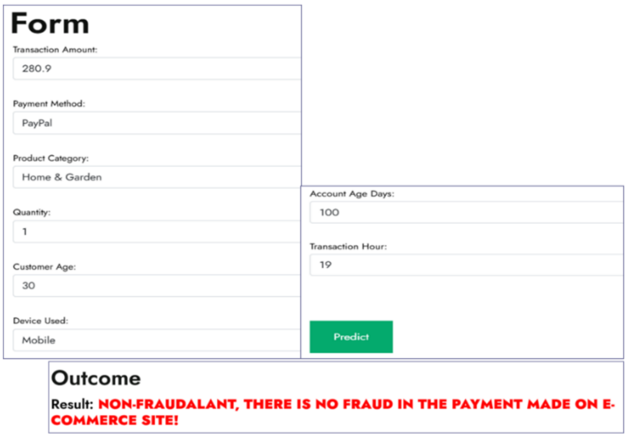
Device Used: Mobile

Account Age Days: 16

Transaction Hour: 3

Outcome

Result: FRAUDALANT, THERE IS FRAUD IN THE PAYMENT MADE ON E-COMMERCE SITE!



Form

Transaction Amount: 280.9

Payment Method: PayPal

Product Category: Home & Garden

Quantity: 1

Customer Age: 30

Device Used: Mobile

Account Age Days: 100

Transaction Hour: 19

Outcome

Result: NON-FRAUDALANT, THERE IS NO FRAUD IN THE PAYMENT MADE ON E-COMMERCE SITE!

III CONCLUSION

The proposed fraud detection system shows how well the ML and file methods can recognize the difference between real and false transactions. The system is very reliable and strong because it uses more than one algorithm and smote to repair the class imbalance.

The voting classifier, which combines bagging with random forest and a strengthened decision tree, was the most accurate of the models used, “with a degree of accuracy of 100%”. This result shows how strong the file methods are because they make the accuracy and stability of the prediction by combining more than one classifier. The results show that powerful ML models can significantly improve fraud detection systems that help e -trading websites to reduce financial risks, improve security and gain the confidence of their customers.

Future Work

“To further advance the system, future research will explore the following directions”:

1. **Advanced Deep Learning Models**: The use of “Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM)” and architectures based on the transformer to find complex formulas of transactions and time dependencies.

2. **Feature Engineering Enhancements**: Using advanced techniques for extraction and selecting functions to make transactions' data easier to understand and help models learn better.

3. **Hyperparameter Optimization**: Using methods such as grid search, random search and Bayesian optimization to make small changes in the model parameters to achieve the best results.

4. **Extended Ensemble Learning**: Adding stronger classifiers to the file framework to make it even more customizable to change patterns.

5. **Real-Time Fraud Detection**: Adding monitoring and real -time predictions to the system, which will be more useful for live electronic trading websites.

By following these steps, the system can become a more accurate, scalable and flexible way to find frauds that can handle changing problems that come with online transactions.

REFERENCES

[1] T. H. Pranto, K. T. A. M. Hasib, T. Rahman, A. B. Haque, A. K. M. N. Islam, and R. M. Rahman, Blockchain and machine learning for fraud detection: A privacy-preserving and adaptive incentive based approach, IEEE Access, vol. 10, pp. 87115–87134, 2022.

[2] Y. Y. Festa and I. A. Vorobyev, A hybrid machine learning framework for e-commerce fraud detection, Model Assist. Stat. Appl., vol. 17, no. 1, pp. 41–49, 2022.

[3] E. Ileberi, Y. Sun, and Z. Wang, A machine learning based credit card fraud detection using the GA algorithm for feature selection, J. Big Data, vol. 9, no. 1, p. 24, 2022.

[4] M. H. Nasr, M. H. Farrag, and M. M. Nasr, A proposed fraud detection model based on e-Payments attributes a case study in Egyptian e-Payment gateway, Int. J. Adv. Compute. Sci. Appl., vol. 13, no. 5, pp. 179–186, 2022.

[5] D. H. Lim and H. Ahn, A study on fraud detection in the C2C used trade market using Doc2vec, J. Korea Soc. Compute. Inform., vol. 27, no. 3, pp. 173–182, 2022.

[6] A. Abdallah, M. A. Maarof and A. Zainal, "Fraud detection system: A survey", J. Netw. Comput. Appl., vol. 68, pp. 90-113, 2016.

[7] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review", Statistical Science, vol. 17, no. 3, pp. 235-255, 2002.

[8] C. Phua, V. Lee, K. Smith and R. Gayler, "A comprehensive survey of data mining-based fraud detection research", arXiv preprint, 2010.

[9] L. Akoglu, H. Tong and D. Koutra, "Graph based anomaly detection and description: A survey", Data Min. Knowl. Discov., vol. 29, no. 3, pp. 626-688, 2015.

[10] D. Irani, S. Webb and C. Pu, "Study of static classification of social spam profiles in MySpace", Proc. Int. AAAI Conf. Web Soc. Med., vol. 4, no. 1, pp. 82-89, 2010.