International Journal of Research and Development in Engineering Science
ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/
Regular Issue, Vol. 7, Issue.1 (January – February), 2025 , Pages: 21 - 24
Received: 29 June 2024 ; *Accepted:* 12 November 2024 ; *Published:* 21 January 2025
https://doi.org/10.63328/IJRDES-V7RI1P4
Original Paper : Policy Research – Report Research Paper

A Predictive Analytics with Machine Learning Algorithms for Real Time Financial Data

B. Lakshmi Prasanna [1] , T. Kesava Rao [2] , M. E Palanivel [3]

Department of Computer Science and Engineering, Sreenivasa Institute of Technology and Management Studies, Chittoor, Andhra Pradesh-517325, India,

* *Corresponding Author:* lakshmiprasanna0026@gmail.com

Abstract: Fraud has come a trillion-bone assiduity moment. Certain fiscal institutions have devoted brigades of sphere experts and data scientists that are assigned with detecting fraudulent exertion. To find frauds, data scientists constantly employ intricate statistical models. But there are a lot of downsides to this strategy. Since fraud discovery isn't real- time, fraudulent exertion are constantly discovered only after the factual fraud has taken place. These methodologies are prone to mortal crimes. In addition, it requires precious, largely professed sphere expert brigades and data scientists. nonetheless, the delicacy of homemade fraud discovery methodologies is low and due to that, it's veritably delicate to handle large volumes of data. More frequently, it requires time- consuming examinations into the other deals related to the fraudulent exertion in order to identify fraudulent exertion patterns.

Keywords: Random Forest Classifier, Decision Tree Classifier, K-Means, Gaussian NB, SVC and K Neighbors Classifier.

1. Introduction

The Dynamic nature of financial transactions drives the use of a comparative predictive analytics method with machine learning for real-time fraud detection in financial data. Traditional methods frequently fail to adapt quickly to new fraud tendencies. This study compares several machine learning models in order to discover the most effective and adaptive solution. Improving fraud detection accuracy and speed is critical for protecting financial systems, creating trust in stakeholders, and limiting potential economic losses, resulting in a more secure and resilient financial ecosystem[1]. Detecting financial fraud in real time is crucial for ensuring economic stability. This project is to use a comparative method in predictive analytics, specifically machine learning (ML), to improve fraud detection in real-time financial data. The study aims to determine the best efficient strategies for detecting fraud in a timely and accurate manner by comparing and assessing several machine learning models. The study's findings will provide vital insights into the industry, allowing financial institutions to create sophisticated systems that proactively minimize fraudulent activity in real-time financial transactions[2].

ML models have long been employed in a variety of application domains where adverse threat variables required to be identified and prioritized. Several prediction algorithms are widely employed to address detection issues. This study highlights the capacity of ML models to detect transaction fraud. Predictive analytics using machine learning (ML) for fraud detection in real-time financial data has considerable potential. This study tries to discover the most effective model for detecting fraud in real time by analyzing multiple machine learning techniques[3]. The study will investigate and compare the performance, accuracy, and efficiency of various ML algorithms, resulting in significant insights for improving financial security measures. This technique advances fraud protection strategies in dynamic financial contexts while ensuring the durability of real-time fraud detection systems[4].

Global networking and the development of new technologies have led to a sharp increase in fraud. There are two main approaches to combating it: detection and prevention. Once prevention has failed, detection steps in to provide a protective barrier against fraudulent actions. Therefore, detection aids in spotting and warning about fraudulent transactions as soon as they happen[5]. Credit card operations are seeing an increase in card-not-present transactions, particularly through online payment gateways [6]. In October 2016, the Nilson Report revealed that worldwide revenue generated by online payment systems in 2015 exceeded. $31 trillion, indicating a 7.3% rise from 2014. At the same time, losses from credit card theft increased to $21 billion globally in 2015 and are

ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, Issue 1,2025 , https://doi.org/10.63328/IJRDES-V7RI1P4

expected to rise to $31 billion by 2020. The notable increase in fraudulent transactions has had a substantial effect on the economy.

Two types of data are investigated in this study: numerical and categorical. The dataset started out with categorical data. Raw data is refined by means of data cleansing and other preparatory processes.

First, numerical data is transformed from category data and appropriately applied analysis techniques are applied. Second, to find the best algorithm, machine learning techniques are used on categorical data[7].

2. Related Work

To improve an accuracy we are going propose an innovative application addressing limitations of traditional methods, aiming to develop a fast and reliable fraud detection system. This study employs powerful algorithms, including Random Forest Classifier and K- Neighbours Classifier, in a Python environment. By leveraging these advanced techniques, the objective is to enhance accuracy in identifying fraudulent transactions.

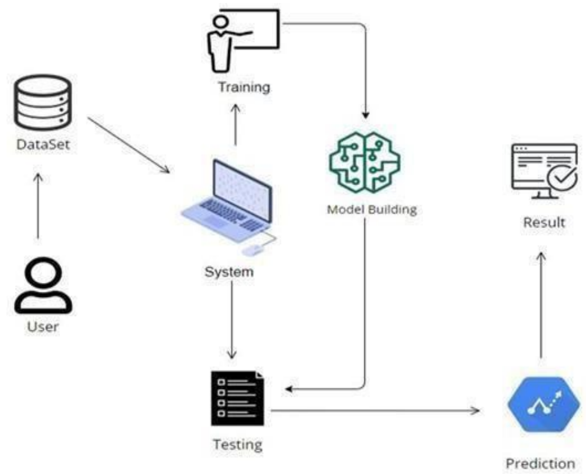

Figure.1 Architecture of Methodology

This comparative approach ensures a comprehensive evaluation, contributing to the creation of an efficient and robust system for real time financial data fraud detection, surpassing the constraints of existing methods. By comparing multiple machine learning models, the approach ensures higher accuracy in identifying fraudulent activities within real-time financial data.

Through comparison, the method allows for the selection of the most robust and effective machine learning model tailored to the unique characteristics of financial datasets, optimizing fraud detection capabilities. The comparative approach enables continuous evaluation and adaptation to evolving fraud patterns, ensuring the model remains effective in addressing new and sophisticated fraudulent techniques in real-time. By evaluating the performance of various models, organizations can allocate resources more efficiently, focusing on the models that demonstrate superior fraud detection capabilities, thereby optimizing operational efficiency[8].

Machine Learning Algorithms

(a). Random forest

One of the machine-learning technique used for regression and classification applications is a random forest. It makes use of ensemble learning, an approach that successfully combines several classifiers to solve complex problems. Many decision trees are used in the random forest technique. Random forest generates a "forest" that is trained by bootstrap aggregation or bagging.

An ensemble meta-algorithm called bagging improves machine learning models' accuracy. By combining the results of each of its own decision trees, the random forest algorithm generates predictions. By averaging or taking the mean of the predictions produced by various trees, it generates forecasts. A higher tree count improves prediction accuracy.

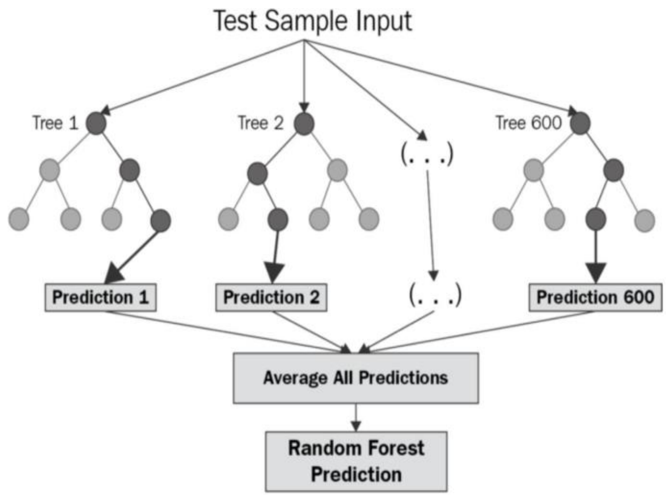

Figure.2 Random Forest workflow

Random forest overcomes the limitations of individual decision trees. It mitigates over fitting issues while improving precision. This method facilitates predictions without requiring extensive parameter adjustments, such as those in Scikit- learn[9].

(b). K-Nearest Neighbor (K-NN)

K-Nearest Neighbor is a straightforward machine learning algorithm that falls under supervised learning techniques. It categorizes new data points based on their similarity to existing data points.

ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, Issue 1,2025 , https://doi.org/10.63328/IJRDES-V7RI1P4

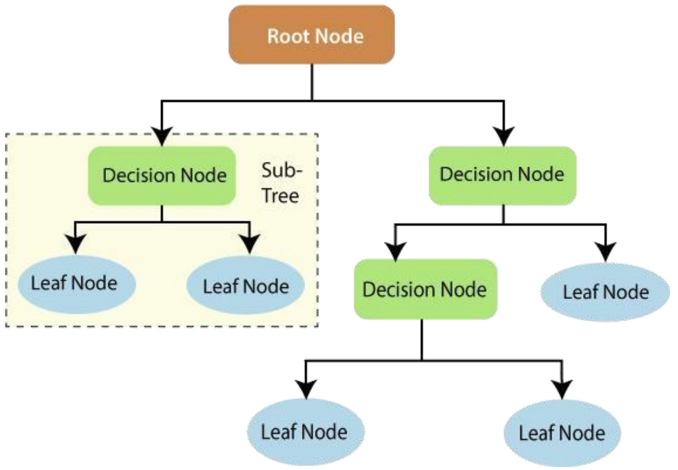

Figure.3 KNN Workflow

This method utilizes the entirety of available data to classify new instances, placing them into appropriate categories based on their resemblance to existing data points. While K-NN can be used for both regression and classification tasks, it is primarily used for classification[10].

Why do we need a K-NN Algorithm?

Imagine we have two categories: Category A and Category B, and we're given a new data point, x1. To determine which category x1 belongs to, we can use the K- NN (K-Nearest Neighbors) algorithm. This algorithm helps us to identify the category or class of a new data point based on the closest data points from the existing dataset. Refer to the diagram below to understand this concept better.

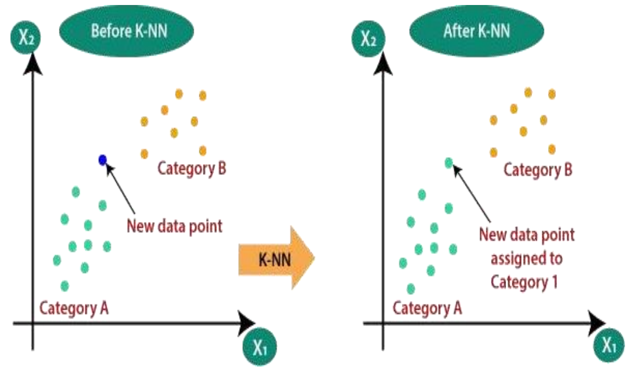

Figure.4 Data Point Category Wise

(c). Evaluation Metrics for Fraud Detection Models

For evaluating the performance of machine learning models used in fraud detection, several key metrics are commonly employed. These metrics help determine how effectively the model can distinguish between fraudulent and non-fraudulent transactions.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+EF+FN} \quad (1)$$

$$\text{Recall (R)} = \frac{TP}{TP+FN} \quad (2)$$

$$F1 = 2 \times \frac{P{\times}R}{P{\times}R} \quad (3)$$

(d). ***Feature Importance***

Random Forest provided insights into which features were most indicative of fraud, aligning with domain knowledge that transaction amount and frequency are crucial indicators.

(e). ***Parameter Sensitivity***

KNN's performance varied significantly with the choice of K and distance metric, while Random Forest required careful tuning of its hyper parameters to avoid over fitting.

(f). Data Description and Preprocessing:

Dataset: The study utilized a real-time financial data set containing X records and Y features. Pre-processing steps included handling missing values, scaling numerical features, and encoding categorical variables to ensure data suitability for machine learning models.

Table.1 Model Performance Metrics

S. No	Algorithm	Accuracy	Precision	F1-Score
1	Random Forest	99.087	0.98	1.0
2	**KNN**	**99.704**	**0.98**	**1.0**

Table.2 Comparison of Algorithms

Feature	Random Forest	K-Nearest Neighbors (KNN)
Feature Importance	High importance of features like transaction amount, transaction frequency, and location.	Not Applicable
Impact of Parameters	Performance significantly affected by tuning parameters like the number of trees and maximum depth.	Highly dependent on the choice of K and distance metric.

3. Conclusion And Future Scope

This paper presents a groundbreaking comparative examination of machine learning algorithms for real-time financial fraud detection, employing a broad set that includes Random Forest Classifier, Decision Tree Classifier, K Means, Gaussian NB, SVC, and K Neighbors Classifier. This study is unique in that it compares its performance ~~to~~ real-world financial data, assessing predicted accuracy in distinguishing

ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, Issue 1,2025 , https://doi.org/10.63328/IJRDES-V7RI1P4

between fraudulent and legitimate transactions. The use of K Means, which is generally disregarded in fraud detection, provides a unique perspective.

This paper advances the area by rigorously analyzing several models, resulting in a thorough grasp of algorithmic efficacy in real-time financial fraud prediction. The innovative insights gained from this multi modal approach make a substantial contribution to improving the robustness and accuracy of fraud detection systems.Future upgrades could include integrating sophisticated ensemble methods, investigating deep learning architectures, and utilizing anomaly detection techniques to provide a more comprehensive fraud detection framework.

Implementing real-time feature engineering and continuous model updating will improve flexibility to changing fraud tendencies. Furthermore, the use of Explainable approaches will increase transparency, which is essential for regulatory compliance and stakeholder trust. Continuous collaboration with industry experts and the incorporation of future algorithms will keep the system at the forefront of fraud detection skills in the ever-changing field of real-time financial data.

References

[1]. Greene, John, and Ian Bone," Understanding neurology: A problem-oriented approach", CRC Press, 2007.

[2]. Pouyanfar, Samira, et al,"A Survey on Deep Learning: Algorithms, Techniques, and Applications", ACM Computing Surveys (CSUR), vol.51, no.5, pp.92, 2018.

[3]. Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla,"Segnet`: A deep convolutional encoder-decoder architecture for image segmentation", IEEE transactions on pattern analysis and machine intelligence, vol.39, no.12, pp.2481-2495, 2017.

[4]. M Karunakar Reddy , M Bavana , M Bharathi , R Anjali , G Anusha , K Lakshmi Kaveri , "Mixed Strategy Game Theory based Clustering Routing Algorithm for Wireless Sensor Networks " ,International Journal of Computational Science and Engineering Research, vol. 1, no. 3, p. 23, July. 2024, doi: https://doi.org/10.63328/IJCSER-V1RI3P6

[5]. Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C,"mining for credit card fraud: A comparative study" by Decision Support Systems, vol.50,no.3,pp.602-613,2011.

[6]. Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G, "Calibrating probability with undersampling for unbalanced classification" by IEEE Symposium Series on Computational Intelligence pp.159-166,2015.

[7]. Chen, C., & Li, Y," Financial fraud detection in online transaction based on machine learning" by International Conference on Big Data Analytics for Cyber-Physical System in Smart City pp.225- 231,2020.

[8]. Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.-E., He-Guelton, L., & Caelen, O, " Sequence classification for credit-card fraud detection" by Expert Systems with Applications vol.100,pp.234-245.

[9]. Phua, C., Lee, V., Smith, K., & Gayler, R. "A comprehensive survey of data mining-based fraud detection research" by arXiv preprint arXiv vol.1009.6119,2010.

[10]. Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & SunX, "The application of data mining techniques in financial fraud detection" by Decision Support Systems, Vol.50,no.3,pp.559-569,2011.

[11]. Zhang, Y., & Zhou, X, " Detecting financial fraud using random forest and support vector machine" by International Journal of Network Security vol.22,no.5, pp.662-669,2020.

[12]. Roy, A., Roy, R., & Dutta, D, "An approach towards financial fraud detection using random forest" by International Journal of Pure and Applied Mathematics, vol.118,no.9, pp.365- 371,2018.

[13]. Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. M, "Transaction aggregation as a strategy for credit card fraud detection" by Data Mining and Knowledge Discovery, vol.18,no.1,pp.30-55,2009.

[14]. Carcillo, F., Dal Pozzolo, A., Le Borgne, Y. A., Caelen, O., Mazzer, Y., & Bontempi, G, "A scalable framework for streaming credit card fraud detection with Spark" by Information Fusion, vol.41, pp.182-194,20

[15]. Jayakarthik R, Srinivasan A, Goswami S, Shivaranjini, Mahaveerakannan R. Fall Detection Scheme based on Deep Learning Model for High-Quality Life. 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC). 2022 Aug 17; https://doi.org/10.1109/ICESC54411.2022.9885 67 5

[16]. An evaluation of the CNN-LSTM model's efficacy in sentiment analysis using the BERT and attendance mechanisms. (2024). International Research Journal of Multidisciplinary Scope, 05(02), 822–829. https://doi.org/10.47857/irjms.2024.v05i02.0659