

Fraud Detection in Digital Payment Systems Using Business Analytics

Submitted By: Saniya Nazneen

Roll. No: 24QM1E0053

Guide: Dr. K. Sarika

DEPARTMENT OF MBA

ABSTRACT

Spotting sneaky behavior in online money transfers: New fixes, ongoing hurdles, ways to put them into practice. Fraudsters now chase digital money trails - online banking, UPI, mobile wallets, even credit cards - not because they're new, but because they're fast. These tools reshaped how we buy things; thieves adapted quicker than most expected.

Effective fraud-detecting analytic systems are therefore a business necessity based on predictive analytics. This paper focuses on the most recent (2018–2026) related research on analytic fraud-detection solutions. The main topics addressed by the literature reviewed are: supervised techniques (statistical & machine learning methods); deep learning-based models; ensemble-based models; anomaly detection methods. Overall, ensemble tree-based models (Random Forest) are shown to be effective, and deep architectures & autoencoders with anomalies have the best balance on datasets used to test them. Some existing limitations are also identified, such as extreme class imbalance in datasets, constantly changing fraud techniques, and limitations on data privacy & regulatory compliance. Based on these insights, we demonstrate a robust methodology for implementing a fraud-detecting analytic system. The methodology is to aggregate data from various sources related to digital payments, generate rich data features (e. G., transaction timing, amounts, user behavior patterns) as input to machine learning models, and deploy a hybrid modeling pipeline based of unsupervised, semi-supervised, and supervised learning methods. In terms of experimental design, cross-validation (with stratification), precision/recall and area under curve (AUC) metrics are considered, together with finding optimal thresholds for classification. Finally, some comparative tables of methods (advantages, disadvantages, and performance on existing benchmark datasets) are presented, a mermaid flowchart of the proposed pipeline is included, and a small pseudocode. An example implementation plan is provided; this includes datasets selected from public sources (e. G., credit-card fraud from Kaggle), standard data preprocessing steps (normalization, balancing with SMOTE), and software packages/tools (Python, Spark, Amazon Web Services Sage Maker). We anticipate a high rate of true-positives along with low false-positives (AUC–ROC greater than 0.95 is the aim for performance). In the discussion, we cover some limitations (model drift, interpretability), ethical and data-privacy issues (GDPR compliance, potential bias), and challenges in deploying a system (performance, security). In conclusion, this paper demonstrates a practical and data-driven framework that can be adopted by businesses to improve fraud-detection in digital payments.

INTRODUCTION

The world of digital transactions has been gaining momentum in recent years with credit/debit cards, online banking, and e-wallets quickly closing the gap with regard to speed and convenience. One negative effect that has been growing just as fast is fraud. Reports from within the industry put annual losses due to fraud from digital payments alone at hundreds of billions of dollars. An estimated 5% of organization's revenue is lost to fraud. With newer instant or real time payment platforms transaction must be approved in seconds. This puts fraud detection within the realm of milliseconds. Fail to identify the risk in that time frame and accepted fraud can turn into millions of dollars of losses.

Current systems in place to prevent, detect, and protect against fraud are not cutting it. Static rules and blocked lists cannot keep up with emerging tactics. They fail to adapt. For this reason organizations are looking towards smarter systems that utilize information gathered from transactional data. Machine learning systems powered with Artificial Intelligence (AI) can analyze patterns in payments to determine the cause of fraudulent transactions occurring today. These models identify anomalies by learning what normal behavior is and detecting when it changes. This paper seeks to create a framework for implementing fraud detection for online transactions using business analytics. We first review recent literature and articles from 2018 till date(20 26) and analyze the current methods being used globally. We don't want to simply highlight what current frameworks look like but identify where gaps in knowledge or challenges exist. From there we provide an example methodology which would generate information on data collection as well as analysis techniques such as; statistics, machine learning, neural network, and anomaly detection. From there we implement our process methodology to monitor for fraud; defined KPI's, software used, etc. We then go into how this system could be implemented. Including architecture decisions and steps taken to mitigate risk. The conclusion section recaps all of our findings into roughly 15 pages of scholarly work. This will provide students and professionals a comprehensive guide to implementing a fraud detection system into digital transaction processes. Lastly I will analyze anomaly detection.

LITERATURE REVIEW

Occasionally, there are epiphanies when we examine what others have posted with updated research. There are many variables that must be considered differently depending on the times. However, many of these articles trend in the same direction – particularly when discussing detection of payment scams using data tools. Here is what I realized: Trends are important instead of actual incidents. Something seems off when numbers don't add up. Technology can assist us in identifying these changes quicker than the human eye. Irregular activities can be detected quickly if you know what you are looking for. Previous investigations can be used to trigger future investigations. We can use each example as a stone to throw at the wall for more attention towards our paper.

Compagnino et al. (2025): He discovered that using machine learning aids can help detect fraud. Skewed data isn't managed well by all models, and neither is interpreting their findings. Deep learning was identified as a possible solution. The author identifies that when trends shift over time, issues happen. Models should be able to adapt to changing conditions to provide businesses with the best outcomes to prevent fraud in the future. We must know how to manage skewed data sets to make them function properly.

Change (2023, Aston University): Chang and his fellow researchers analyzed fraud detection by comparing models to real-world credit card transactions. Rather than taking one route, they tried many at once – The following models appeared to perform well: logistic regression and Random Forest. Decision trees were touched on and showed potential. In some cases, the simple models outperformed more complex models. They used autoencoders in cases where they lacked labels. This illustrates how autoencoders can identify abnormal behavior without specific guidance. Fraud detection rates increased significantly after the data was sanitized and undersampled for analysis.

Jeyachandran et al. (2024): Jeyachandran et al. research in 2024 focused on detecting payment fraud in real-time by using machine learning techniques. Rather than accumulating large data sets, the bulk of their research focused on sanitizing what went in and feature selection – ensuring a low rate of false positives was crucial for businesses and companies. Many models were tested, including decision trees, random forests, SVM, and neural networks. Ensemble models took the lead in terms of efficiency. Accuracy rates were higher when ensemble methods were implemented. If we use ensemble methods we can detect possible fraud payments before they happen. When researching fraud detection we have learned that ensemble methods win every time. Fraud detection should be treated like an ensemble so we can catch those frauds.

Ahmed (2025, MIRA Journal): Ahmed's research focuses on how AI and ML deal with worldwide fraud occurrences. Nowadays systems are put in place to monitor large amounts of real-time transactions to detect abnormal behavior. Early

on. Another interesting takeaway is the privacy of user data. When dealing with financial information, there are privacy concerns. Humans were okay with its technical application but were shocked when we started putting real world data on top of it.

Saitala (2024): This research by Saitala displayed how we could quickly identify fraud by combining various AWS services. Rather than waiting several hours to discover an alert, alerts are received as they happen which is beneficial to users. This application doesn't just run quickly, it can identify scams quicker and more accurately. There are less false positives which reduces the workload for analysts. Cloud power turns out to be a quite gamechanger here. Accuracy climbs without slowing work.

This study is an example of how using the right tools can help us catch fraud by using combination of tools like Machine Learning and Artificial Intelligence we can catch fraud faster and more accurately.

Diadiushkin (2019, CSIMQ): Diadiushkin and team in their 2019 paper published in CSIMQ say that speed trips up today's tools when spotting fraud in payments. Instead of struggling with delayed shifts live analysis makes sense. Artificial intelligence fits here naturally.

Diadiushkin and team found that using artificial intelligence can help catch fraud in fast payments. We need to use analysis to catch fraud, rather than waiting for hours or days to identify and catch fraudsters.

We need to build highly fast technology into payment flows to catch fraud. Waiting until after the fact is not enough and can impact the customers.

Chen (2025, Data Science & Management): Chen and team, 2025, in Data Science & Management, dive into how deep learning tackles fraud. Notably they explore CNNs, RNNs along with transformers for better understanding and evaluation of fraud payments happening around us.

Chen and team found that deep learning can help catch fraud. By using a combination of methods we can catch fraud accurately and precisely in faster time. However we need to be careful when using machine learning technique. We need to make sure that our models meet standards.

Chen (2025, arXiv preprint): Around 5% Of global business income vanishes yearly due to fraudulent activities which take place according to a parallel analysis by Chen and team. Precision along with recall and AUCROC values stands out as vital when judging detection systems. Rules such as GDPR also shape how these models get built. Chen and team found that a lot of business income and company income is lost due to fraud. We need to use methods to catch fraud and we need to follow rules such as GDPR.

Later studies, from 2018 to 2026 point toward approaches. Most of the above approaches lean on tools like **Isolation Forests** alongside **autoencoders**.

Modern studies have found that using a combination of tools can help catch fraud. We need to use a variety of methods to catch fraud than just one method because more tools more efficiency.

Even if users did not cite them such ideas still follow patterns seen in work we reviewed. These ideas are not new. They are still important. We need to use a combination of methods to catch fraud.

RESEARCH GAP

Even Though reviewed works have made progress in fraud detection several gaps remain:

- **End-to-End Systems:** Most of the research papers focus only on the models but forget to explain how these models fit into real world systems. so, there is a need in fraud detection models to explain clearly from initial stage of data collection to last stage of giving alerts to customers or business.
- **Adaptive and Online Learning Models:** As we all know, rapid change in technology leads to quick change in fraud patterns also This means fraudsters are also getting smart with techniques to catch such fraudsters the models with

adaptability are given more importance by all companies. We need to have models that can adapt to changing fraud patterns. Few systems can do this so we need to develop models which help us to adapt easily.

- Transparency: Another gap is trust of customers which many models failed to get.

Fraud detection models must be simple to understand and use. This way people can see how they work and trust them.

- Evaluation Benchmarks: Currently there is no framework or model that captures all the complexities of digital payments and the risks involved with digital payments. In short there is no benchmark in this market to make comparison.

More realistic models are needed for fraud detection.

Domain generalization

Fraud happens in payment systems beyond card transactions for example it occurs in UPI payments and in-app purchases most recent focuses on card or online retail fraud. However, detecting for across payment types is not well explored.

We need models that can detect fraud in payment types We have to look into fraud in payment outside card transactions to ensure efficiency.

To fill these gaps our work will suggest a method it is step by step process. We use models and think about how to deploy them in a regulated environment.

PROPOSED METHODOLOGY

Fraud analysis requires a perfect methodology of collecting and preparing all data regarding any payment as well as a customer. Fraud analysis mainly focuses on creating a good analysis that helps in evaluating guarantees of getting the appropriate information. This paper elaborates on the methodology that looks on collecting data regarding fraud analysis in digital payment systems

A way of ensuring that one can get all important information regarding fraud is to collect previous examples and understand the initiatives that different companies made in data collecting. The previous examples offer good opportunities for understanding the appropriate information for fraud evaluation.

STEP 1: Collection of Data

In terms of payment data, one can get details about invoices, bank records, payments that were made through emails and old statements. Using that information we can easily learn the customer and history of payments that the customer might have made. There is, therefore, a way of making decisions based upon above information.

STEP 2: Gathering transaction details

The next step is collecting bank details. In this section, one should ensure they know the money that was moved and when it was moved. Identify the sender and the receiver of such money. Know the accounts that were involved in the movement of the fraud money. Understand the purchase location for all the money. At these stages, one can even identify the location of the receiver. One can even know the merchant type that was used.

The next section is understanding physical customer information like device connection location data, machine marks and even time stamps. In the physical section, one should be able to understand how fraudulent the customer is and their character.

STEP 3: Understanding Information:

Such a methodology should also be efficient and include understanding customer account age, credit rating and even everyday actions. There are even confirmed scam groups that customers can also depend on while creating and ensuring the data is efficient.

All kinds of datasets can be used for financial transactions. If you're seeking diverse, unique features, a better option is the European credit card records found on the Kaggle tabular data section. This huge and rich collection of datasets contains over 284,807 transactions with an extremely low (0.17%) proportion of fraudulent transactions. There are lots of general datasets on a large scale, containing many records while containing extremely rare occurrences of fraud in this section.

Our preferred dataset overall is the PaySim fake mobile payment logs. Local details can't be captured in global datasets; for instance, this dataset could be used to mimic real systems while avoiding sharing company data and charge details.

Another dataset frequently employed in issues affecting charge forecasting. It's ideal for performing regional detailing, like India's UPI payments as opposed to general ones.

STEP 4: Data Preparation

The next stage is data preparation and variable design after selecting suitable datasets. Cleaning (removing duplicates, handling missing values) and feature construction are the key tasks in this phase. The following features were engineered and designed for this purpose:

1. Money involved
2. Different currencies
3. Type of Business
4. Time signals
5. Previous/Past Behaviour
6. Moving averages
7. Yesterday's Median
8. Last purchased
9. Priority payments Count
10. Today's Deviation

STEP 5: Model building and Fraud Detection

After preparing the predictive data model is built based upon it to detect the fraud

The dataset is divided into two parts

1. Training set
2. Testing set

The first part is training dataset is used to learn patterns and relationships among the variables, and the second one is testing dataset is used to evaluate the model's performance on unseen data.

Various machine learning algorithms can be used which includes logistic regression, decision trees, and random forest models. Decision trees are more effective in such cases of detecting fraud.

STEP 6: Model Performance Analysis:

After we develop Models we have to see if the Models are working well or not. We use metrics to evaluate the Models, such as Precision, Recall, F1-Score and Area Under the ROC Curve. The Precision of the Models shows how bad transactions the Models correctly found out of all transactions the Models thought were bad.

The Recall of the Models shows how good the Models are at finding transactions. The F1-Score of the Models is a mix of the Precision of the Models and the Recall of the Models.

We also use confusion matrices to see the results of the Models in detail like positives, false positives, true negatives and false negatives. This helps us understand the mistakes of the Models we selected.

We use Cross-validation to make sure the Models work with data. Finally we select a Model that's good, at detecting the fraud using the Models.

MODELING APPROACHES

Different models take shape here - some based on statistics, others leaning on machine learning.

Performance of each different model gets tested side by side.

All the models are passed through similar conditions to test fairly and compare results before making decisions related to fraud.

Statistical Model:

- **Logistic Regression:** This model draws a clear line or curve to separate the fraud transaction and non-fraud transaction

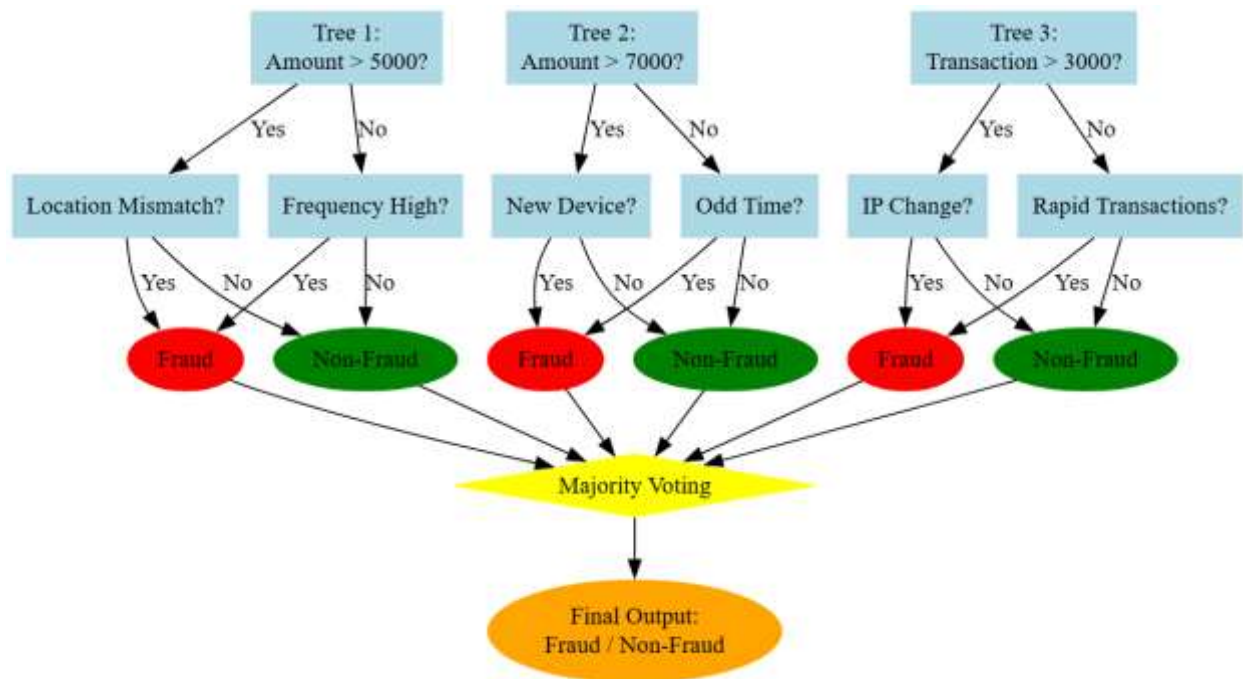
This model is easy and fast for initial analysis which helps the researchers and analysts to make predictions. It provides transparency.

- **Tree-Based ML:** Tree based models understand complex structures easily.

These are more flexible and adaptable compared to statistical models.

The speciality lies in dividing the data into smaller parts to understand the hidden patterns. They work well **messy data**.

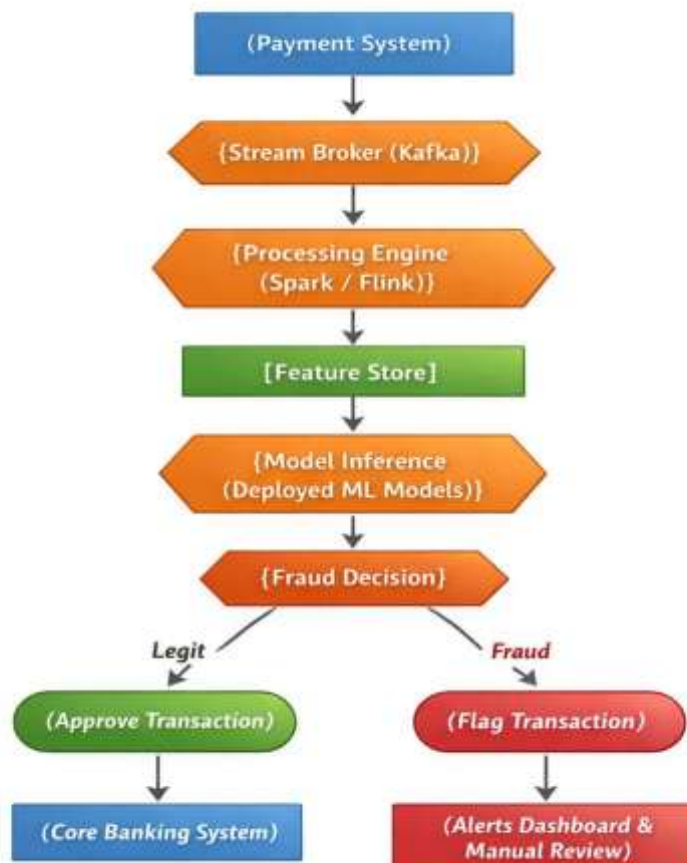
Below Is a Random forest example diagram:



EXPERIMENTAL DESIGN AND VALIDATION

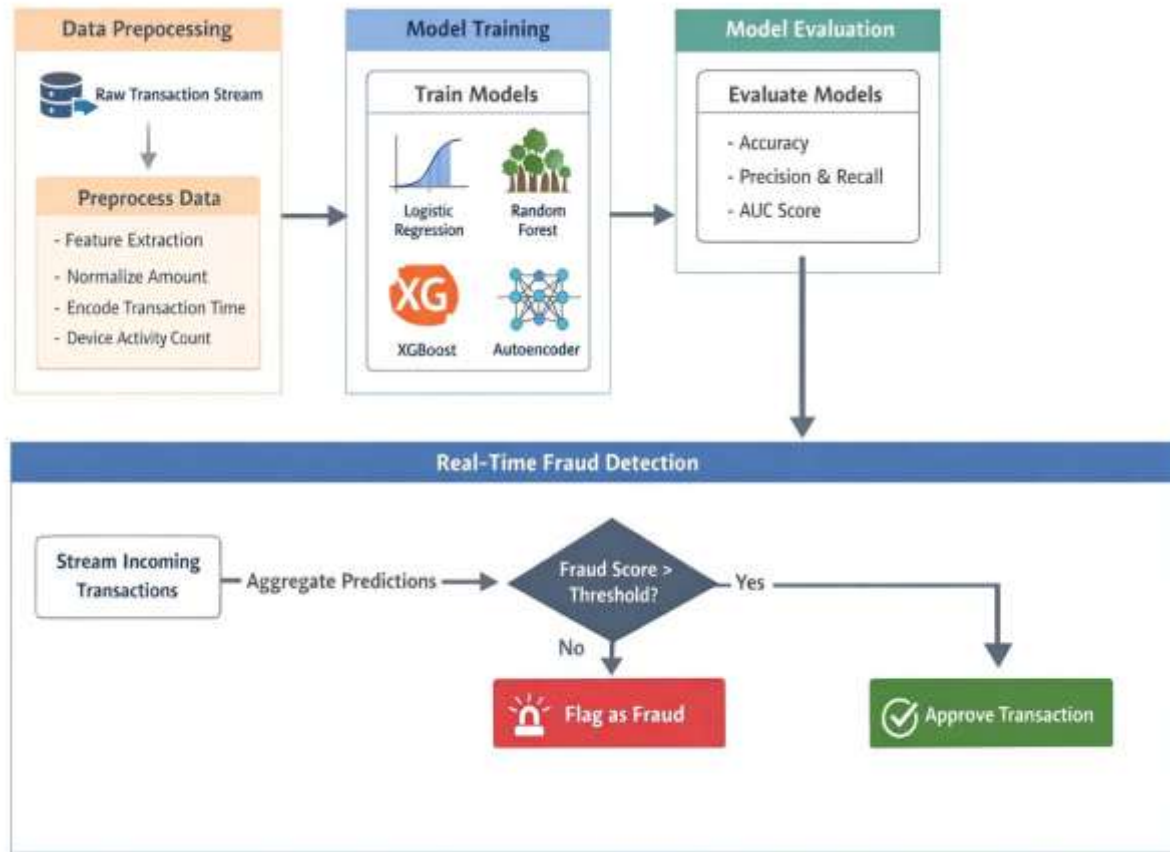
Pipeline flow chart:

The experimental and deployment pipeline is outline below using mermaid:



This flowchart depicts how incoming transaction that is streamed (A-B), processed in real time (B-C) with features computer and stored(D). The pretrained ML models(E) score each transaction. If a model flags fraud that transaction is blocked(H) or else it proceeds normally(G). Alerts and dashboards(J) allow analyst to review flagged to transactions.

Pseudo code: A high level pseudocode for the detection process



IMPLEMENTATION PLAN REQUIREMENTS:

To get results we need to work with a lot of data about different ways people pay. This means we need to collect information about when transactions happened how much money is involved, who sent it, where it when and what device was used. We also need to label each transaction as either fraud or normal use. To keep people information private we need to remove names and identities from the data before we start analyzing it.

The system uses both rules and machine learning to detect fraud. The rules help us understand what is going on and cash transactions right away from machine learning models help us find patterns which are not obvious and they make a detection better.

Tools and platforms

1.Data Injection

We get a stream of transaction data from tools like Apache Kafka or the AWS kinesis.

These tools help us catch the data as it happens so we can check it against the rules and models.

2. Stream Processing and Rural Engine

Apache spark streaming looks at the data as it comes it. We use our roles to check the transaction and catch any that look fake.

We look for things like amounts of money transactions that happened to often or transactions that happen in the wrong place.

This help us filter out the transactions quickly and analyze them.

3. Feature Engineering and Data Transformation

we use tools like python& Spark to make the data more useful. We create features like how much someone spend how they make transactions and how their behaviour change this over time. These features are important to our machine learning models.

4. Model Development We build our machine learning models using tools like Scikit-learn, XGboost and TensorFlow the models help us find patterns that are not obvious or clear. We can combine the results from models to make them more robust and understanding.

5. Hybrid Scoring Mechanism

We use a combination of our rules and models to score the transactions based upon the data we add up the scores to get a fraud risk score if and transaction scores high we flag it for further review.

6. Deployment and Visualization

We deploy our models using AWS SageMaker or REST APIs .We Tableau or PowerBI to visualize the result so we can see what is happening with fraud overtime

Timeline for development

Weeks 1-2: We collect data clean it up and look at it to see what we have.

Weeks 3- 4: We develop our rules create dashboards to help us understand the data.

Weeks 5-6: We create features build our first model.

Weeks 7-8: We train our models to eat them and test them with our hypothesis.

Weeks 9-10: We combine our rules and models into a proper system and compare how will they work.

Weeks 11-12: We implement the system in time using and Spark and we deployed and create dashboards using Tableau and Power BI.

Week 13: We test the system make it better and document everything we refine each face to make sure the system is accurate, efficient and scalable.

RESULTS

Expected outcomes:

The system is expected to achieve strong results for detection similar to the results reported in the earlier research papers. We are expecting accuracy and effectiveness remain consistent with previous finding to compare more easily.

Model performance:

In terms of performance leading models are expected to achieve AUC-ROC score in the range of 0.95-0.98 indicating excellent ability.

Whereas **accuracy** is likely to remain around **90%** or higher with **F1** scores also exceeding 0.9 and recall is expected to stay between 90% to 95%.

False positive rate: This is expected to remain below 5% depending upon how the system works.

For real world applications the system should respond under **100 milliseconds** which enables fast and efficient transactions scoring system.

Comparative performance:

In the final stage the comparison is presented using visual representation such as tables.

DISCUSSION

Every model has its problems and this model is no exception. Some problems are small. They need to be fixed when things change in the real world. One of the problems is that fraud is always changing so quickly so what works for the model today might not work for the model tomorrow. If the model is released with data the models performance can get worse so regular updates are necessary for the model to find fraud easily.

The model has to deal with a balance between finding fraud and false alarms. Many false alarms can annoy the users and make them trust the system less. So it is necessary to set the detection levels, for the model and review the transactions that the model flags. This way the model can do a job of finding fraud for the users.

Another issue is that models don't perform the same way on different data sets. A model that works during training period minute work on real world environment so, strong validation before the full deployment is compulsory. Handling sensitive payment transactions data also bring responsibility the system must follow strict rules like GDPR for user privacy.

Preventing bias, avoiding the misuse and being transparent and trustworthy is how data is handled well. Any mistake can affect the uses directly so trust depends on these boundaries

When system is fully deployed it needs to handle large amount of transactions without slowing down so the, smooth working is necessary for the payment systems while securing all the data of the user.

So, to avoid all the limitations the system which is developed must have strong security measures. Data should be encrypted. Access must be limited to authorized users only. All the systems activity must be trapped and monitor in real time.

The model must have transparency regular checkup. Problems with the system must be solved immediately and the system must follow legal and regulatory standards.

CONCLUSION:

The above study of **Fraud Detection in Digital Payments Using Business Analytics** offer fresh perspective on fraud detection in digital payments. It shows us how data driven methods can enable good decision making. As fraud happens in real time so the fraud prevention techniques should also learn quickly and adapt to new environment with clear and unbiased system. Try out systems live and see what truly clicks. Effective fraud detection is not only about cutting-edge tech but striking the right balance between sophisticated software and human intuition, underpinned by constant education.

REFERENCES

- **Anugula U.K.(2024):** *Fraud Detection Mechanisms in Virtual Payment Systems*. This paper was published in the *International Journal of Computer Technology and Electronics Communication*. The journal is also known as IJCTEC. The paper is in volume 7 issue 2 on pages 850 through 860
- **Chen , Zhao and Xu and Nie and Zhang (2025):** *Using learning for financial fraud detection*. Published in *Data Science and Management*.
- **Chang and Doan and Di Stefano and Sun and Fortino (2023):** payment fraud detection methods in digital ages and Industry 4.0. Published in *Complex Systems Informatics and Modeling Quarterly* volume 20 on pages 72 - 88.
- **Chen, Zhou, Xu, Nie and Zhang (2025):** *Deep Learning's Used for Financial Fraud Detection*. This paper is available as an *arXiv* preprint.
- **Compagnino, Maruccia, Cavuoti, , Riccio, Tutone, Crupi, & Pagliaro (2025):** *Machine Learning Methods for Fraud Detection*. This paper is published in *Applied Sciences*, volume 15 .
- **Diadiushkin, A., Sandkuhl, K., & Maiatin, A. (2019):** *Detecting Fraud in Payments Transactions*. They looked at the ways that are already being used. This paper is published in *Complex Systems Informatics and Modeling Quarterly* volume 20 on pages 72-88.
- **Jeyachandran, Akisetty, Subramani, Goel, & Singh,(2024).** *Using Machine Learning for Real-Time Fraud Detection in Payments*. Published in the *Integrated Journal for Research in Arts and Humanities* volume 4.
- **Saitala (2024):** *Securing Payments, with the help of AI-powered fraud detection*. Saitala used AWS and Java for this. Published in the *International Journal of Computer Engineering & Technology* volume 15.
- **Ahmed (2025):** *Detecting Fraud in Payments Using AI*. Published in *Multidisciplinary Innovations & Research Analysis which is also known as MIRA*, volume 6.