International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Machine Learning techniques for detection of online payment fraud

Fazila Shariff[1] , Gangu Anusha[2] , Gandeti Dilleshwari[3] ,Mrs.S.Shanmathi[4]

B.tech ,Rajam 532127 , India

ABSTRACT :

This abstract discusses an innovative artificial intelligence (AI) model designed to protect against financial fraud, especially in the context of real-time financial transactions. The model, called RXT, combines advanced machine learning techniques to process financial data, detect fraud, and ensure the security of financial transactions.

Here's a breakdown in simpler terms :

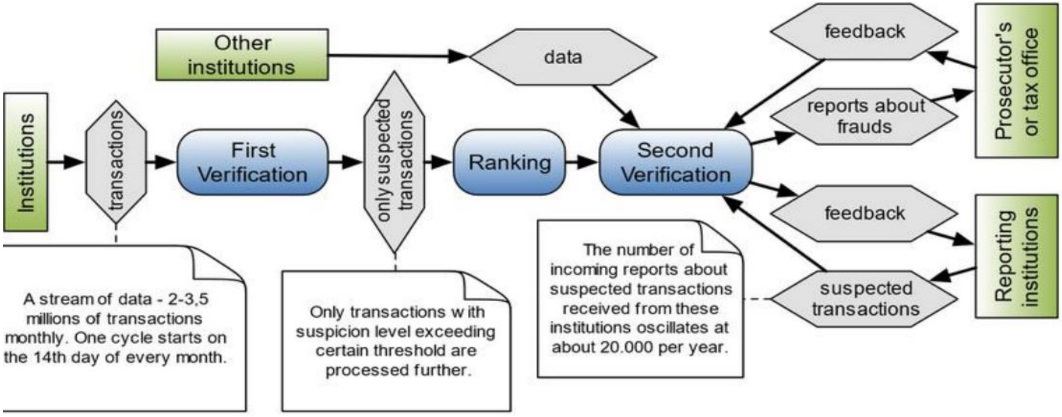
AI Approach: The solution proposed is an AI model called Res-NeXtembedded Gated Recurrent Unit (RXT), which is specifically designed to detect fraud in real-time transactions. The model uses several AI techniques to process data, improve accuracy, and optimize performance.

Keywords: Fraud detection, Deep Learning, Fraud Detection, Financial transaction fraud.

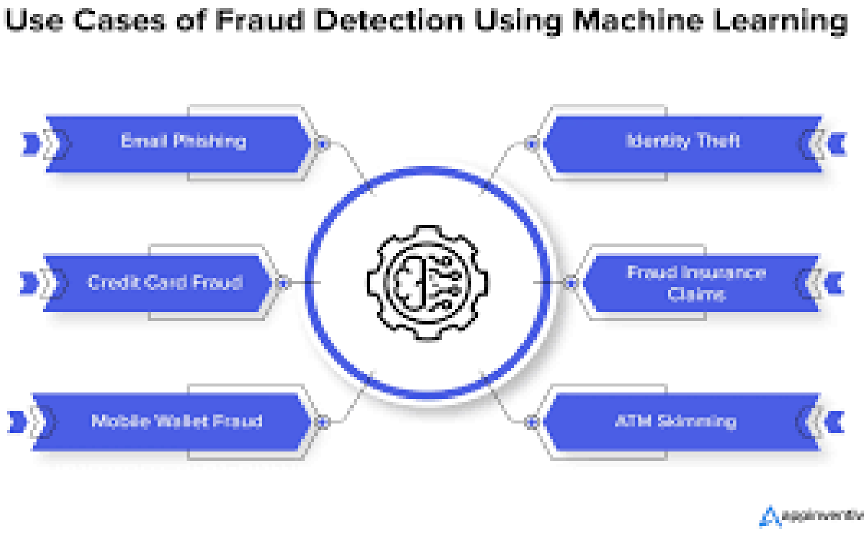
1. Introduction :

Online payment fraud detection using machine learning techniques is a fascinating and crucial area of research. As online transactions continue to grow, so does the risk of fraudulent activities1. Machine learning models can analyze vast amounts of transactional data, user behavior patterns, and contextual information to identify and prevent fraudulent transactions in real time . These models use various algorithms, such as logistic regression, random forest, and neural networks, to detect anomalies and patterns indicative of fraud. By continuously learning from new data, these systems can improve their accuracy and adapt to evolving fraud techniques.

The goal is to protect both businesses and consumers from financial losses and maintain trust in online payment systems. It's a dynamic field that combines data analytics, machine learning, and cybersecurity to create robust fraud detection mechanisms.


Other institutions
data
feedback
Prosecutor's or tax office
Institutions
transactions
First Verification
only suspected transactions
Ranking
Second Verification
reports about frauds
feedback
Reporting institutions
suspected transactions
A stream of data - 2-3,5 millions of transactions monthly. One cycle starts on the 14th day of every month.
Only transactions w ith suspicion level exceeding certain threshold are processed further.
The number of incoming reports about suspected transactions received from these institutions oscillates at about 20.000 per year.

Our model excels in real-time scenarios, delivering rapid and precise results crucial for timely fraud detection. Notably, it exhibits remarkable resilience when faced with data size scalability challenges, ensuring consistent and reliable performance across a diverse range of datasets. This capability makes it a valuable tool for real-world financial applications, where timely and accurate fraud detection is paramount.



This comprehensive approach encompasses feature engineering, dimensionality reduction, and a novel classification method, ensuring the effective identification of fraudulent activities while maintaining operational efficiency. Initially, we have taken dataset features as input. We have applied the preprocessing step, including filling in missing values, normalization, etc. Given the class imbalance in the dataset, where fraudulent activity accounts for only 3.27% of all transactions, addressing this imbalance is crucial for robust model training. To mitigate class imbalance, we commence with data preprocessing using the Synthetic Minority Over-sampling Technique (SMOTE) algorithm. SMOTE is applied to balance the class distribution in the dataset, generating synthetic samples for minority classes and thus enhancing the overall dataset balance. After this critical preprocessing step, we proceed to the feature extraction phase, employing the Ensemble Auto Encoder with Res Net (EARN) model.

In recent years, the proliferation of online payment systems has revolutionized the way we conduct financial transactions. As digital payments become increasingly prevalent, the risk of fraudulent activities has also surged, posing significant challenges to businesses and consumers alike. The need for robust fraud detection mechanisms has never been more critical. Traditional fraud detection methods often fall short in handling the scale and complexity of modern-day transactions. This has paved the way for advanced techniques, particularly machine learning, to play a pivotal role in combating online payment fraud.

Machine learning models offer a dynamic and adaptive approach to fraud detection. Unlike traditional rule-based systems, which rely on predefined patterns and heuristics, machine learning algorithms can learn from vast amounts of data, uncovering intricate patterns and correlations that may indicate fraudulent behaviour. These models can analyses numerous variables and continuously update their understanding of what constitutes normal and abnormal transaction patterns. By doing so, they enhance the accuracy and efficiency of fraud detection, minimizing false positives and negatives.

Several machine learning techniques have shown promise in detecting online payment fraud. **Supervised learning algorithms**, such as logistic regression, decision trees, and support vector machines, utilize labelled data to train models that distinguish between legitimate and fraudulent transactions. **Unsupervised learning methods**, including clustering and anomaly detection, identify unusual patterns without prior knowledge of fraud. These methods are particularly useful in detecting novel fraud schemes that deviate from known patterns. **Deep learning models**, such as neural networks, further enhance fraud detection capabilities by processing complex, high-dimensional data and extracting meaningful features automatically.

The application of machine learning in fraud detection goes beyond the mere identification of fraudulent transactions. It encompasses the entire lifecycle of fraud prevention, from real-time monitoring and risk assessment to adaptive response strategies. **Real-time monitoring systems** powered by machine learning can analyses transaction data as it occurs, providing immediate alerts for suspicious activities. **Risk assessment models** evaluate the likelihood of fraud based on historical data and contextual information, enabling proactive measures to prevent fraud before it occurs. **Adaptive response strategies** leverage machine learning to continuously refine detection models, incorporating feedback from confirmed fraud cases and emerging fraud trends.

Despite its potential, the implementation of machine learning-based fraud detection systems is not without challenges. One of the primary hurdles is the **availability and quality of data**. Machine learning models require large, diverse datasets to learn effectively, yet obtaining such data can be difficult due to privacy concerns and regulatory constraints. Additionally, **fraudsters constantly evolve their tactics**, necessitating continuous model updates and adaptations. Ensuring the **scalability and real-time performance** of machine learning systems is also crucial, particularly in high-volume transaction environments.

By leveraging the power of data analytics and artificial intelligence, these models offer a more flexible, accurate, and efficient means of identifying and preventing fraudulent activities. As online payment systems continue to evolve, so too must the strategies to protect them. Ongoing research and development in machine learning and fraud detection will be essential to staying ahead of increasingly sophisticated fraudsters, ensuring the security and trustworthiness of digital financial transactions.

2.Literature Survey

1. 1.The Future of Money: How the Digital Revolution is Transforming Currencies and Finance.
2. Online payment fraud detection Model Using Machine Learning.

3. Evaluation of deep networks for reducing Credit Card Fraud Alerts.
4. 4.Financial inclusion for SMEs: Role of digital micro-financial services.
5. 5.Graph computing for financial crime and fraud detection: Trends challenges and outlook.
6. 6.Fraud detection in mobile payments utilizing process behaviour analysis.
7. 7.Understanding consumer acceptance of mobile payment services: An empirical analysis.
8. 8.Determining the antecedents of mobile payment loyalty: Cognitive and affective perspectives.
9. 9.Rule-based credit card fraud detection using user’s keystroke behaviour" in Soft Computing: Theories and Applications.
10. 10.Identity application fraud detection using web mining and rule-based decision tree.

3.Methodology :

The Paper describes a system designed to detect fraudulent transactions, such as credit card fraud, using advanced machine learning techniques. Here’s a breakdown of the process:

1]. Preprocessing the Data:

- The raw transaction data is first cleaned and prepared. This involves handling missing values (filling in gaps where data is missing) and normalizing the data (scaling the values to make them consistent).

- Fraudulent transactions make up only a tiny portion (3.27%) of the total transactions to make sure the model learns to detect fraud effectively, despite the imbalance in the data.

2]. Feature Extraction with EARN Model:

- After the data is prepared, the next step is to extract useful patterns or features from the data that help in identifying fraud. For this, the “EARN model” is used, which combines two powerful methods:

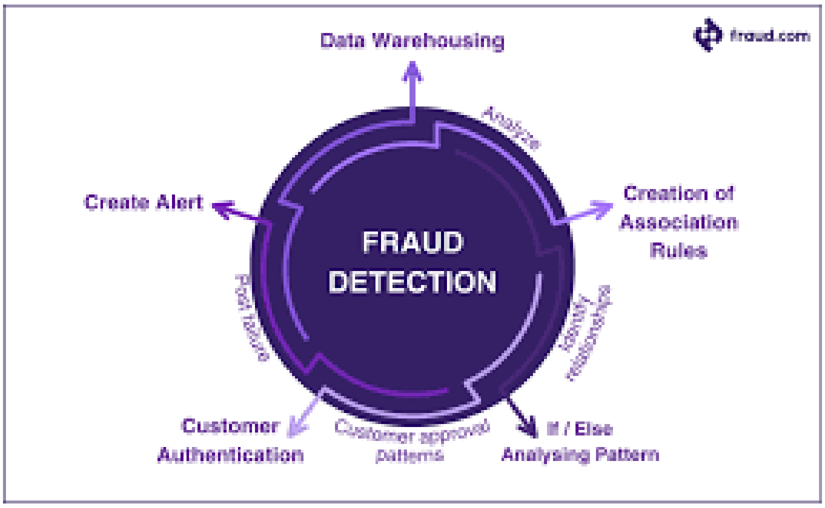
- Auto-encoders: These are neural networks that compress data and help extract important patterns.

- “NEXT”: A deep helpful when fraud patterns can vary a lot.

3. Goal of the System:

- The goal is to build a system that can accurately detect fraud by identifying key patterns in the data.

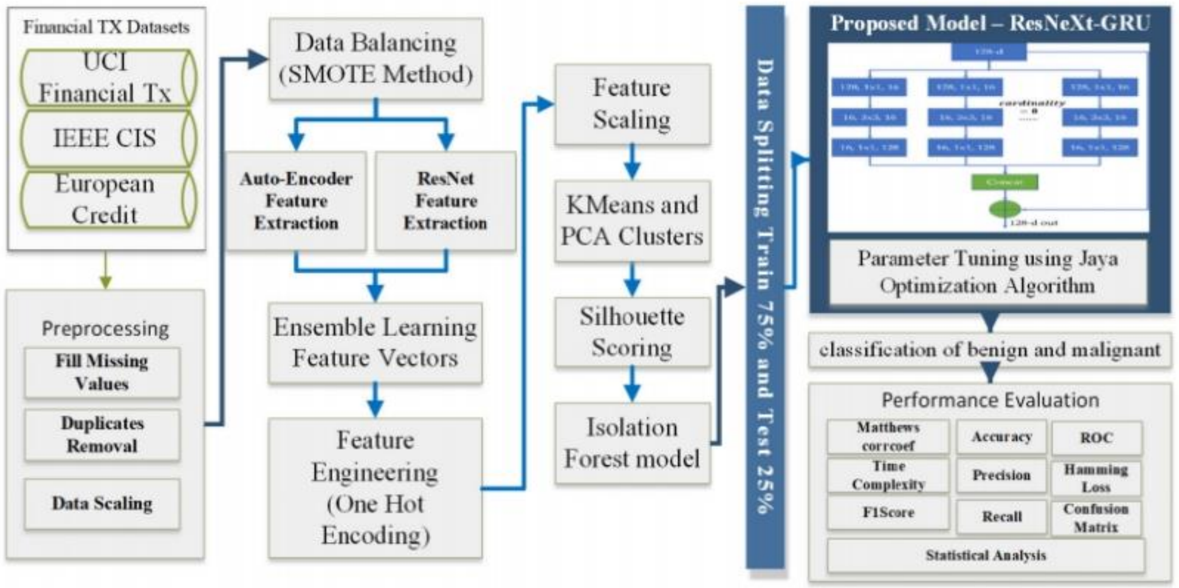
At the same time, it needs to work efficiently and be easy to understand for financial institutions, who need to trust and interpret the results.



TheResNeXt-GRU model represents an advanced and versatile architecture designed to classify financial transaction data, particularly in detecting fraudulent activities .This model amalgamates the robust feature extraction capabilities of the architecture with the sequential learning and count extra understanding existing by GRUs.

The core of our approach is the RXT model, which we fine-tune using a method called the Jaya optimization algorithm (RXT-J) to get the best performance. We tested our model on three real-world financial datasets and found that it consistently outperforms other methods by 10% to 18% in various performance measures while remaining efficient in its calculations.

A deep learning model that looks for more complex, high-level features in the data. The EARN model uses several different autoencoder and models with different setups to capture a wide range of patterns and behaviours, which is particularly helpful when fraud patterns can vary a lot.
Strategically leveraging cardinality-based segmentation, the data is divided into multiple paths to augment the model’s capacity for capturing a wide range of features. Independently, each path processes the data through the architecture .These path outputs ,now enriched with distinct information, are concatenated to form a comprehensive features.



Proposed System Model

The goal of this system is to detect fraudulent financial transactions, like credit card fraud, in an efficient and accurate way. The model combines different techniques to handle the challenges of fraud detection, such as dealing with imbalanced data, reducing complexity, and improving accuracy.

Steps in the Process:

1. **Data Preprocessing :**The first step is to clean the data. This involves filling in any missing information, normalizing the data (so all values are on a similar scale), and addressing the problem of **class imbalance**. Since only a small percentage (about 3.27%) of transactions are fraudulent, this imbalance can make it harder for the model to identify fraud correctly. To fix this, we use a technique called **SMOTE (Synthetic Minority Over-sampling Technique)**, which generates extra examples of fraud (the minority class) to balance the dataset.
2. **Feature Extraction (EARN Model):**
   After preprocessing, we move on to **feature extraction**, where we focus on identifying the most important information in the data. We use a combination of two models for this: **autoencoders** and **ResNeXt**. These models help us reduce the complexity of the data while keeping the most relevant features that can help us identify fraud.

- **Autoencoders** are a type of neural network that learns to compress data into a smaller representation (called the encoding) and then reconstruct it back to the original form. This helps identify patterns in the data that are important for fraud detection.
- **ResNeXt** is another type of deep learning model that helps capture more complex features of the data .We use **multiple autoencoders and ResNeXt models** with different settings to capture a wide range of features, both simple and complex. These models are trained to reduce the data's dimensionality (or number of features) while keeping the most important ones for fraud detection.
- **Feature Fusion (Ensembler Learner):**
  Once we've extracted features using the autoencoders and ResNeXt models, we combine (or **fuse**) these features into a single set. This creates a more complete and rich set of features that can be used for the next step: classification.
- **Value-at-Risk (VAR):**This method is used to model the high-risk fraud features that are rare but extremely costly when they occur. Unlike traditional methods that apply a constant probability weight to fraud, VaR adjusts the fraud risk based on a confidence level. This helps to deal with the imbalance in fraud cases (where fraud is rare) and prevent the machine learning model from being biased by more common, non-fraud cases.
- **Nearest Neighbours (KNN):** This is a distance-based algorithm used to identify fraud by finding clusters of rare fraudulent instances. KNN works by looking at the nearest neighbours (similar data points) to detect fraud. Since fraud is rare, the model is adjusted to focus more on these rare instances by setting the right value for "k" (the number of nearest neighbours to consider). This helps the model focus on the higher-risk fraud cases and improves its ability to detect these rare instances.
- **Adjusting KNN for Skewed Data:** The research adjusts the KNN algorithm by assigning higher weight to data points that are closer to the fraudulent instances. This helps the model to focus on the important, rare fraud cases and not be distracted by the majority of non-fraud cases.This research combines VaR and KNN to better handle rare, high-risk fraud cases by adjusting the model's sensitivity to these instances, making it more effective in fraud detection.
- **Statement of the Problem :** In our research, several deep neural network architectures were trained to automatically discard alerts associated with false positives, thus effectively reducing the number of alerts requiring manual investigation.
- **Dataset Description:** The dataset used for our research contained 446,076 real alerts related to suspicious credit card transactions. They were obtained from a Spanish payments processing organization. The alerts in this dataset span a period of six months.

- **Neural Network Architectures :** The set of neural network architectures tested in our research belong to one of the following types: Multi-Layer Perceptron (MLP) : MLP is type of feedforward, fully connected, neural network with three or more layers of neurons. It uses backpropagation for learning. It contains one input layer, one or more hidden layers, and an output layer. Each neuron applies a non-linear activation.
- 2 Convolutional Neural Network (CNN) . CNN also has an input layer, several hidden layers and an output layer. Hidden layers implement sequences of convolution and max pooling operations and are followed by a fully connected layer. A convolution operation is a sliding dot product of the input and (rectifier linear unit) as non-linear activation function. Max pooling does non-linear down-sampling.

Classification with ResNeXt-GRU (RXT-J)

The **ResNeXt-GRU** model is used to classify whether a financial transaction is fraudulent or not. It's a powerful combination of two machine models: **ResNeXt** (for feature extraction) and **GRU** (for sequential learning).

1. Feature Extraction with ResNeXt

- **ResNeXt** helps in extracting useful features from the raw financial data. It does this by processing the data through several **convolutional layers** and using techniques like **batch normalization** and **activation functions** (like ReLU) to refine the data.
- **Path Cardinality:** ResNeXt splits the data into multiple **paths**, where each path processes the data separately. The outputs from these paths are then combined (concatenated) to form a complete set of features for classification. This approach helps the model capture more detailed and diverse information from the data.

Concatenation Formula:

Concatenation = [RFU1, RFU2, RFU3, RFU4]

Here, RFU1, RFU2, RFU3, and RFU4 are the outputs from different paths. These are combined to form one unified feature.

2. Sequential Modeling with GRU

- **GRU (Gated Recurrent Unit)** is used after the features have been extracted by ResNeXt. It helps the model understand the sequence or **temporal dependencies** between transactions. In simpler terms, GRU helps the model recognize patterns over time, which is essential for detecting fraud, as fraudulent behaviour often follows certain patterns.
- **How GRU works:** The GRU uses several operations (gates) to control how information flows between different time steps:
- **r (reset gate)**: Decides how much of the past information to forget.
- **z (update gate)**: Determines how much of the new information should be kept.
- **h˜ (candidate state)**: Represents potential new information.
- **h (hidden state)**: The output of the GRU, which is a blend of the previous state and the new state.

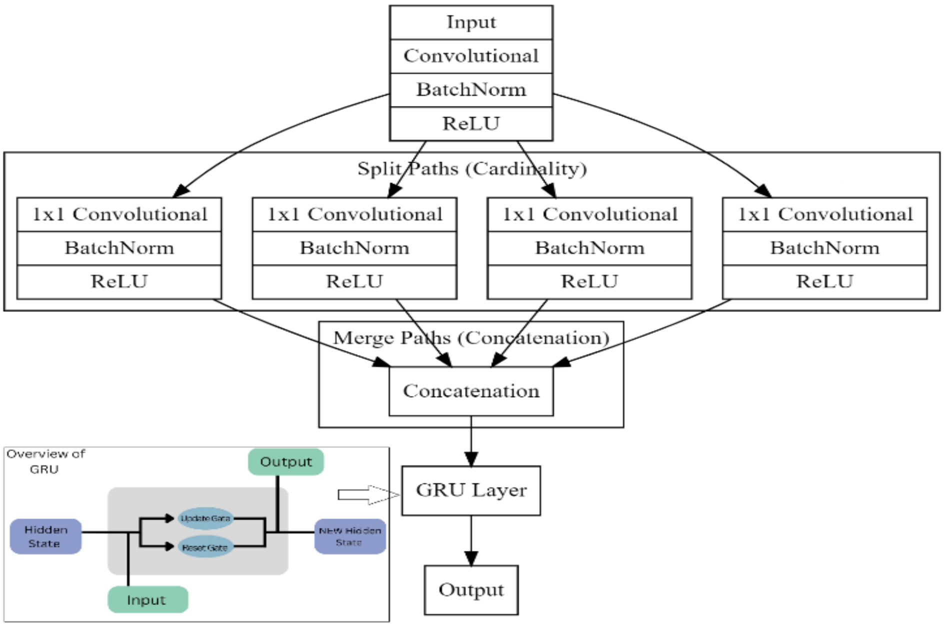


Figure1: ResNeXt Model Fraud Transaction

Random forests examine transaction data to find fraudulent trends in credit card fraud detection. Data collection and preparation, including features like transaction amount and location, is the first step in the process. By building several decision trees, each with random subsets of data and attributes to increase resilience, a random forest model is trained using historical data that has been classified as either authentic or fake. After training, the model uses the trees' majority vote to categorise future transactions.

Analysts can comprehend important patterns in fraud detection by using an ensemble approach, which enhances generalisation and decreases overfitting while offering insights into feature importance. All things considered, random forests successfully use past trends to forecast fraudulent transactions with high accuracy.

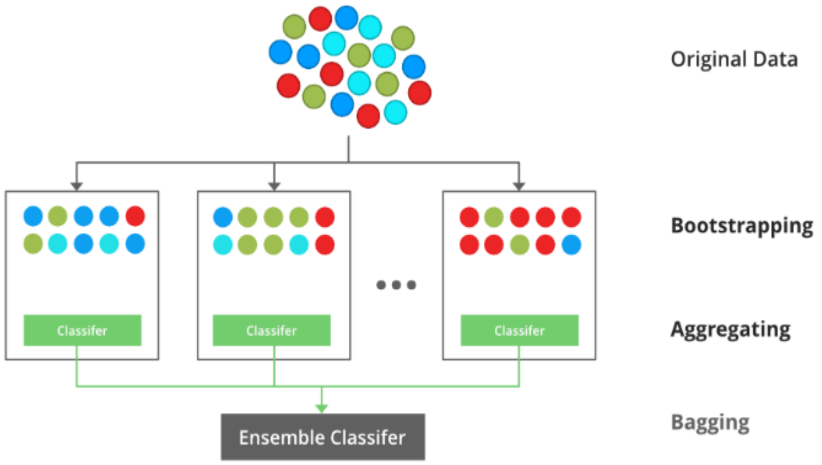
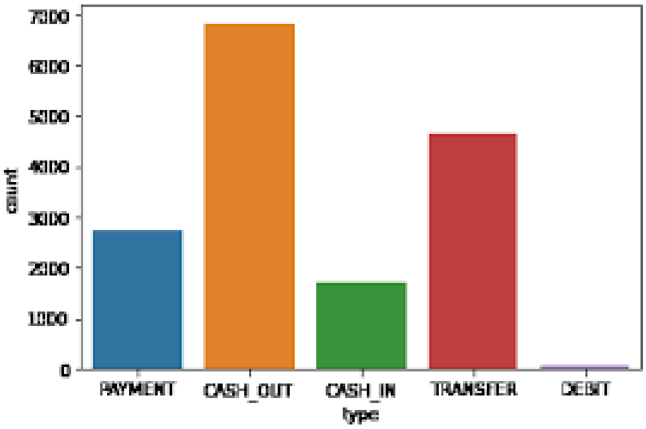


Figure 10: framework of XG Boost

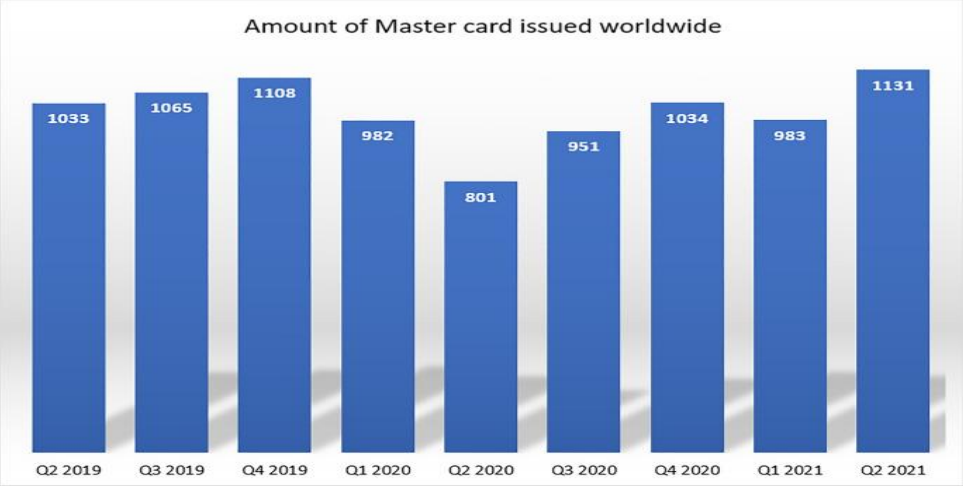
Based on past data, XG Boost is used in credit card fraud detection to categorize transactions as either fraudulent or lawful. First, transaction features including amount, time, location, and user behaviour are gathered and preprocessed. This labelled data is used to train XGBoost, which builds decision trees one after the other to fix mistakes from earlier trees and uses gradient boosting to increase accuracy. XGBoost modifies class weights or uses subsampling techniques to alleviate class imbalance in situations when fraudulent transactions are uncommon. After training, the model flags transactions that above a predetermined threshold and forecasts the likelihood that new transactions would be fraudulent. Furthermore, by offering insights about feature importance, XG Boost helps analysts comprehend the elements that lead to fraud. Because of its great accuracy, effectiveness with big datasets, and capacity to provide insightful information.

RESULTS AND DISCUSSION:

According to a study, machine learning algorithms attained up to 96% accuracy in decreasing ecommerce fraud. For example, Zoho Payments has proven in-house models deployed to analyze payment anomalies and to detect fraudulent payments



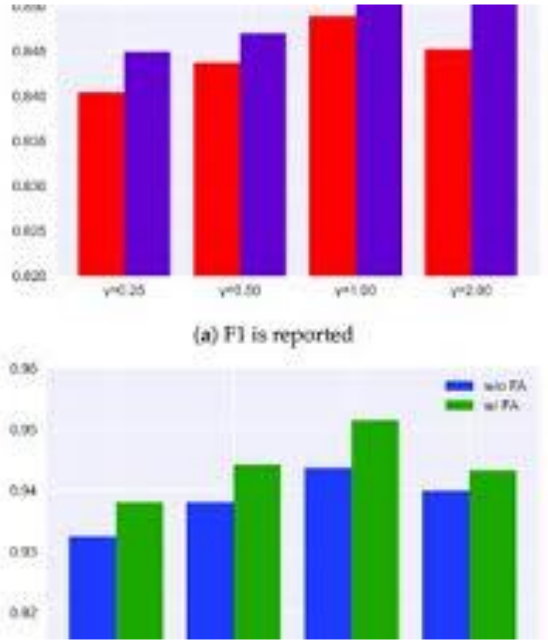
Machine learning models can be used to detect fraudulent transactions while minimizing false positives. Some benefits of using machine learning for fraud detection include: enhanced accuracy, scalability, and real-time monitoring.



The experimental results of IEEE CIS fraud dataset show that the method of this model is better than the benchmark model, such as logistic regression, support vector machine. Finally, the accuracy of our model reached 96.8%, and the AUC ROC score was 92.5%.

BERT outperforms many other architectures in a variety of token-level and sentence-level NLP tasks: Token-level task examples. Tokens refer to labels that are assigned to specific and semantically meaningful groups of characters, like words.

PaySim uses aggregated data from the private dataset to generate a synthetic dataset that resembles the normal operation of transactions and injects malicious behaviour to later evaluate the performance of fraud detection methods.



(a) F1 is reported

CONCLUSION :

Financial fraud is a growing problem that continues to challenge the financial industry despite advancements in technology. To tackle this issue, our study introduces a new model called RXT-J, designed to detect fraud in real-time financial transactions. This model is specifically built to handle the complexities of modern financial fraud, even when analyzing large amounts of data. It performs better than existing solutions, showing higher accuracy and the ability to uncover new and complex fraud patterns that were previously hard to detect .One major advantage of our model is that it overcomes the inefficiencies of traditional fraud detection methods. These older approaches often struggle to keep up with the fast-changing techniques used by fraudsters. To prove the effectiveness of RXT-J, we conducted detailed tests, comparing its performance with other machine learning and deep learning models using real financial transaction data. The results show that RXT-J consistently outperforms these other methods .Looking ahead, our research highlights areas for future improvement. For example, adding features like analyzing where and when fraud happens could make the model even more effective, especially as more data becomes available. Overall, our work represents a big step forward in the fight against financial fraud. By improving detection accuracy and efficiency, RXT-J helps make financial transactions safer and more secure .In a broader sense, this research also supports efforts in other fields, like wireless communication security, where advanced algorithms are used to protect data and systems. By enhancing the tools used to detect and prevent fraud, we contribute to building stronger defenses against these threats, ensuring a safer financial system for everyone.

REFERENCES :

1. 1.E. S. Prasad, The Future of Money: How the Digital Revolution is Transforming Currencies and Finance, Cambridge, MA, USA :Harvard Univ. Press, 2021.

2. 2.*Total Value of Investments Into Fintech Companies Worldwide From 2010 to 2022*, May 2023, [online] Available: https://www.statista.com/statistics/719385/investments-into-fintech-companies-globally/.

3. 3.*Fintech Market to Reach \$324 Billion in 2026*, May 2023, [online] Available: https://www.globaltrademag.com/fintech-market-to-reach-324-billion-in-2026/.

4. 4.*Digital Wallet Users to Exceed 4.4 Billion Globally by 2025*, May 2023, [online] Available: https://www.juniperresearch.com/press/digital-wallet-users-to-exceed-4-4-billion-by-2025/.

5. 5.R. Rasheed, S. H. Siddiqui, I. Mahmood and S. N. Khan, "Financial inclusion for SMEs: Role of digital micro-financial services", *Rev. Econ. Develop. Stud.*, vol. 5, no. 3, pp. 429-439, Jul. 2019.

6. 6.M. A. Hassan and Z. Shukur, "Review of digital wallet requirements", *Proc. Int. Conf. Cybersecurity. (ICoCSec)*, pp. 43-48, Sep. 2019.

7. 7.Y. Kou, C.-T. Lul Y.-P. Huang, "Survey of fraud detection techniques", *Proc. IEEE Int. Conf. Network . Sens. Control*, vol. 2, pp. 749-754, Mar. 2004.

8. 8.H. Shen, "Graph computing for financial crime and fraud detection: Trends challenges and outlook", *Int. J. Semantic Compute.*, vol. 14, no. 4, pp. 565-589, Dec. 2020.