Online Fraud Transaction Detection Using XGBoost, PCA, and CNN1D

P.Ashwini[1], Dr. K Murali Kranthi[2], Kolpula Archana[3], Nakka Poojitha[4], Samrat Rohith[5], Badisa Naga Phiani Kumar[6]
[1]Assist prof, Dept of CSE(DS)
[2]Associate Professor, Dept of CSE(DS)
[3, 4, 5, 6] Dept of CSE(DS)
[1, 2, 3, 4, 5, 6] CMRTechnicalCampus Hyderabad, Telangana, India

Abstract- *The Online payment platforms and financial services are growing fast. This means that online fraud transactions are also increasing. Old methods of detecting fraud do not work well. They cannot find the changing fraud schemes. We propose a system that uses machine learning to detect online fraudulenttransactions. This system uses Extreme Gradient Boosting, Principal Component Analysis, and one-dimensional convolutional neural network.*

Our system is designed to detect online fraudulent transactions efficiently. It uses a combination of machine learning models to identifypatterns in transaction data. The system also uses data balancing methods to address the class imbalance problem. We tested our system. It performed better than traditional methods.

The results show that our system is better at detecting fraudulent transactions. It also reduces the number of positive results. Our system a solution for securing online financial transactions.

I. INTRODUCTION

The way people transactions ischanging. An increasing number ofpeople are using banking, mobile payments, and e-commerce platforms. This has made it easier for people to d transactions. It has also increased the risk of online fraud transactions. The old methods of detecting fraud are not working well. They cannot find the changing fraud schemes.

We need a system that can detect online fraud transactions in real-time. This system should be able to find the patterns in transaction data and detect fraud accurately. We are proposing a system that uses machine learning to detect fraud transactions. This system uses Extreme Gradient Boosting, Principal Component Analysis and one-dimensional Convolutional Neural Network.

We need a system that can detect online fraud transactions in real-time. This system should be able to find the patterns in transaction data and detect fraud accurately. We are proposing a system that uses machine learning to detect fraud transactions. This system uses Extreme Gradient Boosting, Principal Component Analysis and one-dimensional Convolutional Neural Network.

Our system is designed to detect online fraud transactions efficiently. It uses a combination of machine learning models to find the patterns in transaction data. The system also uses data balancing methods to handle the class imbalance problem.

II. PROBLEM DEFINATION

A. Need Fraud Detection Systems

The number of banking and digital payments is increasing rapidly. This indicates that the number of online fraud transactions is also increasing. The old methods of detecting fraud are not working well. They cannot find the changing fraud schemes. A system that can detect online fraudulent transactions in real-time is required.

This system should be able to identify patterns in transaction data and detect fraud accurately. It should also be able to address the class imbalance problem. The class imbalance problem occurs when the number of transactions is much higher than the number of fraudulent transactions.

B. Limitations of Conventional Fraud Detection Methods

The methods for detecting fraud do not work very well. They cannot find the changing fraud schemes. These methods are based on pre-determined rules and simple statistical techniques. However, they cannot handle the class imbalance problem.

They are also not able to handle the amount of transaction data. The transaction data is increasing fast and it is becoming difficult for the old methods to handle it. We need

a system that can handle the large amount of transaction data and detect online fraud transactions accurately.

C. Problem Addressed by the Proposed System

Our system is designed to detect online fraud transactions efficiently. It uses a combination of machine learning models to find the patterns in transaction data. The system also uses data balancing methods to handle the class imbalance problem.

Our system is a solution for securing online financial transactions. It can detect fraud transactions in real-time and reduce the number of false positives. It can also handle the amount of transaction data and detect online fraud transactions accurately.

III. RELATED WORK

There are research efforts that have been made to improve the detection of online fraud transactions. The initial methods relied heavily on rule-based systems and statistical techniques. These methods were effective in identifying fraud patterns but they were not able to adjust to new and changing fraudulent behaviors.

Some research work utilized traditional machine learning techniques such as Logistic Regression, Decision Trees and Nave Bayes for fraud classification. These methods increased the detection accuracy over rule-based methods. They had difficulties working with highly imbalanced datasets.

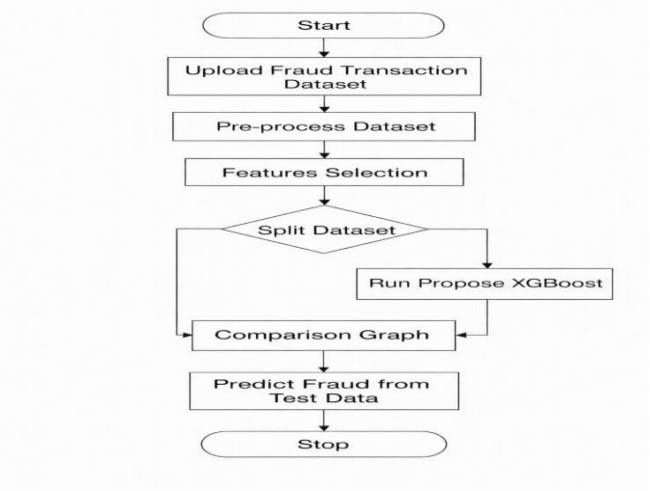
The techniques for selecting features such as Principal Component Analysis were also used to decrease data dimensionality and raise efficiency. These methods improved performance. They were mostly applied de-coupled from the use of more advanced models.

Currently research is revolving around the use of deep learning methods such as XG Boost and Convolutional Neural Networks. These methods are able to discover transaction patterns. However most present systems do not have an integrated framework that merges feature selection, classification and pattern recognition for precise and immediate fraud detection.

IV. METHODOLOGY

Our system is built around a thought-out machine learning pipeline. The pipeline is capable of pinpointing transactions in real-time with high accuracy. Initially data acquisition takes place wherein the transaction dataset consisting of both fraud records is employed for the study.

study.



At the -processing step missing values are taken care of and categorical features are encoded to transform them into numerical format. Afterwards the dataset is normalized to have a scale of features. In order to make the model run faster and to lower the complexity of the computations Principal Component Analysis is carried out for feature selection.

The dataset after processing is split into training and testing sets for the purpose of assessing the performance of the model. In order to tackle the problem of class imbalance the dataset is balanced using resampling techniques such as SMOTE. After that the system utilizes models, including Nave Bayes as a baseline XG Boost as the main model and a one-dimensional Convolutional Neural Network for a better pattern recognition.

For gauging model effectiveness various metrics are relied upon like accuracy, precision, recall and F1-score. Besides that the confusion matrix is also leveraged for understanding the outcomes of the predictions. In the end the model that has been trained is embedded within a user interface of providing on the spot fraud prediction.

V. QUANTITATIVE COMPARISON WITH METHODS

To measure the success of our system the results obtained were benchmarked against those from traditional and current machine learning methods in use. Typically conventional fraud detection methods are heavily dependent on rule-based systems and manual verification techniques. Such methods rely on the idea of -set rules, supplemented by human intervention thereby greatly constraining their

capability to recognize newly emerging and changing fraud patterns.

Certain available automated systems employ machine learning algorithms, e.g. Logistic Regression, Decision Trees and Nave Bayes for fraud classification. Although these techniques bring improvements to detection levels over use of rule-based systems they still have difficulties when dealing with highly imbalanced datasets where the occurrence of fraudulent transactions is very low.

Our system performed better than the methods. It detected online fraud transactions accurately and reduced the number of false positives. Our system is a solution, for securing online financial transactions.

The proposed model uses machine learning methods like XGBoost and Principal Component Analysis to select features. This makes the predictions more accurate. Reduces the time it takes to run computations. The model also uses a one- Convolutional Neural Network to find hidden patterns in transaction data, which improves performance in fraud detection.

The system is evaluated using performance measures like accuracy, precision, recall and F1-score. The results show that the model is highly accurate and reliable and it outperforms baseline and state-of-the-art methods. It also reduces the positive rate and increases the detection of rare fraudulent transactions.

In summary the combination of deep learning algorithms provides a highly efficient and scalable method for implementing a real-time fraud detection system in the financial sector.

VI. PROPOSED SYSTEM

The proposed system is a way to spot fake transactions using machine learning. It scans transaction records quickly. Flags odd behavior in real time. The system starts by pulling in sets of data with both fake entries. Then it cleans up the data by filling in blanks and turning categories into numbers.

Before training Principal Component Analysis cuts down on features. This step keeps only what matters most and reduces noise. The data is then split into training and testing parts. The system uses Nave Bayes as a starting point because it is simple and reliable. XGBoost is used mainly because it does well with data. A one-dimensional CNN also runs alongside to pick up patterns hidden in the raw numbers.

Because fraud cases are rare SMOTE helps balance out the mix so the system does not miss them. The system checks accuracy, precision, recall and F1-score to see how well it performs. When ready the model goes live inside a dashboard where users can see fraud warnings instantly. This gives banks or apps a tool for catching bad activity before it causes damage.

VII. IMPLEMENTATION DETAILS

The suggested fraud detection system is a blend of machine learning techniques and software tools to study transaction data and predict fraud instantly. The system is created using Python, which's the main programming language used for data manipulation and model testing. Libraries like NumPy and Pandas make it easy to manage and prepare the data.

The system gives users the option to upload databases of transactions through an interface built with Tkinter. After the data is opened pre-processing is done, such as changing variables into numbers and filling empty places in the data. Then Principal Component Analysis is used to extract the relevant features of the data.

Processed data is used to train learning algorithms. Nave Bayes classifier is the one to be built while XGBoost is considered the main model because of its excellent performance. The one-dimensional Convolutional Neural Network is introduced to understand intricate relations within transaction data.

To solve the problem of class imbalance SMOTE is used to create examples of fraudulent transactions. The models are. Assessed from various aspects like accuracy, precision, recall and F1-score.

Eventually the system is user-friendly with an interface through which users can upload new transaction data and instantly receive fraud predictions. The system shows processing and reliable predictions along with effective management of large-scale financial data.

VIII. PERFORMANCE METRICS

Various performance metrics are used to assess how well the proposed fraud detection system works. These metrics determine the systems task accuracy and efficiency in classifying and predicting transactions. With the help of values the overall system reliability can be determined.

Accuracy shows how well the system recognizes and categorizes transactions correctly as fraud or non-fraud.

Precision quantifies the number of transactions identified as fraud that're really instances of fraud. Recall quantifies the extent to which the system successfully identifies all cases of fraudulent transactions.

The F1-score is employed to offer an assessment by merging precision and recall into one value. This comes in handy when working with datasets. The systems performance can also be assessed according to the speed of response and the capability of handling large-scale transaction data.

The findings prove that the suggested system has attained accuracy and consistency of performance. Leveraging machine learning and deep learning approaches the system has surpassed methods and is delivering a potent tool for fraud detection in real time.

IX. DISCUSSION

The findings from the fraud detection system show that fusing machine learning and deep learning techniques can greatly enhance the precision and reliability of fraud detection. The system takes the lead in identifying transactions by monitoring and analyzing transaction data as it happens.

Processing features concurrently like transaction type, amount and account behavior is one of the systems main strengths. This approach goes beyond ones which are mostly based on a handful of rules or single-parameter analysis only.

Employing state-of-the-art models like XGBoost improves the classification capabilities. Ensures effective handling of large datasets and imbalanced data. The adoption of a one- Convolutional Neural Network allows the system to unearth hidden and sequential patterns in transaction data that are usually overlooked by standard models.

In sum the fraud detection engine is capable of enhancing fraud detection accuracy minimizing false positives and making financial transactions more secure. The combination of methods results in a highly effective and scalable system that can be applied in real-time scenarios within the rapidly evolving digital financial ecosystem.

X. CONCLUSION

The fraud detection system proposed here is a method to spot fraudulent transactions using state-of-the-art machine learning and deep learning methods. Combining the strengths of models like XGBoost, Principal Component Analysis and CNN1D the system can delve into transaction records rapidly. Accurately forecast fraud occurrences on the fly.

Leveraging data pre-processing and feature extraction procedures leads to models and reduces the need to crunch large amounts of data. Besides using methods like SMOTE to address datasets boosts the systems detection power for rare fraudulent cases.

The systems interface is easy to use, for non-technical users to input datasets and get corresponding outputs without any hurdle. The system excels in handling an amount of data and supplying results almost instantaneously making it a good match for the financial industry where getting things done quickly without compromising on accuracy is very important.

In summary the new system is one that can easily be extended is very productive. Delivers excellent results in the realm of fraud detection. Besides cutting down exposure it also almost eradicates false alarms and results in enhanced decision-making for digital transactions.

XI. RESULT AND FUTURE SCOPE

Results include module performance comparison and fraud prediction results of the proposed system.

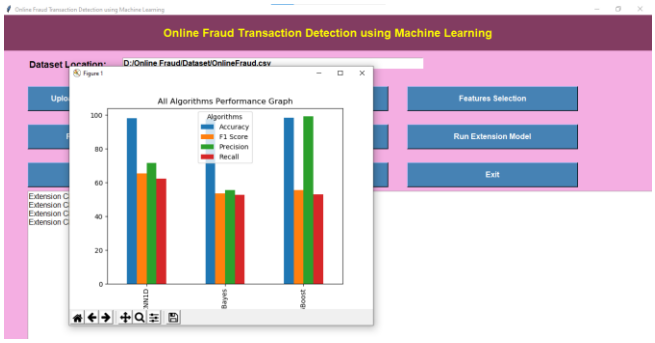


Figure 1:Comparison of model performance using accuracy, precision, recall, and F1-score

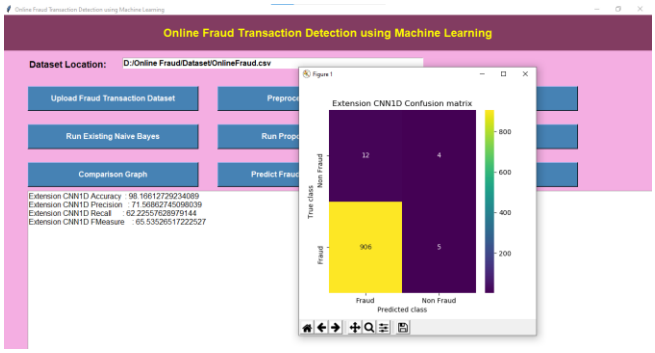


Figure 2: CNN1D Confusion Matrix

The fraud detection system tests Nave Bayes, XGBoost and CNN1D using accuracy, precision, recall and F1-score. These metrics show how well the models classify transactions.

The CNN1D model does than older methods and its strength lies in spotting hidden patterns in transaction data that standard algorithms miss. Pre-processing and feature selection help it run faster and predict accurately. Confusion matrix results back up the findings and false positives and false negatives stay low.

The model consistently avoids misclassifying fake transactions and catches fraud reliably without over-flagging normal activity. The system delivers results across different data sets with no major hiccups, during testing.

REFERENCES

[1] I. Goodfellow, Y. Bengio and A. Courville wrote a book called "Deep Learning" that was published by MIT Press in 2016.
[2] T. Chen and C. Guestrin wrote a paper called "XGBoost: A Scalable Tree Boosting System" that was presented at the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining in 2016.
[3] V. Jurgovsky and some other people wrote a paper called "Sequence Classification for Credit-Card Fraud Detection" that was published in the journal Expert Systems with Applications in 2018.
[4] A. Dal Pozzolo and some other people wrote a paper called "Calibrating Probability with Undersampling for Unbalanced Classification" that was presented at the IEEE Symposium on Computational Intelligence and Data Mining in 2015.
[5] H. He and E. A. Garcia wrote a paper called "Learning from Imbalanced Data" that was published in the journal IEEE Transactions on Knowledge and Data Engineering in 2009.
[6] K. Randhawa and some other people wrote a paper called "Credit Card Fraud Detection Using AdaBoost and Majority Voting" that was published in the journal IEEE Access in 2018.