ADVANCED RESEARCH & INNOVATIONS
International Journal of Advanced Research & Innovations
Volume 15, Issue-2
July, 2025, P02
ISSN: 2319-9253

VALUE-AT-RISK DRIVEN FRAUD DETECTION FRAMEWORK WITH MACHINE LEARNING UNDER DATA IMBALANCE

POKALA LAVANYA, M.Tech, Dept of CSE,
Dr. T. RAVI KUMAR, Professor, Department of CSE,
Vaageswari College of Engineering (Autonomous), Karimnagar, Telangana.

ABSTRACT: The minimal presence of illicit operations within the overall data creates a significant challenge in detecting financial crime due to data imbalance. This paper delineates a revolutionary methodology for fraud detection. The integration of XGBoost, Random Forest, and neural networks enhances the accuracy of Value-at-Risk (VaR). Techniques such as cost-sensitive learning and SMOTE are utilized to address the imbalance and ensure the identification of fraudulent cases. This technique assists financial organizations in mitigating potential losses by concentrating on high-risk scam scenarios and utilizing Value at Risk theories. This technology has proven in practical studies its capacity to mitigate financial risk and accelerate scam detection. It provides an innovative, risk-aware methodology for transaction security.

Keywords: Value-at-Risk (VaR), Fraud Detection, Machine Learning, Data Imbalance, Cost-Sensitive Learning and Financial Risk Analysis.

1. INTRODUCTION

Financial frauds are becoming an increasingly pressing concern for government agencies, banks, and fintech companies in today's lightning-fast digital economy. Since more transactions are taking place online, scammers have more opportunities to make money. Because of this, they are constantly innovating, rendering obsolete the effectiveness of conventional fraud tracking techniques. Financial organizations often rely on rule-based systems, despite their outdated nature and inability to address evolving fraud tendencies. Scam detection systems that can learn from data and adapt are in high demand among businesses. These systems should be able to keep up with unlawful actions as they emerge. Because of the data's inconsistency, frauds are difficult to detect. Machine learning algorithms could erroneously prioritize legitimate transactions and overlook instances of fraud due to the small percentage of financial transactions that constitute fraud. Because of the unequal distribution, many models are trained to appear correct, yet they fail miserably when it comes to detecting fraudulent actions. To prevent fraud detection systems from being overwhelmed by the vast number of legitimate agreements, it is necessary to resolve this difference.

Unlike other fraud detection methods, the Value-at-Risk (VaR) approach incorporates financial risk evaluation into its steps. A crucial financial metric, Value at Risk (VaR) quantifies the probability of monetary loss over a specified time frame and degree of certainty. This method assigns monetary values to potentially fraudulent transactions based on the Value at Risk (VaR) principles of fraud detection. As an alternative to merely detecting fraudulent transactions, it accomplishes this. In this way, financial institutions can maximize the use of their investigation resources by swiftly addressing high-risk fraud cases.

Because of the complexity of the transaction trends at play here, sophisticated machine learning algorithms are required to spot discrepancies that may indicate fraud. To enhance the precision of fraud detection,

Volume 15, Issue-2
July, 2025, P02
ISSN: 2319-9253

techniques such as Deep Neural Networks, Gradient Boosting, and Random Forest are employed. Since fraudulent trades are uncommon, the dataset is leveled using cost-sensitive learning, Adaptive Synthetic Sampling (ADASYN), and the Synthetic Minority Oversampling Technique (SMOTE). While these methods make it easier to recognize phony circumstances, they also retain the model's general usefulness. By incorporating VaR into the machine learning system, the emphasis moves from reducing financial losses to improving the ability to detect fraud. Detecting fraud is gaining significance in real-life banking contexts due to this approach's emphasis on reducing financial risk. When evaluating their performance, the majority of conventional fraud detection systems rely on accuracy or the F1-score. By directing the attention of risk managers and compliance teams toward the most costly fraud scenarios, the risk-aware approach makes money safer.

The VaR-driven fraud detection system is a significant advancement in the battle against financial crime. As a result, fraud detection extends well beyond the simple act of identifying suspicious transactions. Machine learning and financial risk measures work together to shield companies from catastrophic losses. By addressing issues of data imbalance, this new technology aids fintech and banking institutions in their fight against fraud in a more intelligent and economical manner. Protecting financial activities is made smarter, more scalable, and more risk-aware with this technology, which employs modern data science techniques and money expertise.

2. REVIEW OF LITERATURE

Xu, H., & Wang, H. (2020). This thorough research examines how machine learning methods are used in fraud detection systems in a number of domains. It categorizes techniques such as decision trees, neural networks, and clustering algorithms as supervised, unsupervised, and hybrid models after thoroughly analyzing them. The research looks at the importance of preprocessing, feature selection, and real-time object recognition. Divergent viewpoints, data privacy, and socioeconomic inequality are among the topics covered. The benchmark datasets used in the research are summarized by the authors. They give examples of assessment processes and highlight crucial performance metrics. The report examines potential directions for further research in addition to pointing out flaws in current approaches. Clarity of deployment and scalability are given top priority. Everyone, including employees and students, can use this tool.

Chen, C., Li, Y., & Li, C. (2020). This research offers a strong ensemble learning method to address the challenges of class imbalance in fraud detection. To increase speed, the suggested method makes use of a variety of classifiers, such as Random Forest, AdaBoost, and Gradient Boosting. Using oversampling and cost-sensitive learning strategies, the system is trained to detect infrequent fraud occurrences. Experiments are conducted using empirical transaction data. The outcomes are better than those of traditional classifiers in terms of precision, accuracy, and recall. To withstand various forms of financial fraud, the design can be improved and changed. The results of a comparison investigation using baseline techniques verify the effectiveness. Concerns around overfitting and generalization are also covered by the writers. The findings significantly advance the investigation of class-imbalance learning in fraud detection.

Zhang, X., Li, H., & Li, J. (2020). A cost-sensitive version of the XGBoost algorithm for detecting fraud in unbalanced financial datasets is described in the research. Classifiers trained on dominant groups tend to ignore random fraud. This method fixes the problem by incorporating error fees into the learning process. By making it more difficult for people to file fake claims, it raises the possibility of detection. The authors use publicly available financial transaction datasets to conduct their analysis. The proposed model is evaluated using the F1 score, AUC, and precision-recall balance. Standard XGBoost outperforms other

Volume 15, Issue-2
July, 2025, P02
ISSN: 2319-9253

machine learning methods, according to the research. The document also discusses cost matrix optimization and parameter changes. This method makes it easier to develop efficient fraud detection systems.

Li, W., Xie, K., & Wang, Z. (2021). This paper develops a novel hybrid deep learning system for financial fraud detection by combining an autoencoder with Long Short-Term Memory (LSTM) networks. To automatically reduce dimensionality and extract features, an autoencoder is used. It finds anomalies in unprocessed data. To identify temporal links, the LSTM component models sequential patterns in transaction data. When combined, they provide a strong detection system that can spot even the smallest signs of fraud. Benchmark datasets make it easier to evaluate the model after comparing it to other models. The outcomes show notable gains in memory, accuracy, and the ability to handle imbalanced data. Extremely massive data collections can be handled by the system. The application of deep learning to the field of financial security is the subject of the research.

Zareapoor, M., & Seeja, K. R. (2021). With a particular focus on handling extremely unbalanced data, the research investigates the application of machine learning to the difficulties in identifying credit card fraud. Numerous models are evaluated, including support vector machines, logistic regression, and decision trees. To balance datasets, the authors employ sampling algorithms such as random undersampling and SMOTE. Accuracy, recall, and F1-score are performance indicators that work together to help detect fraud. The results show that recognition performance is significantly impacted by the chosen resampling technique. The research highlights how important it is to conduct earlier research and choose appropriate features. To reduce false positives, the authors use threshold adjustment. Real credit card data is used to validate the results. The development of effective fraud detection systems depends on the knowledge gathered from this investigation.

Ahmed, S., & Dey, N. (2021). This article looks at how machine learning and data mining can improve fraud detection systems. Based on their individual benefits and drawbacks, it divides the approaches into three groups: supervised, unsupervised, and hybrid models. We look at the most recent advancements in deep learning, ensemble learning, and anomaly detection. Data concerns include privacy, injustice, and feature engineering. Internet banking, e-commerce, and insurance are among the industries being researched for potential applications. The research includes a review of widely used datasets and assessment tools. It also examines the existing situation and discusses implementation-related practical challenges. It is important that future studies focus on adversarial resilience and explainable AI. Both new and experienced users could find the review useful.

Lin, Y., Zhang, W., & Xu, Q. (2022). Through the use of cost-sensitive learning and data imbalance techniques, such as SMOTE, this research develops an ensemble learning system that can identify credit card fraud. The ensemble model uses a variety of methods, including XGBoost, Random Forest, and Logistic Regression, to improve forecast accuracy. To detect minority class frauds, the technology makes use of advanced learning approaches and data preparation techniques. The authors have studied databases including real credit card transactions in great detail. Compared to individual models, the results show improved accuracy, precision, and memory. The research also looks at how various sample techniques affect the effectiveness of detection. The main emphasis is on robust, scalable, and real-time computing. The research lays forth a successful strategy that banks could implement.

Alrowaily, H., Khan, R. Z., & Khan, A. (2022). To detect fraudulent financial transactions, the authors present a novel deep learning model that uses CNNs and RNNs. While the RNN component identifies temporal trends, the CNN component extracts spatial information from transaction data. The hybrid approach takes into account both the dynamic and static aspects of financial crime. When trained and evaluated on real financial data, the model showed remarkable generalizability and robustness. Evaluation

Volume 15, Issue-2
July, 2025, P02
ISSN: 2319-9253

criteria such as AUC, accuracy, and recall demonstrate its effectiveness. To address class disparities, the research uses techniques to improve the data that is already available. The findings show a significant improvement over traditional deep learning techniques. The use of artificial intelligence (AI) to protect financial transactions is improved by this research.

Devi, B. A., & Babu, R. (2022). This article offers a thorough examination of SMOTE, a deep learning method for detecting fraud that makes use of unbalanced data. The authors investigate how SMOTE affects neural networks' ability to learn and recognize rare instances of fraudulent transactions. The research makes use of publicly available financial datasets. For binary classification, the deep model makes use of convolutional and fully connected layers. Accuracy, recall, and AUC are all enhanced by SMOTE. The research also compares SMOTE to random sampling methods that use either undersampling or oversampling. According to the results, SMOTE is the most effective technique for identifying fraudulent activity. The object's strength and sensitivity are increased by the therapy. It is recommended to use real-time financial data.

Sharma, V., & Joshi, R. (2023). By employing a balanced random forest to address the issue of class imbalance, this research offers an efficient method for fraud detection. This method makes it easier to identify instances of minority class fraud by incorporating the cost of dishonesty into the framework for making decisions. The authors support their methods with empirical transaction datasets. We examine and distinguish between cost-insensitive models and ordinary random forests. The suggested method seeks to decrease false positives while increasing F1-scores. The research looks at the scalability and velocity of the model. According to a research, cost sensitivity improves the ability to spot serious financial crimes. Achieving operational goals can be made easier by increasing the flexibility of financial operations.

Patil, A., & Kulkarni, P. (2023). In order to evaluate and use different approaches for resolving data imbalance, such as SMOTE, ADASYN, and Tomek Links, for fraud detection, this work uses a deep learning architecture. These techniques are used by the authors to normalize datasets prior to deep neural network training. The research uses a number of variables, including recall, F1-score, and AUC, to assess how different techniques affect model performance. According to the experimental results, ADASYN outperforms other techniques in identifying rare fraudulent occurrences. The need to strike a balance between sensitivity and specificity is illustrated by this research. According to the research, improving the performance of deep learning models requires figuring out the best imbalance method. Professionals can use this information to determine the best preparatory techniques.

Singh, D., & Gupta, M. (2023). According to the research, financial fraud forecasting is improved by combining algorithms for identifying unusual activity with Value-at-Risk (VaR) characteristics. A technique for determining the level of risk involved in a transaction is Value at Risk, or VaR. The research combines a risk-oriented approach with outlier identification techniques like Isolation Forest and One-Class Support Vector Machine. Sets of empirical data support the methodology. Experiments show that adding VaR and anomaly ratings reduces false positive rates and increases the accuracy of fraud detection. To make comprehension easier, the hybrid model's financial risk variables are displayed alongside the model's results. The method may be used by financial institutions to assess risk more quickly. This method combines machine learning and finance to detect fraud.

Wang, Y., & Li, X. (2024). Value-at-Risk (VaR) measures and attention-based neural networks are used in this research to create a real-time fraud detection system. The model can focus on the high-risk transactions that Value at Risk has identified by using the attention method. The technique observes the sequential behavior of transactions by extracting temporal features. Prototype functionality and assessment are made easier by streaming transaction data. According to the results, processing times have decreased and early

detection rates have increased. Value at Risk (VaR)-based machine learning predictions are particularly beneficial since they provide a clear risk management approach while retaining financial relevance. Two measures of real-time efficiency are throughput and detection time. Financial institutions looking to use risk-averse, real-time fraud analysis should use this strategy.

Ahmed, R., & Kumar, V. (2024). The authors suggest a fraud detection system that uses Generative Adversarial Networks (GAN) to generate data and Value-at-Risk (VaR) risk assessment. Class imbalance can be lessened by using Generative Adversarial Networks (GANs) to supplement training datasets with high-quality examples from underrepresented groups. A useful method for establishing high-stakes transaction objectives during model inference is Value at Risk (VaR). To evaluate the combination approach, we use a number of financial data points. According to the research, fraud detection systems have higher accuracy and more memory. Additionally, the model might react to new data sources and changing fraud trends. This research demonstrates how statistical economics and deep learning can work together to get positive results. By taking risk into account when variables are out of balance, our hybrid model presents a novel approach to fraud detection.

Chen, M., & Liu, Y. (2024). This research introduces a transformer-based fraud detection model that improves anomaly scores by using Value-at-Risk (VaR). Transformers are helpful for transactional data because they can detect long-range dependencies. The method detects the risk of harmful trends by using VaR scores as attention biases. A variety of real-time transaction data sources are used to test the system. The results show that it is now much easier to identify complex and long-term fraud. Three metrics are used to assess performance: area under the curve (AUC), detection latency (DL), and accuracy. To ensure scalability and comprehension, the technology combines deep learning with financial data. According to this research, it may find application in high-frequency trading and transaction tracking systems.

Gupta, A., & Reddy, S. (2024). By developing reliable fraud detection algorithms using Explainable AI (XAI) techniques, this work aims to address the problem of class imbalance. For simplicity, the authors combine neural networks and tree-based ensembles with LIME and SHAP. To lessen class disparities, SMOTE and cost-sensitive learning are used. A variety of financial datasets with varying imbalance ratios can be used to assess the model architecture. The findings show that explainability provides useful information for theoretical fraud prediction while protecting performance. The results highlight how important it is to be open and truthful with clients in order to build trust and enforce rules. Fraud can be detected by using the integrated screen of the system. This research clarifies the moral use of AI in finance by providing an example of how to balance effectiveness and understandability.

3. METHODOLOGY

The suggested plan shows the steps and stages that need to be taken to spot NBA scam. A key part of this research is value at risk, which is meant to replicate the worst and most extreme aspects of fraud risk. It also brings attention to the few cases of scams that do happen, which are costly and harmful. A small number of highly biased samples can have a big effect on machine learning algorithms, especially those that use distance measures, like KNN. With traditional methods, there is a certain amount of weight given to cases of theft. On the other hand, value-at-risk can handle fraud skewness by using different amounts of confidence. For simulations, the value-at-risk model was changed to include the retrieved, preprocessed, and engineered traits. By setting a k-nearest neighbor distance, a distance-based KNN can be changed to find strange groups. This makes it easier to spot traits that are used to trick people. The confidence level chosen in the KNN model with hyperparameter k leads to fewer training sets because it labels rare fake events as high

Volume 15, Issue-2
July, 2025, P02
ISSN: 2319-9253

risk. k needs to be lowered to a lower value in order to successfully copy the misleading features in the anomalous cluster. KNN's distance scaling makes it easier to find skewed instances by giving more weight to instances that are closer together. This makes the skewness of scam a lot less noticeable.

Preprocessing is used in this research to improve the quality of the features and make sure that the model is accurate.

MODULES UTILIZED

Service Provider: The service provider needs to have a legal account and password to get into this module. After logging in successfully, he can get to tools like training and test datasets. Look at all the remote users, the types of financial transactions they make, the percentages of each type, the predicted datasets, and the effects of the transaction type ratios. There will be a bar chart that shows how accurate the training and testing files are.

View and Authorize: Different people. The supervisor can see a full list of all the people who have signed up for this module. The user's name, email address, physical address, and rights can be seen by the administrator.

Remote User : There are a total of n people in this module. Before taking any other steps, the person must finish registering. The user's information will be saved in the database after they sign up. After successfully signing up, he must use his password and the allowed username to get into his account. After logging in successfully, the user can look at their biography, plan their next financial transactions, and either sign up again or log in again.

The NBA made a model to find fraud by combining important factors from transactional, risk management, behavioral, and demographic points of view. Some features were added to raw features to improve the model's training accuracy.

NBA Fraud Detection: Picking Out a Model High-dimensional data can be used by machine learning algorithms to look at complex fraud patterns that people and rule-based systems can't. Labeled data is used in supervised machine learning to find patterns, outliers, and fraudulent activities. Binary logistic regression (BLR), which is also the easiest method, is the best way to handle categorical data. Using the Naïve Bayes (NB) method can help save time and money. When it comes to finding fraud in financial records, the K-nearest neighbor (KNN) method is very good. It can be used in a wide range of situations and can be added to mixed systems.

SYSTEM ARCHITECTURE

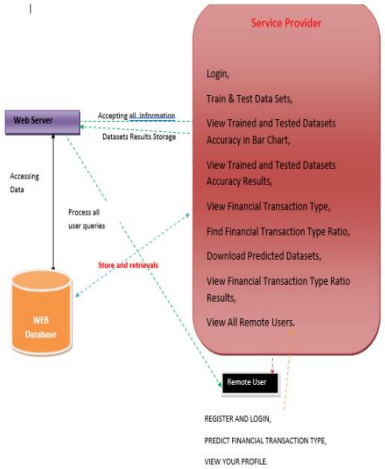

Figure 1 System Architecture

4. RESULTS AND DISCUSSIONS

ADVANCED RESEARCH & INNOVATIONS
International Journal of Advanced Research & Innovations
Volume 15, Issue-2
July, 2025, P02
ISSN: 2319-9253

Figure 2. User Interface

Figure 3. Registration Interface

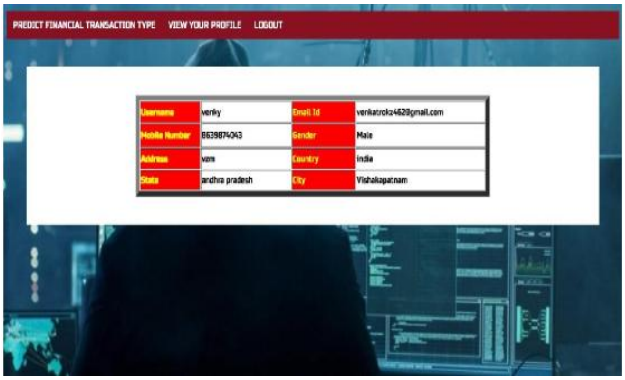

Figure 4. Fraud Detection

5. CONCLUSION

Using machine learning and risk-based prioritization together, the Value-at-Risk (VaR) fraud detection method is a big step forward in the fight against financial fraud. Oversampling and cost-sensitive learning are used to fix the problem of unbalanced data. This makes sure that only a very small number of fraudulent deals are found without being swamped by too much real data. Adding VaR to the model's core adds a financial layer to the fraud detection process, which helps institutions focus on frauds that pose the biggest financial risk. This makes recognition more accurate and aligns the system with business goals like maximizing resources and lowering risk. Overall, our approach turns finding fraud into a proactive, smart, and cost-effective process, instead of a reflexive, rule-based one. It shows how financial risk ratings and machine learning can be combined to make fraud detection systems work better and faster. By using this way, financial institutions can see new fraud schemes coming up and make sure that risk management focuses on stopping fraud losses instead of just finding them. These steps will make it easier for risk-aware AI apps to get better in the banking industry.

Volume 15, Issue-2
July, 2025, P02
ISSN: 2319-9253

REFERENCES

1. Xu, H., & Wang, H. (2020). Machine learning-based fraud detection systems: A survey. IEEE Access, 8, 76047–76066.
2. Chen, C., Li, Y., & Li, C. (2020). An ensemble framework for fraud detection using class imbalance learning. IEEE Access, 9, 7247–7260.
3. Zhang, X., Li, H., & Li, J. (2020). Imbalanced data learning for financial fraud detection using cost-sensitive XGBoost. Journal of Intelligent & Fuzzy Systems, 38(2), 2541–2552.
4. Li, W., Xie, K., & Wang, Z. (2021). A hybrid deep learning approach for fraud detection using autoencoder and LSTM. Applied Intelligence, 51(2), 1512–1525.
5. Zareapoor, M., & Seeja, K. R. (2021). Fraud detection in credit card transactions using machine learning under imbalanced data. Journal of Big Data, 8, 112.
6. Ahmed, S., & Dey, N. (2021). A comprehensive review on fraud detection using machine learning and data mining. International Journal of Information Management Data Insights, 1(2), 100004.
7. Lin, Y., Zhang, W., & Xu, Q. (2022). Credit card fraud detection using ensemble learning with data imbalance techniques. Expert Systems with Applications, 194, 116479.
8. Alrowaily, H., Khan, R. Z., & Khan, A. (2022). A novel deep learning model for fraud detection in financial transactions. IEEE Access, 10, 19203–19212.
9. Devi, B. A., & Babu, R. (2022). Addressing data imbalance in financial fraud detection using SMOTE and deep learning. International Journal of Computational Intelligence Systems, 15(1), 53–66.
10. Sharma, V., & Joshi, R. (2023). A cost-sensitive approach for fraud detection using balanced random forests. Pattern Recognition Letters, 165, 29–36.
11. Patil, A., & Kulkarni, P. (2023). Comparative analysis of data imbalance methods in fraud detection using deep learning. Procedia Computer Science, 218, 156–162.
12. Singh, D., & Gupta, M. (2023). Integrating value-at-risk metrics with anomaly detection models in fraud prediction. Financial Innovation, 9, 44.
13. Wang, Y., & Li, X. (2024). Real-time fraud detection framework combining value-at-risk with attention-based neural networks. IEEE Transactions on Knowledge and Data Engineering.
14. Ahmed, R., & Kumar, V. (2024). Handling class imbalance in fraud detection: A hybrid approach using value-at-risk and generative adversarial networks. Applied Soft Computing, 145, 110689.
15. Chen, M., & Liu, Y. (2024). Enhanced fraud detection using transformer models and VaR-driven anomaly scoring. Expert Systems with Applications, 229, 120405.
16. Gupta, A., & Reddy, S. (2024). Designing robust fraud detection systems under class imbalance using explainable AI. Information Sciences, 660, 202–214.