International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0125
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 819 - 822

Financial Fraud Detection in Banking Transactions

Harshath Vijay V B[1]
[1]*UG – Information Technology, SIST University, Navalur, Tamil Nadu*
Emails: *n.vanchi@gmail.com*[1], *vbharshathvijay@gmail.com*[2]

Abstract
The study focuses on detecting anomalies and fraudulent activities in banking transactions using a variety of machine learning and deep learning models. It utilizes a dataset, often referred to as the Financial Anomaly Detection dataset, which is typically sourced from a reliable dataset repository in either '.csv' or '.xlsx' format. This dataset contains transaction records that include features such as transaction amounts, timestamps, customer details, and other transactional metadata, making it ideal for identifying patterns that deviate from typical behavior. Detecting fraud and anomalies is crucial for enhancing the security of financial systems, and the models implemented in this study aim to address this challenge effectively. To achieve anomaly detection, the study employs a range of machine learning algorithms. One of the primary models used is the Random Forest algorithm, which is particularly effective in handling large datasets and complex feature interactions. Random Forest is trained to classify transactions as either legitimate or potentially fraudulent, based on historical patterns of known fraudulent transactions. Another key model utilized is the Isolation Forest, which is designed to identify anomalies by isolating outliers within the dataset. This model is well-suited for detecting rare, unusual, or malicious activity that deviates from the majority of transactions. In addition to traditional machine learning models, the study also explores the use of deep learning techniques, specifically Long Short-Term Memory (LSTM) networks. LSTMs are a type of recurrent neural network (RNN) that excels in capturing sequential patterns in time-series data, such as transactions over time. By analyzing sequences of transaction records, LSTMs are able to identify subtle patterns and predict anomalous or fraudulent transactions with a high level of accuracy. These deep learning models are particularly advantageous when dealing with large amounts of time-dependent transactional data, as they can retain long-term dependencies and capture complex relationships that simpler models might miss.

Keywords: *Keywords- Fraud Detection, Banking Transactions, Financial*

1. Introduction
Anomaly detection in banking transactions is a critical task in the modern financial ecosystem. With the rapid advancement of digital banking, mobile payments, and online financial services, the volume of transactions has significantly increased. This, in turn, has raised the importance of accurately detecting fraudulent activitiesand outlier transactions to ensure the safety and security of banking systems. Anomaly detection techniques can be defined as methods that aim to identify unusual or abnormal behavior in data that significantly deviates from the norm.In the context of banking,this involves identifying fraudulent transactions, systemmal functions, operational errors, and other anoles that could potentially lead to financial loss, fraud, or breaches of customer privacy. Fraudulent activities in banking can have disastrous consequences, not only for the affected customers but also for the financial institution's reputation and the broader financial market. As a result, banks and financial institutions have increasingly turned to data analytics, machine learning, and deep learning models to enhance their fraud detection capabilities. These models allow for the automatic detection of fraudulent activities in real time or near-real time, preventing further damage and loss. Anomaly detection techniques are integral to these systems, as they help in recognizing patterns that fall outside the usual range of behavior, which might indicate a fraudulent transaction, a cyberattack, or a malfunctioning system. The rapid growth of digital financial services and the corresponding rise in cybercrimes have made it increasingly difficult for traditional methods of fraud detection to keep up. Historically, fraud detection systems in banks relied

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0125
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 819 - 822

on rule-based methods or statistical techniques, where predefined rules or thresholds were used to flag unusual transactions. While these systems were effective in detecting some types of fraud, they were limited in their ability to handle complex patterns and identify novel forms of fraudulent activity. For instance, traditional systems could only recognize fraud patterns that had been previously recorded, making them vulnerable to evolving types of fraudulent activity, which could bypass the predefined rules. In recent years, machine learning (ML) and deep learning (DL) techniques have revolutionized the field of anomaly detection in banking transactions. Machine learning is a subset of artificial intelligence that allows systems to learn from data and improve over time without explicit programming. In anomaly detection, ML algorithms can be trained on historical transaction data to learn the normal behaviour of customers, and then flag transactions that deviate from this learned pattern. With the use of advanced algorithms, such as Random Forest, Isolation Forest, and Support Vector Machines (SVM), machine learning models can detect not only known types of fraud but also previously unseen or emerging fraudulent schemes. The increasing complexity and volume of financial transactions require more advanced tools to handle the diverse nature of banking data. This is where deep learning techniques, specifically Long Short-Term Memory (LSTM) networks, come into play. LSTMs are a type of recurrent neural network (RNN) that excels at processing sequential data, making them particularly suited for time- series data, such as transactional records in banking. LSTMs can recognize sequential patterns in transaction histories, which allows them to detect anomalies that unfold over time, such as sudden changes in spending patterns, unusual account activity, or potential identity theft. By capturing these temporal dependencies, LSTMs can provide highly accurate predictions and detect previously unseen fraudulent activities that might be missed by traditional machine learning models.

2. Methods

2.1.Random Forest

In the context of fraud detection[1], Random Forest is particularly effective at identifying fraudulent transactions in banking systems. Since fraud detection involves distinguishing between legitimate and fraudulent transactions, which can involve complex patterns and subtle anomalies[2], Random Forest's ability to handle large datasets with many features and its robustness to noise makes it a valuable tool for this task. When applied to banking transactions, Random Forest can analyse various features such as transaction amount, time, location, and account history to classify transactions as either fraudulent or legitimate[3]. It can also handle imbalanced datasets, where fraudulent transactions make up a small fraction of the total transactions, by ensuring that the model is not biased towards the majority class (legitimate transactions)[4]. The model learns to detect the underlying patterns associated with fraud by training on labelled data containing both fraudulent and non- fraudulent transactions. Random Forest's ability to generalize well makes it suitable for detecting new, previously unseen fraudulent activities. Fraud patterns often evolve over time, and Random Forest’s ensemble nature allows it to adapt to these changes by aggregating the results from different trees that might capture different aspects of fraud[5][6].

2.2.Isolation Forest

Isolation Forest is a machine learning algorithm specifically designed for anomaly detection, with a primary focus on identifying outliers or anomalous data 1) Methods[7][8] points in large datasets. Unlike traditional anomaly detection methods, which typically model the distribution of data and try to classify instances based on their deviation from the norm, Isolation Forest takes a different approach. It is based on the concept of isolating anomalies rather than profiling normal data, making it particularly well-suited for high- dimensional datasets where traditional methods might struggle[9]. The core idea behind Isolation Forest is that anomalies are "few and different," meaning that they are easier to isolate compared to normal data points, which tend to be more frequent and clustered together[10][11]. The algorithm isolates anomalous points by randomly selecting a feature and then randomly selecting a split value between the minimum and maximum values of

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0125
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 819 - 822

that feature[12][13]. This process of randomly partitioning the data continues recursively until the point is isolated or until a specified number of partitions (or "splits") is reached.Shows Figure 1 Flow Diagram[14][15]

2.3.Figures

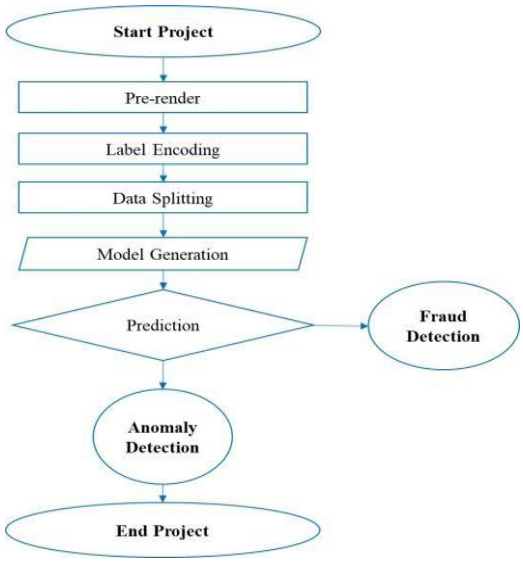

Figure 1 Flow Diagram

3. Results

80% of the data in the process is used for training and 20% of the data in the process is used for testing. A new user account is created successfully and the user is redirected to the login page. The user is able to log in successfully and access the fraud detection features.Shows Figure 2 Flow Of The Dataset.

4. Discussion

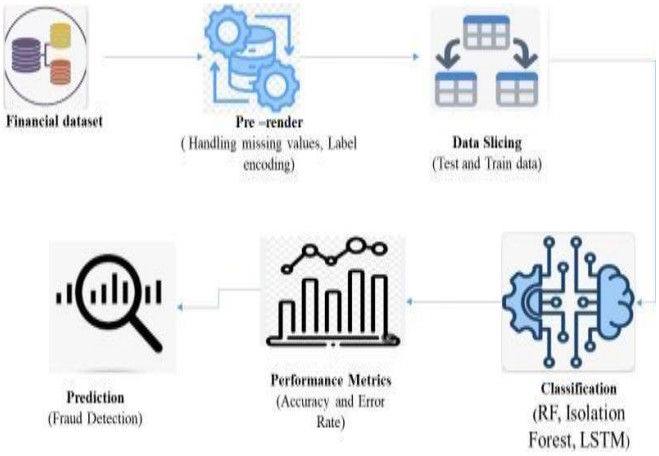

Figure 2 Flow Of The Dataset

The above architecture dataset shows the process by which the result is achieved. The architecture diagram outlies the workflow for predicting financial data using machine learning models. It begins with data selection where the AI ability dataset is sourced in CSV format. The data then undergoes a pre-render phase to address missing values and apply label encoding, transforming categorical variables into numerical form. Following data preparation, data splitting separates the dataset into training and testing subsets. The classification phase employs ML algorithms to build predictive models. Result generation involves evaluating evaluating the models using performance metrics. Finally, the models are used to predict anomaly.

Conclusion

In conclusion, the use of advanced machine learning and deep learning models, such as Random Forest, Isolation Forest, and LSTM, plays a critical role in improving fraud detection and anomaly identification in financial transactions. These models allow for the processing of large datasets to uncover hidden patterns and trends, which are crucial for identifying fraudulent activities or outliers in real time. By leveraging preprocessing techniques, such as handling missing values and normalizing data, the accuracy of these models is enhanced, ensuring more reliable predictions. Random Forest's ability to handle complex classification tasks and Isolation Forest's strength in anomaly detection offer significant advantages in dynamic financial environments. Moreover, LSTM's capability to capture long-term dependencies and sequential patterns makes it particularly effective in fraud detection, where behaviours change over time. As financial institutions continue to face increasing security risks, implementing these models can provide a proactive approach to detecting and preventing fraudulent transactions. Ultimately, these advanced techniques not only increase the efficiency of fraud detection but also contribute to the overall security and stability of financial systems. Real-time prediction capabilities ensure that potential threats are identified and addressed swiftly, minimizing the impact on users and institutions alike.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0125
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 819 - 822

References

[1]. Hastie, T., Tibshirani, R., & Friedman,J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media.
[2]. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer Science & Business Media.
[3]. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
[4]. Goodfellow, I., Bengio, Y., & Courville,(2016). Deep Learning. MIT Press.
[5]. Chollet, F. (2017). Deep Learning with Python. Manning Publications.
[6]. Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. MIT Press.
[7]. Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction. MIT Press.
[8]. Kohavi, R., & Provost, F. (1998). "Glossary of terms." Machine Learning, 30(2-3), 271-274.
[9]. Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2007). "Supervised machine learning: A review of classification techniques." Emerging artificial intelligence applications in computer engineering, 160-166.
[10]. Jordan,M.I.,&Mitchell,T.M.(2015). "Machinelearning:Trends,perspectives,and prospects." Science, 349(6245), 255-260.
[11]. Caruana, R., & Niculescu-Mizil, A. (2006)."An empirical comparison of supervised learning algorithms." Proceedings of the 23rd international conference on Machine learning, 161- 168.
[12]. Langley, P. (1996). "Elements of machine learning." Morgan Kaufmann Publishers Inc.Mitchell, T. M. (1997).
[13]. Machine Learning. McGraw-Hill.Raudys, S. J., & Jain, A. K. (1991).
[14]. "Small sample size effects in statistical pattern recognition: Recommendations for practitioners." IEEE Transactions on pattern analysis and machine intelligence, 13(3), 252-264.
[15]. Friedman, J., Hastie, T., & Tibshirani,R. (2001). "The elements of statistical learning." Springer series in statistics.