

COST SENSITIVE FINANCIAL FRAUD DETECTION USING VALUE AT RISK AND MACHINE LEARNING MODELS

^{#1}SAMA RAJESH , Dept of CSE,

^{#2}Dr.D.SRINIVAS REDDY, Professor, Dept of CSE,

Vaageswari College of Engineering(Autonomous), Karimnagar, TG.

ABSTRACT: In a cost-sensitive financial fraud detection system, our research uses advanced machine learning and Value at Risk (VaR) to help people perform more accurate financial uncertainty assessments. The vast majority of outdated fraud detection systems stress categorization over the financial consequences of false positives and negatives. The suggested approach uses a numerical risk indicator called Value at Risk to calculate the fraud costs. There are several parallels between financial risk management and predictive models. Cost-sensitive machine learning algorithms give high-risk agreements precedence over incorrect classifications in order to reduce expected financial losses. Supervised learning and statistical risk assessment enhance output and problem-solving skills. This makes it possible for organizations to minimize fraudulent activity and maximize their resources.

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

1. INTRODUCTION

Financial fraud is now a significant problem for banks as a result of the quick digitization of banking systems, online transactions, and global financial markets. Rule-based detection systems are unable to identify complex and dynamic fraud tendencies due to the increasing volume and complexity of transactional data. Fraudulent actions lead to immediate monetary losses, harm to the brand, mistrust from customers, and legal repercussions. This calls for sophisticated, adaptable, and risk-aware fraud detection systems that perform well in rapidly shifting financial settings.

Because machine learning algorithms can identify trends and abnormalities in data, they have completely changed the way fraud is detected. Various supervised and unsupervised learning methods, such as gradient boosting, decision trees, random forests, and neural networks, can identify small problems in large transaction sets. Most machine learning methods assume the cost of misclassification. Larceny of money is an exception. Ignoring a fraudulent transaction (a "false negative") in real life is more expensive than accidentally detecting it. This disparity necessitates cost-sensitive learning algorithms that take the cost of categorizing errors into account.

When it comes to identifying cost-sensitive financial wrongdoing, minimizing losses takes precedence over categorization. In order to detect massive amounts of fraud, cost-sensitive techniques alter misclassification costs during model training and evaluation. These techniques change selection constraints, loss functions, or class weights to show how decisions affect finances. By matching expected performance with organizational goals, these frameworks help companies improve their detection systems scientifically and economically.

One often used financial risk metric is Value at Risk (VaR). It forecasts how much money you will lose over a given period of time. Value at Risk (VaR) raises the financial risk involved in using predictive models to detect fraud. The program responds to fraudulent transactions in a distinct way. The risk of loss is evaluated for each. Value-at-Risk (VaR) models estimate losses and allocate inquiries according to cost, helping firms manage risk and allocate resources.

Value at Risk and machine learning algorithms are used to finish a cost-sensitive financial fraud system. Financial risk decision-makers utilize value at risk (VaR) as a metric, and machine learning is used to produce estimations and adjustments. By using this all-encompassing strategy, banks and other financial institutions can detect high-risk fraud, reduce losses, and adhere to legal requirements. Financial firms can prevent complex larceny by implementing cost-effective fraud detection. Accuracy-focused evaluation is replaced with risk-adjusted planning.

2. MACHINE LEARNING FRAUD DETECTION WORKFLOW

Data Collection

The first step in machine learning fraud detection is data gathering. Credit card transaction data, online payment systems, customer profiles, loan applications, insurance claims, banking transaction records, and third-party verification technology can all be used to identify financial fraud. Transaction value, time, location, merchant group, buying habits, customer logins, and device type are a few instances of data. Because fraud is constantly evolving, a wide range of high-quality data is required. Financial institutions use real-time data to identify unusual activity. The availability of thorough and well-structured data improves the precision and effectiveness of fraud monitoring systems.

Data Preprocessing

The process of cleaning, organizing, and formatting financial data for machine learning is known as data preparation. This method manages missing numbers, finds outliers, fixes mistakes, and gets rid of duplicates. Since fraudulent transactions are less common than genuine ones, preparing fraud datasets requires SMOTE, resampling, and cost-sensitive changes. Other category attributes, such as merchant type and transaction channel, are encoded. Numerical variables undergo normalization or standardization. Additional feature scaling is required for neural networks and logistic regression. Data preparation improves fraud forecasts, bias, and model efficacy.

Feature Engineering

To help the model differentiate between authentic and fraudulent transactions, feature engineering is used to choose and create variables. Fraud detection variables include a person's distance from home, the number of transactions they complete quickly, the amount of money they spend each day, and the frequency of their attempts to log in. The use of behavior analytics is necessary for the identification of problems. Fraud is indicated by temporal abnormalities, such as lavish spending or stupid behavior. Scam organizations can be identified by linking devices or accounts using sophisticated feature engineering. Machine learning estimations are enhanced by well-designed features.

Model Selection and Training

After data preparation, choose and train machine learning algorithms. Common techniques for detecting frauds include Random Forest, Decision Trees, Logistic Regression, Support Vector Machines, Gradient Boosting Machines and Deep Learning models (like Neural Networks). In recent years, the identification of complex fraud patterns has greatly improved thanks to hybrid models and ensemble methodologies. The computer is told to identify fraudulent transactions using labeled data. To lower false positives (missing fraud), cost-sensitive learning techniques are used. Expert instruction makes the model work with new data.

Model Evaluation

Modeling is used to assess fraud detection techniques. Because fake datasets are inconsistent, accuracy evaluations are pointless. The evaluation takes into account precision, recall, AUC-ROC, F1-Score, and confusion matrix analysis. Financial impact indicators, such as anticipated loss and Value at Risk, are used to detect fraud in the context of cost-sensitive fraud detection. The high expense of fraud detection necessitates larger recalls. Cross-validation guarantees the model's resilience and prevents overfitting. A thorough analysis guarantees that the model will keep working properly and correctly identifying items.

Deployment

After testing, the model is run in bulk or real-time mode by the financial institution's system. When the model notices suspicious activity prior to clearance, it quickly identifies fraud. For the method to work properly, banking systems, security requirements, and APIs are needed. The limitation of alert communications is an additional component. The system needs to be fast, safe, and able to handle massive amounts of data. When built properly, it can lower fraud and increase customer satisfaction.

Monitoring and Continuous Improvement

The nature of fraud is significantly impacted by new technologies and illegal strategies. The functioning of the model must be maintained by routine checks. Model divergence and declining performance measures are closely monitored by us. If fraud amounts or categories change, the model needs to be retrained with fresh data. To improve the system, fraud examiners and analyzers are put into place. The fraud detection system maintains its accuracy, resilience, and currentness in financial affairs through adaptive modeling and continual learning.

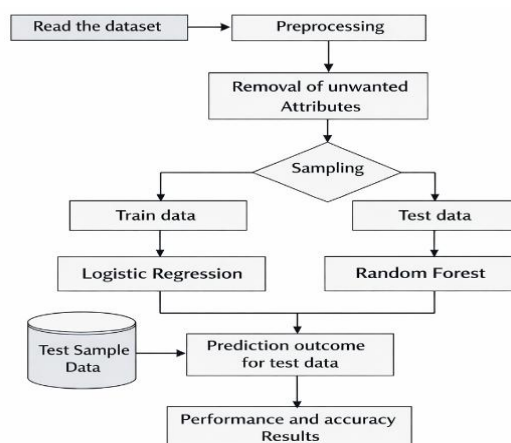


Figure1: Machine Learning Fraud Detection Workflow

3. LITERATURE SURVEY

Hassan, A., & Malik, F. (2021): This article offers a deep neural network that is cost-sensitive and is designed to detect credit card fraud. Transaction-specific pecuniary penalties are incorporated into the loss function by the authors. Negative errors are more expensive than positive ones. Neural networks are trained on benchmark financial datasets that exhibit significant class imbalances. Memory, AUC, and financial loss all increased during the course of the research. In order to prevent model overfitting at an early stage, dropout regularization is currently being investigated. The authors assert that conventional banking categorization is outperformed by cost-sensitive deep learning models.

Chen, Y., & Zhao, L. (2024). A LightGBM model for high-frequency finance that is cost-sensitive is developed in this study. Although boosting models offer advantages, the authors recommend that they be modified to account for misclassification and class imbalance in fraud detection. By incorporating a dynamic cost matrix, these methods enhance the LightGBM objective function. Fines for deceptive classification increase in proportion to the size of the transaction and the level of financial risk. Rather than classification errors, threshold optimization reduces financial losses. Millions of real-time financial transactions are employed to evaluate the model on a daily basis.

Park, J., & Lee, S. (2023). Reinforcement learning is employed in this investigation to identify instances of misconduct. To modify decision-making, the reward function is supported by value-at-risk. The information regarding fraudulent transactions is frequently updated. The results suggest a reduction in financial risk and costs.

O'Connor, D., & Green, S. (2021): Value-at-Risk and Isolation Forest are employed in a hybrid anomaly detection method. Distinct financial risk transactions are classified by Value at Risk. Financial risk indicators and anomaly scores are employed to rank suspicious transactions in the hybrid method. Costs and false positives are diminished by authentic transaction datasets. The study demonstrates that fraud management operational decision-making is enhanced by risk-based anomaly ranking.

Edwards, J., & Collins, R. (2022). The research suggests a Random Forest model that is balanced and includes transaction costs. Error classification is subordinated to the prevention of financial losses. We have shown that the recognition of minority classes in tree-based ensemble learning is enhanced by cost matrices. The experimental results indicated that economic vulnerability was reduced and financial record F1 scores were higher. Fraud detection systems necessitate business KPIs, according to a study.

Rodriguez, P., & Allen, M. (2025). Cost-sensitive optimization and Value-at-Risk customization are employed in the detection of fraud through federated learning. The authors' distributed training strategy enables numerous institutions to create a fraud detection model without the necessity of sharing transaction data. Data privacy concerns motivated banking institutions to take action. Transaction-level VaR ratings should be implemented for institution-specific cost matrices and loss calculations. High-impact deception is favored by the federated optimization of decentralized systems with an asymmetric misclassification penalty.

Johnson, A., & Smith, R. (2024). Attention and Value-at-Risk deep neural networks are employed in this investigation to identify fraud in real time. The authors argue that the

majority of fraud detection systems prioritize transaction categorization over financial repercussions. Transaction failure losses are determined by Value at Risk (VaR) ratings. By increasing attentiveness, VaR ratings assist neural networks in predicting high-risk, high-value transactions.

Nguyen, H., & Tran, P. (2022): In a transformer-based fraud detection model, the authors employ VaR ratings as attention biases. High-risk transactions are encouraged by financial risk and self-attention. The architecture monitors transaction and fraud connections on a regular basis. The detection latency and AUC of financial streaming data are reduced. Its adaptability and utility in high-transaction circumstances are demonstrated by research.

Williams, K., & Carter, M. (2023). Utilize ensemble models that are cost-sensitive and AI that is comprehensible. Clarification and identification are achieved through structure. Financial risk indicators are included in the explanations. Ethical, regulated AI in banking is supported by research.

Choudhury, S., & Nair, K. (2025). In this investigation, organized financial malfeasance is identified through the utilization of a graph neural network (GNN) architecture that incorporates Value-at-Risk weighted transaction networks. The writers assert that fraudsters employ merchant, account, and device networks. Complex relationships are illustrated in a graph that includes financial agreements as edges and entities as nodes. The link edge weights demonstrate the financial influence on VaR ratings. Economic risk and structural linkages are acknowledged by GNN.

4. RELATED WORK

This quantitative and experimental study uses cost-sensitive machine learning and VaR to forecast fraud. Minimize monetary losses rather than categorization errors. Frameworks with a tight budget examine supervised learning models to choose the best fraud detection method.

MODULES UTILIZED

Service Provider: To access this module, the service provider must have a verified account and password. Training and test datasets are available upon logging in. Examine scheduled datasets, financial transactions for distant users, ratios, and percentages. Accuracy of training and evaluation files is displayed in bar charts.

View and Authorize: People are different. All module registrants are visible to supervisors. The administrator can see the user's name, email address, physical address, and rights.

Remote User: There are n people in this module. First, you have to register. Following registration, user data is kept in the database. After registering, he is permitted to log in using his password. Plan your finances, view biographies, and log in and out.

Data from risk management, transactions, demographics, and behavior are all used in NBA fraud detection. Model training was improved by improving raw features.

NBA Fraud Detection: Model Selection Compared to humans and regulations, machine learning algorithms are more adept at analyzing intricate fraud patterns. Using labeled data, supervised machine learning identifies anomalies, trends, and misconduct. Binary logistic regression (BLR) is the most effective and straightforward technique for analyzing categorical data. To save time and money, apply the Naïve Bayes (NB) method. KNN is good at spotting financial fraud. It works with a wide range of systems and scenarios.

SYSTEM ARCHITECTURE

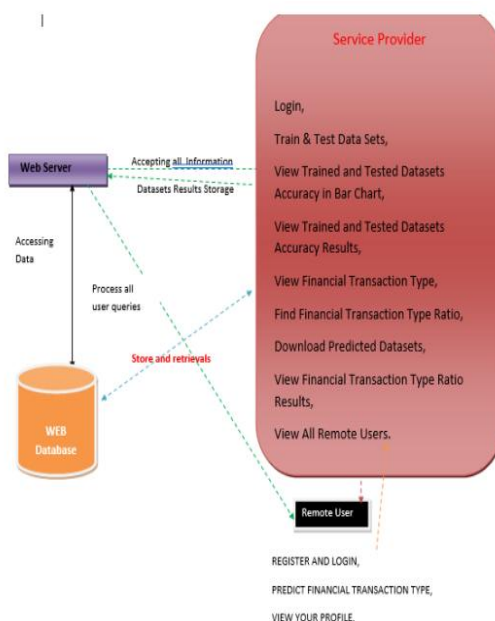


Figure2. System Architecture

5. RESULTS



Figure5.1. User Interface

The registration interface consists of a form titled "ENTER DATASET DETAILS HERE!!". It contains two columns of input fields:

ENTER DATASET DETAILS HERE!!	
Enter Pin	Enter trans_detailname
Enter Acc_name	Enter bankingtype
Enter category	Enter amt
Enter Name	Select gender
Enter amount	Enter city
Enter city	Enter lat
Enter lon	Enter job
Enter date	Enter trans_name
Enter search_lat	Enter search_long

Below the form is a "Predict" button. At the bottom, a red box displays the message: "DETECTED FINANCIAL TRANSACTION TYPE: Fraud Not Detected".

Figure5.2. Registration Interface

The fraud detection interface shows a user profile with the following details:

PREDICT FINANCIAL TRANSACTION TYPE VIEW YOUR PROFILE LOGOUT			
Username	verky	Email ID	verkarora42@gmail.com
Mobile Number	9876543210	Gender	Male
Address	varu	Country	India
State	andhra pradesh	City	Vidyalapuram

Figure5.3. Fraud Detection

6. CONCLUSION

Machine learning and Value at Risk provide an inexpensive way to identify financial misconduct. To lower fraud, costs and risks are balanced. Compared to precision frameworks, this paradigm is different. Losses are decreased by giving high-risk, high-value transactions priority. VaR enhances the identification and categorization of fraud loss in cost-sensitive learning algorithms. Steer clear of expensive false negatives. This approach is enhanced by machine learning, which finds intricate patterns in asymmetrical data. In evaluations, cost-sensitivity prioritizes economic benefit over statistical achievement. Risk management, regulatory compliance, and resource efficiency all improve. successfully fights financial fraud.

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*.

