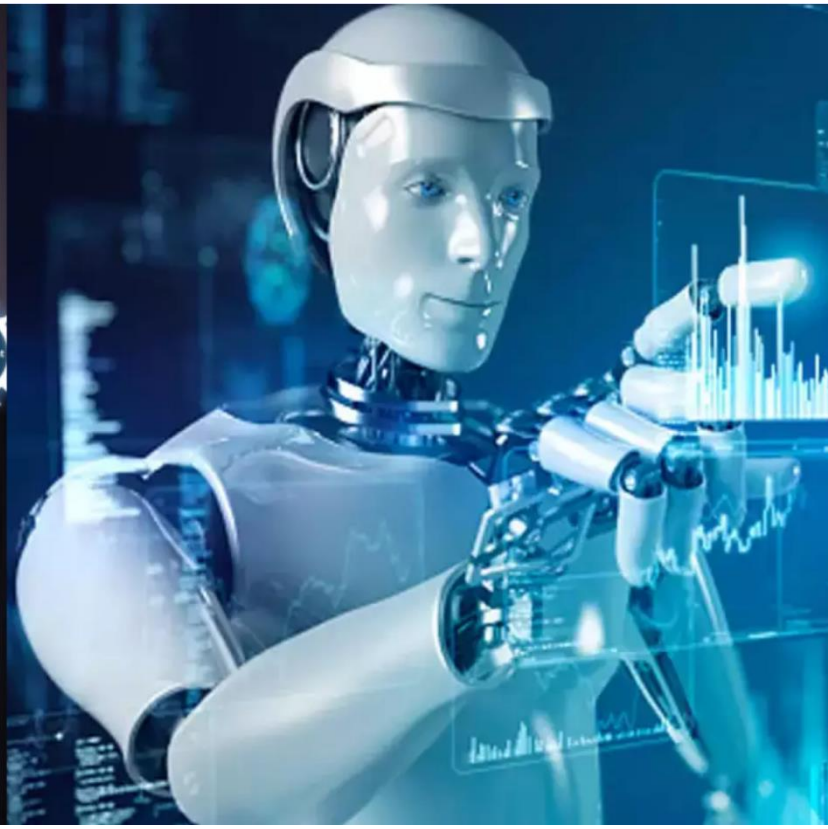
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404545|

Machine Learning Approach for Detection of Financial Fraud Using Value at Risk

Bandaru Harika Hasini, Jayanthi Venkata Satya Sai Suresh Kumar, Yanagala Gowthami, Akkisetty Akash Chaithanya, Gantyada Prasanth

UG Scholar, Dept. of Computer Science Engineering (Artificial Intelligence and Data Science), Satya Institute of Technology and Management, Vizianagaram, India

Kuppili Yasoda Krishna

Assistant Professor, Dept. of Computer Science Engineering (Artificial Intelligence and Data Science), Satya Institute of Technology and Management, Vizianagaram, India

ABSTRACT: As more people utilise online banking services, the large losses that banks and other financial institutions sustained because of new bank account (NBA) fraud are concerning. Machine learning (ML) models have faced significant challenges because to the intrinsic skewness and rarity of NBA fraud cases. This occurs when the number of non-fraud instances exceeds the number of fraud instances, causing the ML models to miss and mistakenly regard fraud as non-fraud instances. Customers' confidence and trust may be damaged by such mistakes. Existing research addresses the skewness of fraud datasets by considering fraud patterns rather than possible losses of NBA fraud risk characteristics. This study suggests NBA fraud detection in the framework of value-at-risk, a risk metric that views fraud cases as the worst-case situation. Value-at-risk models risk features as a skewed tail distribution and estimates possible losses of such attributes using historical simulation. ML was used to classify the risk-return characteristics derived from value-at-risk on the bank account fraud (BAF) dataset. The value-at-risk assigns weight to the skewed NBA fraud cases by managing the fraud skewness with an adjusted threshold probability range. The effectiveness of the fraud detection algorithm was assessed using a unique detection rate (DT) metric that takes risk fraud characteristics into account. A K-nearest neighbour with a true positive (TP) rate of 0.95 and a DT rate of 0.9406 is used to create an enhanced fraud detection model. Value-at-risk offers a clever way to create data-driven standards for fraud risk management within a reasonable loss tolerance in the banking industry.

KEYWORDS: Financial fraud, value at risk, Synthetic Minority Over-sampling Technique, machine learning

I. INTRODUCTION

According to a financial fraud report published by the Association of Certified Fraud Examiners (ACFE) 2022, 2,110 fraud incidents involving financial sectors in 133 countries caused losses of around $3.6 billion [1]. The intentional use of illegal methods or strategies to generate financial benefit is known as financial fraud [2]. Economic disruption, increased living costs, and weakened customer confidence are all possible outcomes of financial fraud [3]. Financial fraud may take many different forms, such as mortgage fraud, credit and debit card fraud, money laundering, insurance fraud, and new bank account fraud [4-5].

"New bank account (NBA) fraud" refers to the practice of creating an account with the intention of committing fraud at banks or other financial institutions [6].

Fraud has wider repercussions, impacting consumers and financial systems through market volatility and fuelling more significant macroeconomic downturns, in addition to causing immediate financial losses and undermining public trust in institutions [7]. Typical characteristics of fraudulent datasets include skewness, changing trends, high dimensionality, and limited access to pertinent data. Studies have been particularly concerned by fraud skewness, which represents the majority fraud class over the non-fraud class, as it has an impact on the effectiveness of fraud detection models. Machine learning systems, including distance-based algorithms, may be negatively impacted by skew

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404545|

fraud cases [8]. Rule-based expert systems, statistical techniques, machine learning, and risk-based approaches have all been developed in the past to combat fraud [9-10].

Numerous studies in the literature assess financial fraud using statistical techniques. In particular, it was discovered that important research used autoregressive (AR) and ordinary least squares (OLS) regression models to assess financial fraud. The study [11] examines the relationship between auditor attributes and fraud detection in emerging economies using a regression model using the Tehran Stock Exchange dataset. The authors offer helpful details to increase the findings' dependability. The study [12] provides insights into the relationship between politics and fraud by examining the impact of political alignment on corporate fraud convictions using pooled OLS and panel regressions. The authors examine state-level data on party affiliation and convictions for corporate fraud in the United States from 2003 to 2018. According to the fraud triangle, the study [13] uses OLS to look at the financial aspects of financial fraud. Using logistic regression, the study [14] found that financial stability and external factors positively impacted financial reporting fraud. False financial reporting, however, is not significantly impacted by hubris, inadequate monitoring, director changes, collusion, or arrogance. Using a bivariate probit model, the study [15] shows how gender diversity contributed to the commission and detection of fraud in Chinese listed companies from 2007 to 2018. According to the authors, female corporate leaders are associated with a greater capacity to identify fraudulent activity, hence reducing the probability of fraudulent activity by enterprises. The study [16], which employs regression analysis to examine Lebanese data, provides insight into the reasons behind fraud and the role of forensic accounting from the perspective of external auditors. According to the study [17], the number of fraud cases and the total amount lost to fraud had a positive impact on the performance of Nigerian money institutions, while the overall number of workers involved in fraud had an adverse effect. The authors' conclusions are reinforced by their use of statistical techniques such OLS regression, Pearson correlation, and descriptive analysis. This research suggests NBA fraud detection in the framework of risk management that treats skewed fraud cases as a worst-case scenario using value-at-risk. Value-at-risk was supplemented with anticipated loss and expected shortfall of frauds, which further quantifies the mean and severe loss impacts, respectively, in order to accurately evaluate the losses of fraud risks. The combination of these risk measurements will make it possible to quantify hazards between the average, worst-case, and extreme situations. Value-at-risk estimates possible losses of risk features using historical simulation. Their risk exposure to fraud risk is the basis for the risk-return characteristics derived from valueat-risk. The NBA fraud detection model receives the risk-return features as input. The K-nearest neighbour fared better than the other machine learning models that were developed. This study makes the following contributions:

- Rather than modelling the fraud pattern, it employed an extreme value theorem to represent the tails (possible losses).
- To more effectively represent the skewness of fraud cases, this article employed value-at-risk.
- Since this study made no assumptions about any distribution, it used historical simulation to determine value-at-risk.

II. METHODOLOGY

The proposed design implemented in which it describes the steps and process involved in NBA fraud detection. Value-at-risk being an important part of this research is designed to model the severe and extreme fraud risk features; it also focuses on rare fraud instances that are detrimental and very costly when occurred. However, the rare cases that are mostly skewed can distort machine learning algorithms especially distance based like KNN. The value-at-risk can handle the fraud skewness through the utilization of adjustable threshold probability ranges (confidence level) unlike the conventional methods that employ constant fraud probability weight that's attached to the skewed fraud instances. The preprocessed, extracted and engineered features were sent as input to value-at-risk for simulation. Meanwhile, a distance based KNN is designed for adjustability to detect fraudulent features through identifying rare clusters with nearest neighbor distance *k*. The confidence level chosen considers the rare fraud cases as higher risk features that would result in fewer training sets, particularly for the KNN model with hyperparameter *k*. The fraud detection model requires the optimization of *k* to a lower setting to sufficiently model the fraudulent features in the rare cluster. The distance weight of KNN is imperative in inhibiting fraud skewness by assigning a higher weight to near instances which in turn facilitates efficient detection of skewed instances.

This paper carries out preprocessing tasks to improve the quality of features and ensure model accuracy. The redundant feature device_fraud_count contains zero instances all of which were manually removed from the data

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404545|

making the model less complex. The categorical features device_os, employment_status, payment_type, and housing_status were labeled to make them easier to learn because machine learning cannot process features that are non-numeric. The features zip_count, keep_alive_session and bank_months_count were eliminated to avoid noise and collinearity issues. The features foreign_request, has_other _card, and email_is_free were eliminated as they give too much undefined log returns.

A. Modules Utilized

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Train & Test Data Sets, View Trained and Tested Datasets Accuracy in Bar Chart, View Trained and Tested Datasets Accuracy Results, View Financial Transaction Type, Find Financial Transaction Type Ratio, Download Predicted Datasets, View Financial Transaction Type Ratio Results, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address, and admin authorize the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like register and login, predict financial transaction type, view your profile.

The development of a NBA fraud detection model was based on the selection of relevant features from demographic, behavioral, risk management and transactional perspective. Demographic features such as *income*, *customer_age*, and *employment_status* were selected. Behavioral features such as *bank_branch_count_8w* and *housing_status* were selected. Risk-based features that include *credit_risk_score* and *proposed_credit_limit* were selected. Transactional features such as *days_since_request*, *total_velocity*, and *payment_type* were also selected. The features engineered are based on location, velocity of transactions, default risk, and ability to repay loans to determine the likelihood of fraudulent behaviors. The
selected features were used along with the other raw features for accurate model training.

B. NBA Fraud Detection Model Selection

Machine learning models can utilize high dimensional data to analyze complex fraud patterns that humans or rule-based systems would not. Supervised ML has unique significance in employing labeled data to find the patterns, anomalies, and fraudulent activity. Binary logistic regression (BLR) is suitable in handling categorical data and is good due its interpretability. Naïve bayes (NB) has efficiency and simplicity in terms of cost and time. K-nearest neighbor (KNN) has high effectiveness in fraud detection when managing transactional information, adaptability, and as well as its potential use in hybrid form.

III. EXPERIMENTAL FINDINGS

A. System Environment

The proposed model is developed using said hardware system configuration and software required to implement the design is shown below.

Hardware System Configuration:

- Processor: Intel i5
- RAM : 8GB
- Hard Disk : 256 GB
- Key Board : Standard Windows Keyboard
- Mouse : Two or Three Button Mouse
- Monitor : SVGA

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404545|

Software Requirements:

- Operating system : Windows 7 Ultimate.
- Coding Language: Python.
- Front-End : Python.
- Back-End : Django-ORM
- Designing : Html, css, javascript.
- Data Base : MySQL (WAMP Server).

B. Experimental Results

The processed results and user interface results are shown in below figures.

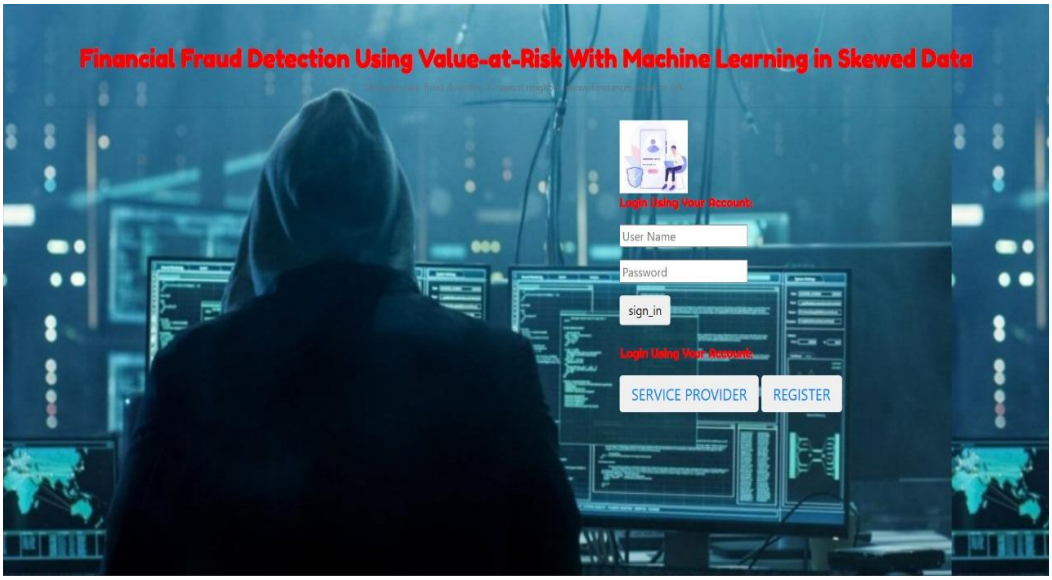


Figure 1. User Interface

Figure 2. Registration Interface

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404545|

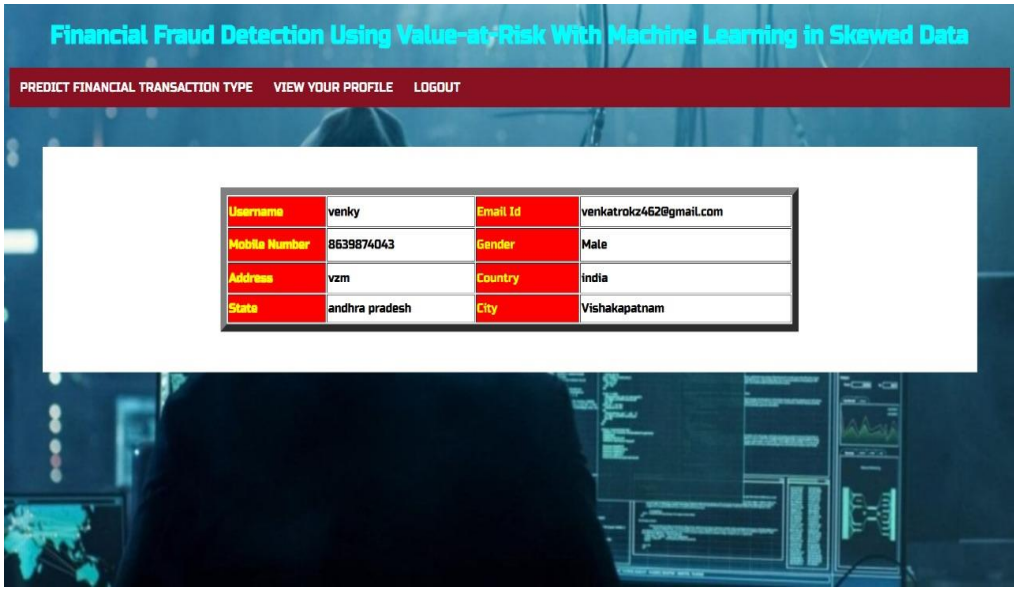


Figure 3. Fraud Detection

IV. CONCLUSION

The value-at-risk-based fraud detection approach described in this work makes it possible to quantify and mitigate fraud risk characteristics while also overcoming the impact of skewed fraud cases, both of which are critical in addressing financial fraud issues. The value-at-risk gives the infrequent fraud scenarios with nearest neighbour distance k a confidence probability weight. By giving nearby examples a larger weight, the distance weight of KNN effectively reduces class skewness and makes it easier to identify skewed instances. By using projected shortfall and expected loss by value at risk, risk may be quantified in mean, worst-case, and extreme situations, allowing their strengths to be aggregated. As a result, a precise fraud detection system helps businesses make wise decisions and lower the total cost of fraud detection and prevention. The experiment's time frames are not considered in this research. The main obstacle, though, is the scarcity of data for NBA fraud detection.

REFERENCES

[1] ACFE. *Association of Certified Fraud Examiners (ACFE) 2022 Report to the Nations*. Accessed: 2023.
[2] T. Ashfaq, R. Khalid, A. S. Yahaya, S. Aslam, A. T. Azar, S. Alsafari, and I. A. Hameed, ‘‘A machine learning and blockchain based efficient fraud detection mechanism,’’ *Sensors*, vol. 22, no. 19, p. 7162, Sep. 2022.
[3] N. S. Alfaiz and S. M. Fati, ‘‘Enhanced credit card fraud detection model using machine learning,’’ *Electronics*, vol. 11, no. 4, 662, 2022.
[4] A. Alfaadhel, I. Almomani, and M. Ahmed, ‘‘Risk-based cybersecurity compliance assessment system (RC2AS),’’ *Appl. Sci.*, vol. 13, no. 10, p. 6145, May 2023.
[5] D. Sarma, W. Alam, I. Saha, M. N. Alam, M. J. Alam, and S. Hossain, ‘‘Bank fraud detection using community detection algorithm,’’ in *Proc. 2nd Int. Conf. Inventive Res. Comput. Appl. (ICIRCA)*, Jul. 2020, pp. 642–646.
[6] A. Pagano, ‘‘Digital account opening fraud on demand deposit accounts: An assessment of available technology,’’ Ph.D. thesis, Utica College, Utica, NY, USA, 2020.
[7] Shuftipro. *New Account Fraud—A New Breed of Scams*. Accessed: 2023.
[8] R. Sasirekha, B. Kanisha, and S. Kaliraj, ‘‘Study on class imbalance problem with modified KNN for classification,’’ in *Intelligent Data Communication Technologies and Internet of Things*, vol. 101. Singapore: Springer, 2022, pp. 207–217.
[9] P. Vanini, S. Rossi, E. Zvizdic, and T. Domenig, ‘‘Online payment fraud: From anomaly detection to risk management,’’ *Financial Innov.*, vol. 9, no. 1, p. 66, Mar. 2023.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404545|

[10] X. Zhu, X. Ao, Z. Qin, Y. Chang, Y. Liu, Q. He, and J. Li, ''Intelligent financial fraud detection practices in post-pandemic era,'' *Innovation*, vol. 2, no. 4, Nov. 2021.
[11] J. Khaksar, M. Salehi, and M. Lari DashtBayaz, ''The relationship between auditor characteristics and fraud detection,'' *J. Facilities Manage.*, vol. 20, no. 1, pp. 79–101, Jan. 2022, doi: 10.1108/jfm-02-2021-0024.
[12] A. Cordis, ''Political alignment and corporate fraud: Evidence from the United States of America,'' *J. Appl. Accounting Res.*, Oct. 2023.
[13] M. J. Rahman and X. Jie, ''Fraud detection using fraud triangle theory: Evidence from China,'' *J. Financial Crime*, vol. 31, no. 1, pp. 101–118, Jan. 2024, doi: 10.1108/jfc-09-2022-0219.
[14] T. Achmad, I. Ghozali, and I. D. Pamungkas, ''Hexagon fraud: Detection of fraudulent financial reporting in state-owned enterprises Indonesia,'' *Economies*, vol. 10, no. 1, p. 13, Jan. 2022.
[15] Aragani, Venu Madhav and Maroju, Praveen Kumar and Mudunuri, Lakshmi Narasimha Raju, Efficient Distributed Training through Gradient Compression with Sparsification and Quantization Techniques (September 29, 2021). Available at SSRN: https://ssrn.com/abstract=5022841 or http://dx.doi.org/10.2139/ssrn.5022841
[15] Y. Wang, M. Yu, and S. Gao, ''Gender diversity and financial statement fraud,'' *J. Accounting Public Policy*, vol. 41, no. 2, Mar. 2022, Art. no. 106903.
[16] J. Hendieh, M. Schneider, and T. Sakr, ''Fraud detection and prevention,'' *Middle-East J. Sci. Res.*, vol. 31, no. 1, pp. 44–52, 2023.
[17] A. Maniatis, ''Detecting the probability of financial fraud due to earnings manipulation in companies listed in Athens stock exchange market,'' *J. Financial Crime*, vol. 29, no. 2, pp. 603–619, Mar. 2022.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details