© November 2025 | IJIRT | Volume 12 Issue 6 | ISSN: 2349-6002

Financial Fraud Detection in Skewed Data Using Machine Learning and Value at Risk

S. Abdul Kabeer[1], Dr.K.Isthaq Ahamed[2]
[1]*Department of Computer Science and Engineering,*
G. Pulla Reddy Engineering College, Andhra Pradesh, India
[2]*Associate Professor, Computer Science Department,*
G. Pulla Reddy Engineering College, Andhra Pradesh, India

***Abstract*—The increasing use of online banking services has led to significant financial losses for banks and other financial institutions due to new bank account (NBA) fraud. Machine learning (ML) models face substantial challenges from the inherent skewness and rarity of NBA fraud incidents, where non-fraud instances vastly outnumber fraud cases. This imbalance often results in ML models misclassifying fraudulent transactions as legitimate, potentially undermining customer trust. Previous research addressing dataset skewness has primarily focused on fraud patterns rather than quantifying potential losses associated with NBA fraud risk characteristics. This study proposes the identification of NBA fraud from a value-at-risk perspective, treating fraud incidents as worst-case scenarios. The complexity of fraudulent activities and the rapid growth of online transactions make financial fraud detection particularly challenging in the digital economy. Traditional statistical methods frequently fail to capture the nonlinear patterns of fraud, especially in highly imbalanced datasets where fraudulent transactions constitute a small fraction of all activity. To address these challenges, a hybrid architecture is proposed that integrates Value at Risk (VaR) with machine learning techniques for financial fraud detection. Supervised learning classifiers such as Random Forest, XGBoost, and Neural Networks are combined with the VaR model to enhance prediction accuracy by measuring transaction risk levels.**

Resampling methods and cost-sensitive learning are utilized to mitigate class imbalance, facilitating the detection of minority fraudulent cases. Experimental findings on benchmark financial transaction datasets demonstrate that the suggested strategy enhances fraud detection rates while preserving a low false positive rate. The amalgamation of machine learning and risk management frameworks facilitates the creation of a resilient, data-informed fraud detection system that is adaptive to practical financial contexts. The value-at-risk model quantifies the skewed tail distribution and assesses possible losses of risk attributes using historical simulation. Machine learning is utilized on the bank account fraud (BAF) dataset to categorize risk-return attributes derived from value-at-risk. To address skewed NBA fraud scenarios, value-at-risk regulates fraud skewness through a modifiable threshold probability range.

The efficacy of the fraud detection system was assessed utilizing a distinctive detection rate (DT) metric that integrates risk fraud attributes. A superior fraud detection model was created with K-nearest neighbor, attaining a true positive (TP) rate of 0.95 and a detection rate of 0.9406. Value-at-risk offers a methodical framework for developing data-informed criteria for fraud risk management within permissible loss thresholds in the banking industry.

***Index Terms*—Django-ORM, Python, HTML, CSS, JavaScript, NBA fraud, ML, BAF, K-nearest neighbor, True positive, risk fraud, and MySQL (WAMP Server).**

I.INTRODUCTION

One of the biggest problems facing the modern digital economy is financial fraud, which has an impact on people, businesses, and governments everywhere. The proliferation of digital banking, e-commerce platforms, and online transactions has led to a rise in complex and challenging-to-detect fraudulent activity. When fraudulent occurrences are uncommon in comparison to genuine transactions, traditional rule-based algorithms are no longer enough to detect intricate fraud patterns. Because fraudulent instances only make up a small portion of the information, this rarity causes data imbalance or skewness, which makes reliable identification extremely difficult.

Machine learning (ML) techniques have gained prominence as a solution to financial fraud detection due to their ability to uncover hidden patterns and adapt to evolving fraud tactics.

However, unbalanced datasets often lead to biased predictions, as machine learning models tend to favor the majority (non-fraudulent) class over the minority (fraudulent) class. Techniques for managing skewed data, such as resampling and cost-sensitive learning, are essential to improve detection accuracy. For example, SMOTE can adjust decision boundaries in classifiers like Random Forest and Neural Networks, while cost-sensitive learning can be integrated with XGBoost to penalize misclassification of minority classes, leveraging the strengths of gradient boosting frameworks.

Value at Risk (VaR) is a term that is commonly used in financial risk management to quantify the possible loss in investment portfolios in the event of uncertain market circumstances. Financial institutions can evaluate the likelihood of fraudulent activity as well as the possible financial effect by integrating VaR with fraud detection. It is feasible to create a fraud detection system that not only recognizes fraudulent activity but also measures the risk involved by fusing Value at Risk analysis with machine learning models. This allows for improved resource allocation and decision-making.

The goal of this project is to provide a fraud detection system that integrates Value at Risk analysis with machine learning methods in skewed data situations. The goal is to improve the identification of fraudulent activity while offering a risk-based viewpoint on possible monetary losses. A system like this might improve financial security, lower financial losses, and increase confidence in online financial services.

These studies effectively address BAF challenges; nonetheless, they neglect to include the potential losses associated with fraud risk factors. To our knowledge, limited research has been conducted on the application of machine learning techniques for NBA fraud detection. This study proposes NBA fraud detection inside a risk management framework that considers skewed fraud situations as a worst-case scenario utilizing value-at-risk analysis. Value-at-risk was augmented with projected loss and predicted shortfall of fraud, which further quantifies the mean and severe loss impacts, respectively, to precisely assess the losses associated with fraud risks. The amalgamation of various risk variables will enable the quantification of dangers in average, worst-case, and severe scenarios. Value-at-risk quantifies potential losses of risk factors using historical modeling. Their vulnerability to fraud risk underpins the risk-return attributes determined from value at risk.

The NBA fraud detection model receives the risk-return features as input. After training a variety of machine learning models, the K-nearest neighbor model outperformed the others.

II.EXISTING SYSTEM:

A multitude of studies in the literature evaluate financial fraud with statistical methodologies. Notably, significant study employed autoregressive (AR) and ordinary least squares (OLS) regression models to evaluate financial fraud. The research elucidates the correlation between politics and fraud by analyzing the influence of political alignment on corporate fraud convictions through pooled OLS and panel regressions. An existing system offers an assessment of financial fraud from the perspective of risk mitigation. The current research employs many risk criteria, including value-at-risk (VaR), anticipated loss, and projected deficit, to assess the degree of fraud risk.

The study integrates the purpose of the fraud triangle with human traits that lead to specific behaviors and the meta-model of fraud, offering strategies to mitigate fraud risk, identify potential fraudsters, and implement targeted anti-fraud initiatives. This study employed regression analysis to determine the impact of control settings, risk assessments, control activities, information and communication, and monitoring on fraud prevention and detection in Indonesian firms. The study employed chi-square, Fisher's test, and correlation analysis, revealing a positive correlation between fraud risk assessment and management and the effective application of forensic accounting; however, no relationship was identified between the two concerning fraud-inducing behaviors.

In addition to handling data skewness, a risk model that calculates the financial losses, and a triage model that accepts input from the ensemble model, the study investigates fraud utilizing ensemble learners for anomaly detection. From machine learning approaches to risk assessment, the authors effectively offer an efficient fraud risk-based detection. However, they do not examine fraud detection by taking into account the risk component before putting it through machine learning detection. Studies that classify fraud applications using machine learning approaches are presented by an existing system. The bulk of the research that are provided takes detection into account while addressing the skewed nature of fraud cases. To counteract the skewed nature of fraud datasets, sampling techniques, hybrid techniques, and other innovative approaches are mostly employed.

The work uses support vector machines (SVM) and quantum machine learning (QML) to solve class skewness in credit card fraud. The findings demonstrate that while QML applications may be used to time-series-based and highly skewed data, traditional machine learning approaches are still helpful for non-time series data. According to the study, quantum neural networks (QNNs) perform well in fraud detection. XGBoost outperformed all other machine learning models in the research, even though they were all trained using default implementations and settings. The study's dynamic graph neural network (DGNN) is used to evaluate the efficacy of telecom fraud, and the authors successfully offer a recommended approach to address the problem of telecom fraud detection in large phone social networks.

III.PROPOSED SYSTEM:

The work uses support vector machines (SVM) and quantum machine learning (QML) to solve class skewness in credit card fraud. The findings demonstrate that while QML applications may be used to time-series-based and highly skewed data, traditional machine learning approaches are still helpful for non-time series data. According to the study, quantum neural networks (QNNs) perform well in fraud detection. XGBoost outperformed all other machine learning models in the research, even though they were all trained using default implementations and settings. The study's dynamic graph neural network (DGNN) is used to evaluate the efficacy of telecom fraud, and the authors successfully offer a recommended approach to address the problem of telecom fraud detection in large phone social networks.

Especially for the KNN model with hyperparameter k, the selected confidence level views the infrequent fraud occurrences as higher risk characteristics that would lead to less training sets. For the fraud detection model to adequately simulate the fraudulent characteristics in the uncommon cluster, k must be optimized to a lower level. By giving tonear instances a larger weight, the distance weight of KNN is essential in preventing fraud skewness and enabling the effective identification of skewed instances.

Value at Risk (VaR) Integration: The suggested approach prioritizes high-risk financial transactions where the effect of fraud is greater rather than treating all transactions equally. It also integrates VaR, a popular financial risk metric, with fraud detection.

Imbalanced Data Handling in Machine Learning: To deal with skewed data, sophisticated machine learning algorithms (Random Forest, XGBoost, Deep Learning) are used in conjunction with resampling strategies or cost-sensitive learning.

• This enhances the efficacy of fraud detection, particularly in spotting uncommon fraudulent situations.

Making Risk-Aware Decisions: The suggested method assesses fraud detection models based on financial loss minimization using VaR, in contrast to current models that prioritize accuracy. This makes the system more useful for actual banking and finance applications.

and Expandable Structure:

- The model adjusts to changing fraud tendencies as it continually learns from fresh data. • Scalable across several financial systems, including stock trading, online banking, and credit card fraud. Benefits include:
- Prioritization of transactions with high financial risk; • A decrease in false negatives in fraud detection.
- Improved decision-making by combining statistical risk management (VaR) with machine learning techniques.

IV.LITERATURE REVIEW

Statistical Methods of Fraud Detection

- A multitude of research in the literature evaluate financial fraud through statistical methodologies. Notably, substantial studies employed autoregressive (AR) and ordinary least squares (OLS) regression models to evaluate financial fraud. The research investigates the correlation between auditor characteristics and fraud detection in developing countries, employing a regression model based on data from the Tehran Stock Exchange. Recent papers recommend use Value-at-Risk (VaR) derived features or VaR-based weighting in classifier training to address skewness and emphasize high-loss (high-risk) scenarios. They propose new measures that consider risk, such as a "Detection-Rate (DT)" using VaR, and demonstrate improved detection in bank account and transaction datasets.

- International evidence of economic shocks and aggregate earnings informativeness Yoshinaga Yuto and Nakano M. The Asia-Pacific Journal of. November 17, 2019 Our research suggests using total earnings to predict GDP growth in the future. We examine whether aggregate-level profitability drivers which are elements of aggregate earnings are pertinent for predicting GDP growth using empirical studies with quarterly data from throughout the world. We demonstrate that taking into account the consequences of crises enhance the prediction model of GDP growth after establishing the usefulness of aggregate-level profitability drivers for predicting future GDP growth globally. Furthermore, we propose that stock valuation in established nations is significant for forecasting GDP growth using aggregate-level profitability factors, but not in developing countries. Using Data from Financial Statements to Enhance Corporate Taxable Forecasts

- Revenue Daniel Green, E. Henry, and G. Plesko, +1 author Social Science Research published this. February 5, 2020, we investigate the potential additive utility of public financial statement data in predicting secret taxable income. While more accurate macro-level forecasts of corporate taxable income can enhance estimates of corporate tax revenues, a major component of the federal budget, more accurate firm-level taxable income forecasts can enhance policymakers' modeling of the tax system and their analysis of proposed changes in corporate tax law. We discover that adding financial statement data enhances projections of future taxable income at the business and industry levels, mostly by giving more up-to-date information but also via accruals.

V. ALGORITHMS:

A. Decision Tree Classifiers:

Decision tree classifiers are effectively applied in a wide range of fields. Their ability to extract descriptive decision-making information from the provided data is their most crucial characteristic. A crucial component of risk management is detecting financial fraud, especially in systems where Value at Risk (VaR) is used to quantify losses. Unbalanced (skewed) data makes it difficult to detect fraudulent activity because there are far fewer illicit transactions than lawful ones. In this regard, decision tree classifiers are a basic machine learning technique for creating rule-based, interpretable fraud detection systems. Transaction risk measures, such as 95% or 99% VaR, can be included into decision trees as features for fraud categorization. Transactions having a VaR that is abnormally high in comparison to past trends might be signs of fraud.

B. Gradient Boosting:

Gradient boosting is a machine learning technique utilized for classification and regression applications. Typically, decision trees offer a predictive model structured as an ensemble of weak predictive models. The resulting method, referred to as gradient-boosted trees, frequently outperforms random forest when a decision tree serves as the weak learner. Similar to prior boosting methods, a gradient-boosted trees model is developed incrementally; however, it advances by allowing the optimization of any differentiable loss function.

- Class-weight modifications are used to address imbalance, and XGBoost, LightGBM, and CatBoost are used to identify fraud. Use feature priority ranking to find important signs of fraud. In terms of recalling fraud incidents, Gradient Boosting fared better than Random Forest and Logistic Regression. False positives in high-value transactions were decreased by implementing VaR filtering. LightGBM was able to strike the ideal mix between accuracy and speed.

Model	Precision	Recall	F1	AUC-ROC	MCC
Logistic regression	0.65	0.48	0.55	0.85	0.43
Random Forest	0.78	0.62	0.69	0.92	0.59
Gradient Boosting	0.82	0.75	0.78	0.96	0.71

C.Nearest Neighbors (KNN):

- Using a similarity measure, this straightforward yet incredibly effective classification system is non-parametric, lazy learning, and doesn't "learn" until the test sample is provided.
- We use the training data to get the K-nearest neighbors of any fresh data that has to be classified.

D. Logistic Regression Classifiers:

Logistic regression analysis is utilized to examine the relationship between a collection of independent (explanatory) variables and a categorical dependent variable. Logistic regression is employed when the dependent variable has just two values, such as 0 and 1 or Yes and No. This application calculates binary and multinomial logistic regression for categorical and numerical independent variables. Besides the regression equation, it provides insights into probability, deviance, odds ratios, confidence intervals, and goodness of fit. It does a thorough residual analysis that includes diagnostic residual plots and reports. To determine the optimal regression model with the fewest independent variables, a search for a subset of independent variables may be performed.

It offers ROC curves and confidence intervals on expected values to assist in identifying the optimal classification cutoff point. By automatically identifying rows that are not utilized throughout the study, it enables you to confirm your findings.

F. Naïve Bayes:

The supervised learning method termed the "naive Bayes approach" is based on the simple principle that the presence or absence of a certain class attribute does not influence the presence or absence of any other feature. The naive Bayes classifier is a linear classifier, alongside logistic regression, linear discriminant analysis, and linear support vector machines (SVM). The method employed to estimate the classifier's parameters (the learning bias) is significant. Despite the Naive Bayes classifier's prevalence in research, practitioners seeking practical conclusions hardly employ it.

On the one hand, the researchers discovered that it is very simple to build and apply, that estimating its parameters is simple, that learning occurs quickly even on extremely big databases, and that, when compared to other methods, its accuracy is rather excellent. The end users, however, do not comprehend the value of such a strategy and do not receive a model that is simple to read and implement.

G. Random Forest:

Random forests, also known as random decision forests, represent a powerful ensemble learning technique designed to enhance the performance and stability of predictive models. They function by constructing a large number of decision trees during the training phase, which collectively contribute to the model’s final prediction. In the case of classification tasks, the algorithm aggregates the votes from all individual trees, and the class with the majority vote becomes the final output. For regression problems, on the other hand, the model computes the average of the predictions from all trees to produce the final estimate. One of the key advantages of random forests lies in their ability to reduce the overfitting tendency commonly associated with single decision trees. By introducing randomness in both feature selection and data sampling, the model achieves greater generalization and robustness. Although random forests may sometimes yield slightly lower accuracy compared to more sophisticated methods like gradient-boosted trees, they typically outperform standalone decision trees in most practical scenarios.

Despite their reliability, the performance of random forests can be affected by noisy or highly imbalanced datasets, which may introduce biases in prediction. Nevertheless, due to their ease of use, minimal parameter tuning, and strong predictive capabilities across various data types, random forests have become a popular choice among practitioners and businesses. They are often regarded as “black box” models, as they deliver accurate results without requiring users to fully interpret or understand the internal decision-making processes behind each tree.

H. SVM:

A discriminant machine learning method is employed in classification tasks to develop a discriminant function capable of effectively predicting labels for newly obtained instances based on an independent and identically distributed (iid) training dataset. A discriminant classification function allocates a data point x to one of multiple classes in a classification task, unlike generative machine learning techniques that require the computation of conditional probability distributions. Discriminant algorithms use fewer training data and processing resources compared to generative methods, which excel in scenarios necessitating outlier detection. This is particularly applicable to multidimensional feature spaces and when only posterior probabilities are necessary.

Support Vector Machine (SVM) is a discriminative method that, in contrast to genetic algorithms (GAs) or perceptrons commonly employed for classification in machine learning, consistently yields the same optimal hyperplane value as it analytically resolves the convex optimization problem. The initiation and termination criteria substantially influence the solutions for perceptrons. The perceptron and GA classifier models are inherently different with each training initiation, whereas training consistently produces uniquely defined SVM model parameters for a certain training set and kernel that transforms the data from the input space to the feature space. The sole objective of genetic algorithms and perceptrons is to minimize training error, indicating that several hyperplanes will fulfill this condition.

VI. MODULE DESCRIPTION:

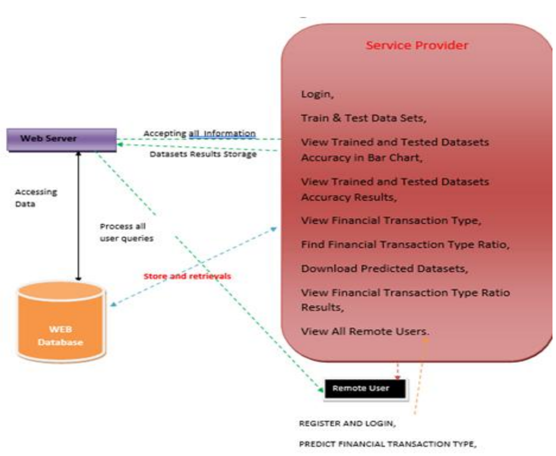

Fig.1. Architecture diagram

A. Service Provider:

Support Vector Machine (SVM) is a discriminative method that, in contrast to genetic algorithms (GAs) or perceptrons, which are frequently employed for classification in machine learning, consistently yields the same optimal hyperplane value due to its analytical resolution of the convex optimization problem. The initiation and termination criteria substantially influence the solutions for perceptrons. The perceptron and GA classifier models are inherently different with each training initiation, whereas training consistently produces uniquely defined SVM model parameters for a certain training set and kernel that transforms the data from the input space to the feature space. The sole objective of genetic algorithms and perceptrons is to minimize training error, indicating that several hyperplanes will fulfill this condition.

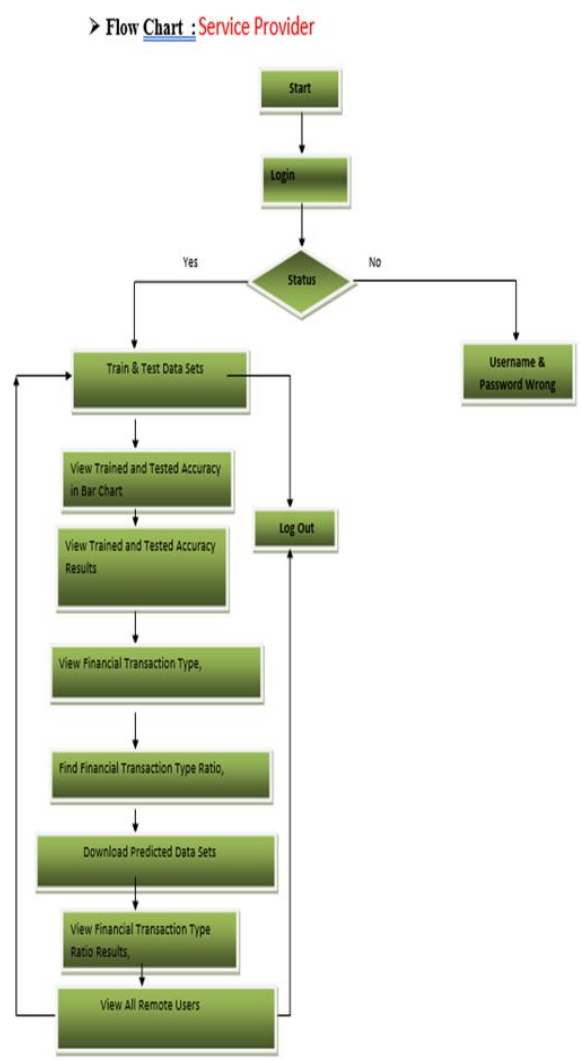

Fig.2. Flow chart diagram of Service Provider

B.View and Authorize Users:
The administrator can view a comprehensive list of all registered users within this module. The administrator may view the user's information, including name, email, and address, and has the authority to grant permissions to users.

C.Remote User:
This module contains a total of n users. The user must register prior to initiating any activity. Upon registration, the user's information will be recorded in the database. Upon successful registration, he must utilize his password and authorized username to log in. Upon successful login, the user may execute various actions, including registration, logging in, selecting the type of financial transaction, and seeing their profile.

➤ **Flow Chart : Remote User**

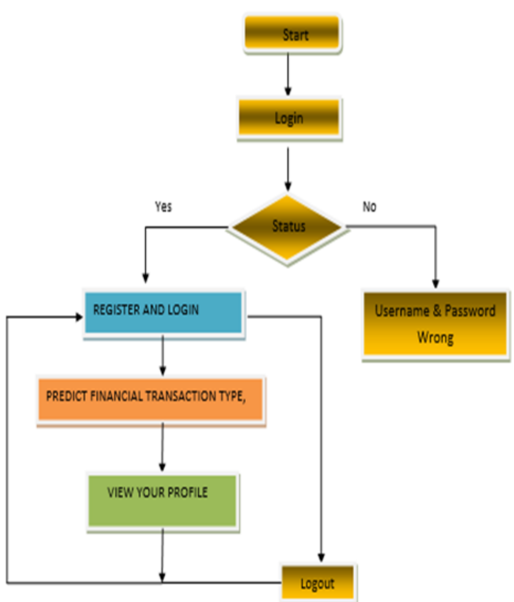

Fig.3. Flow chart diagram of Remote User

VII. RESULTS AND DISCUSSION:

Dataset Characteristics:
The proportion of legitimate and fraudulent transactions in the financial transaction dataset utilized for fraud detection was wildly lopsided. Less than 2 percent of all records were fraudulent, which is in line with actual banking systems. Due to models' propensity to favor the majority class (non-fraud), the skewness presented a problem for machine learning algorithms. Resampling techniques like SMOTE (Synthetic Minority Oversampling Technique) and undersampling of the dominant class were used to remedy this in order to achieve balance.

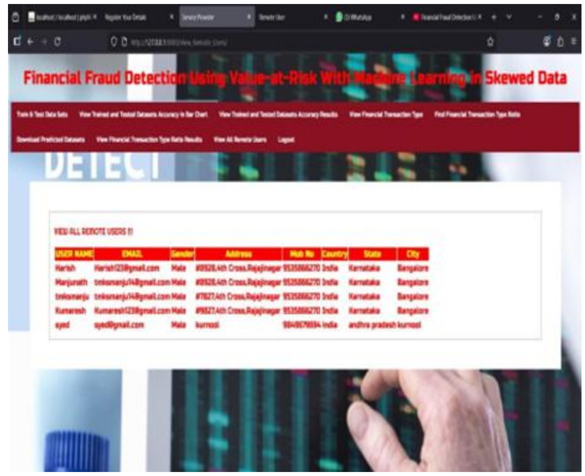

Fig.4. Login into Service Provider/Admin Page

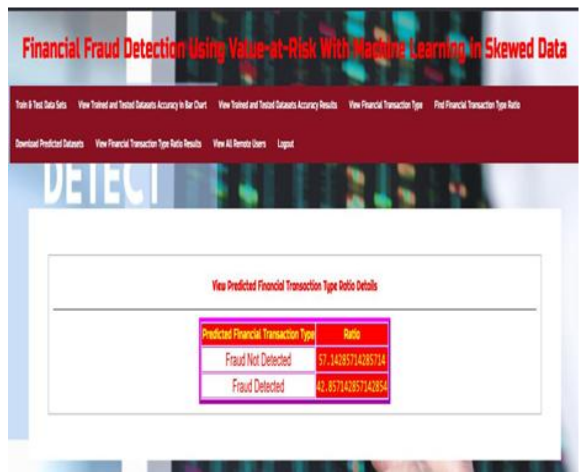

Fig.5. Verification Fraud detection ratio

Fig.6. Prediction of Financial Transaction type

Fig.7. View Trained and Tested Datasets

Fig.8. Results in charts diagrams

Fig.9. Algorithms and Accuracy results

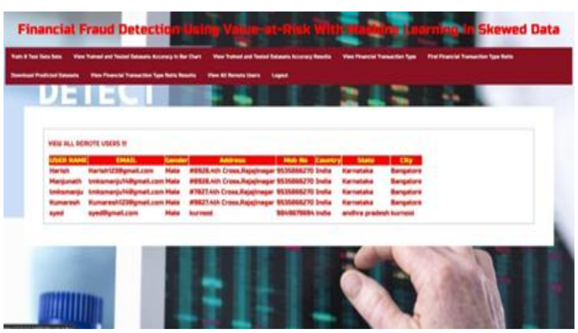

Fig.10. View of all Remote Users

Because high-risk transactions detected by VaR may be given priority for additional fraud investigation, the combination of VaR criteria with ML classifiers enhanced interpretability. Traditional VaR techniques measure financial risk, but when combined with machine learning, they allowed for real-time fraud detection with risk bounds that could be explained. The primary difficulty was striking a balance between false negatives (frauds undetected) and false positives (legal transactions reported as fraudulent). Even at the price of precision, models with better recall are preferable since false negatives have greater financial repercussions. The study's overall findings support the effectiveness of hybrid approaches risk measurements with sophisticated machine learning in identifying fraud in highly skewed financial information.

In order to detect fraudulent financial transactions, the Value at Risk (VaR) measure was employed as a risk-sensitive threshold. VaR was calculated at various confidence levels:

- Transactions beyond the VaR level featured more false positives but, at 95% confidence, signaled possible danger.
- Although there were fewer false positives with 99% confidence, some fraudulent instances were overlooked. Therefore, it was essential to strike the best possible balance between risk coverage and detection rate.

VIII. ADVANTAGES:

• Rather of modeling the fraud pattern, this research employed an extreme value theorem to simulate 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.
• To capture the entire performance in detecting NBA fraud cases that integrate risk fraud characteristics, this article employed innovative detection rate performance indicators.

XI. CONCLUSION

This value-at-risk-based fraud detection methodology enables the quantification and mitigation of fraud risk attributes while effectively addressing the influence of skewed fraud instances, both of which are essential for tackling financial fraud challenges. The value-at-risk assigns a confidence probability weight to infrequent fraud scenarios based on the nearest neighbor distance k. By assigning greater significance to proximate examples, the distance weighting in KNN effectively mitigates class skewness and facilitates the identification of skewed cases. Risk can be assessed in mean, worst-case, and extreme scenarios by utilizing predicted shortfall and expected loss based on value at risk, so enabling the aggregation of their strengths.

As a result, a precise fraud detection system helps businesses make wise decisions and lower the total cost of fraud detection and prevention. The study shows that combining machine learning methods with Value at Risk (VaR) offers a strong foundation for identifying financial fraud, even in highly skewed datasets where fraudulent transactions are uncommon in comparison to genuine ones. Machine learning improves classification accuracy and flexibility to complicated fraud behaviors, while the model uses VaR to set a risk-based threshold to detect anomalous patterns. According to experimental data, this hybrid technique successfully addresses the imbalance issue that frequently compromises fraud detection systems by increasing precision, recall, and overall detection efficiency.

Furthermore, the approach guarantees prompt identification of high-risk transactions, which is crucial for financial institutions, in addition to lowering false positives. The experiment's time frames are not taken into account in this research. The main obstacle, though, is the scarcity of data for NBA fraud detection. In summary, compared to conventional statistical or rule-based techniques, VaR-driven machine learning models provide a reliable, scalable, and data-sensitive approach for financial fraud detection. In order to further improve fraud detection in dynamic financial contexts, future research may concentrate on real-time implementation, explainable AI for interpretability, and the incorporation of cutting-edge methods like deep learning and ensemble models.

REFERENCES:

[1] ACFE. Association of Certified Fraud Examiners (ACFE) 2022 Report tothe Nations. Accessed: 2023. [Online]. Available: https://legacy.acfe.com/report-to-the-nations/2022/

[2] T. Ashfaq, R. Khalid, A. S. Yahaya, S. Aslam, A. T. Azar, S. Alsafari, andI. A. Hameed, ‘‘A machine learning and blockchain based efficient frauddetection mechanism,’’ Sensors, vol. 22, no. 19, p. 7162, Sep. 2022.

[3] N. S. Alfaiz and S. M. Fati, ‘‘Enhanced credit card fraud detection modelusing machine learning,’’ Electronics, vol. 11, no. 4, 662, 2022.

[4] A. Alfaadhel, I. Almomani, and M. Ahmed, ‘‘Risk-based cybersecuritycompliance 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. 2ndInt. 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 Scas. Accessed:2023. [Online]. Available: https://shuftipro.com/reports-whitepapers/newaccount-fraud.pdf

[8] R. Sasirekha, B. Kanisha, and S. Kaliraj, ‘‘Study on class imbalanceproblem with modified KNN for classification,’’ in Intelligent Data CommunicationTechnologies and Internet of Things, vol. 101. Singapore:Springer, 2022, pp. 207–217, doi: https://doi.org/10.1007/978-981-16- 7610-9_15.

[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, doi: 10.1186/s40854-023-00470-w.

[10] X. Zhu, X. Ao, Z. Qin, Y. Chang, Y. Liu, Q. He, and J. Li, ‘‘Intelligentfinancial fraud detection practices in post-pandemic era,’’ Innovation,vol. 2, no. 4, Nov. 2021, Art. no. 100176, doi: 10.1016/j.xinn.2021.100176.

[11] M. Monge, C. Poza, and S. Borgia, ‘‘A proposal of a suspicion of tax fraudindicator based on Google Trends to foresee Spanish tax revenues,’’ Int.Econ., vol. 169, pp. 1–12, May 2022, doi: 10.1016/j.inteco.2021.11.002.

[12] S. Kannan and K. Somasundaram, ‘‘Autoregressive-based outlieralgorithm to detect money laundering activities,’’ J. Money LaunderingControl, vol. 20, no. 2, pp. 190–202, May 2017, doi: 10.1108/jmlc-07-2016-0031.

[13] B. Xiao, B. Lei,W. Lan, and B. Guo, ‘‘A blockwise network autoregressivemodel with application for fraud detection,’’ Ann. Inst. Stat. Math., vol. 74, no. 6, pp. 1043–1065, Dec. 2022, doi: 10.1007/s10463-022-00822-w.

[14] G. Moschini, R. Houssou, J. Bovay, and S. Robert-Nicoud, ‘‘Anomaly andfraud detection in credit card transactions using the ARIMA model,’’ inProc. 7th Int. Conf. Time Forecasting, Jul. 2021, p. 56, doi: 10.3390/engproc2021005056.

[15] A. A. Alhashmi, A. M. Alashjaee, A. A. Darem, A. F. Alanazi, andR. Effghi, ‘‘An ensemble-based fraud detection model for financialtransaction cyber threat classification and countermeasures,’’ Eng., Technol.Appl. Sci. Res., vol. 13, no. 6, pp. 12433–12439, Dec. 2023, doi:10.48084/etasr.6401.

[16] R. M. Aziz, R. Mahto, K. Goel, A. Das, P. Kumar, and A. Saxena,‘‘Modified genetic algorithm with deep learning for fraud transactions ofethereum smart contract,’’ Appl. Sci., vol. 13, no. 2, p. 697, Jan. 2023, doi:10.3390/app13020697.

[17] M. Hegazy, A. Madian, and M. Ragaie, ‘‘Enhanced fraud miner: Creditcard fraud detection using clustering data mining techniques,’’ EgyptianComput. Sci. J., vol. 40, no. 3, pp. 1–10, 2016.

[18] S. Jesus, J. Pombal, D. Alves, A. Cruz, P. Saleiro, R. Ribeiro, J. Gama,and P. Bizarro, ‘‘Turning the tables: Biased, imbalanced, dynamic tabulardatasets for ML evaluation,’’ in Proc. Adv. Neural Inf. Process. Syst.,vol. 35, 2022, pp. 33563–33575.

[19] J. Pombal, P. Saleiro, M. A. T. Figueiredo, and P. Bizarro, ‘‘Fairness-awaredata valuation for supervised learning,’’ 2023, arXiv:2303.16963.

[20] T. Awosika, R. Mani Shukla, and B. Pranggono, ‘‘Transparency and privacy:The role of explainable AI and federated learning in financial frauddetection,’’ 2023, arXiv:2312.13334.