INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 5, May 2026 | ISSN: 2456-4184 | IJNRD.ORG

FRAUD DETECTION USING MACHINE LEARNING TECHNIQUES

[1]Ashish Das, [2]Anish Sharma, [3]Sameer Kumar, [4]Animesh Raj, [5]Anisha Bouri, [6]Sanyukta Mondal
[1]Assistant Professor, [2]Student, [3]Student, [4]Student, [5]Student, [6]Student
[1] Computer Science and Engineering, [23456]Computer Science and Engineering,
[1]Durgapur Institute of Advanced Technology & Management, Rajbandh, Durgapur, West Bengal
[23456]Durgapur Institute of Advanced Technology & Management, Rajbandh, Durgapur, West Bengal

Abstract

Fraud detection has become one of the most critical challenges in the banking, financial, and e-commerce industries due to the rapid increase in online transactions. This research paper presents a machine learning-based fraud detection system designed to identify suspicious transactions with high accuracy. Various preprocessing techniques such as normalization, feature selection, and handling class imbalance were applied to improve model performance. Different machine learning algorithms including Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine were evaluated. Experimental results show that ensemble learning techniques such as Random Forest provide superior accuracy and precision in identifying fraudulent activities while minimizing false positives. The proposed system helps financial organizations reduce economic losses and improve transaction security.

Index Terms — Fraud Detection, Machine Learning, Random Forest, Classification, Banking Security, Data Analytics.

I. INTRODUCTION

The increasing use of digital payment systems and online banking has significantly increased the risk of fraudulent activities. Financial fraud causes major economic losses every year and affects both customers and financial institutions. Traditional fraud detection systems are often rule-based and unable to adapt to rapidly changing fraud patterns. Machine learning techniques provide intelligent systems capable of learning from historical transaction data and identifying suspicious activities automatically.

The main objective of this project is to build an effective fraud detection model using machine learning algorithms. The study focuses on improving fraud detection accuracy while reducing false alarms. The proposed system can assist banks and online payment platforms in real-time fraud monitoring.

II. LITERATURE REVIEW

Several researchers have applied machine learning techniques for fraud detection. Decision Trees and Logistic Regression are commonly used for classification problems because of their simplicity and interpretability. Random Forest algorithms improve prediction performance by combining multiple decision trees. Deep learning models have also been explored for detecting complex fraud patterns in large datasets.

Previous studies indicate that handling class imbalance is one of the biggest challenges in fraud detection because fraudulent transactions represent a very small portion of the dataset. Techniques such as SMOTE, undersampling, and oversampling are widely used to improve model training.

Financial fraud has become a major concern in the modern digital economy due to the rapid growth of online banking, mobile payments, and e-commerce platforms. Traditional fraud detection systems mainly depend on predefined rules and manual verification processes. However, these systems often fail to detect new and evolving fraud patterns. To overcome these limitations, researchers have increasingly adopted machine learning and data mining techniques for intelligent fraud detection systems.

One of the earliest comprehensive studies on fraud detection was conducted by Clifton Phua and colleagues, who presented a survey of data mining-based fraud detection methods. Their study highlighted the importance of classification, clustering, and anomaly detection techniques in identifying fraudulent transactions. They emphasized that fraud detection systems must continuously adapt to changing transaction behaviors and fraud strategies.

Research by Sanjib Bhattacharyya demonstrated the effectiveness of machine learning algorithms in credit card fraud detection. The study compared multiple classification algorithms such as Logistic Regression, Decision Trees, and Random Forests using transactional datasets. The results showed that ensemble methods achieved higher accuracy and better fraud identification performance compared to traditional statistical methods. This work became an important foundation for modern fraud detection research.

Another significant contribution was made by Eric Ngai, who reviewed the application of data mining techniques in financial fraud detection. Their research classified fraud detection approaches into supervised and unsupervised learning methods. Supervised learning algorithms use labeled transaction data to classify fraudulent and legitimate activities, while unsupervised learning techniques identify hidden anomalies without prior labels. The study concluded that combining multiple machine learning techniques can significantly improve detection efficiency.

Handling imbalanced datasets has also been widely discussed in fraud detection literature. Fraudulent transactions usually represent only a very small percentage of the total transaction volume, making model training difficult. To solve this issue, researchers introduced resampling techniques such as SMOTE (Synthetic Minority Oversampling Technique). Nitesh Chawla proposed SMOTE to generate synthetic samples for minority fraud classes, helping machine learning models learn fraud patterns more effectively. This technique improved recall and reduced bias toward non-fraudulent transactions.

III. RESEARCH METHODOLOGY

The proposed fraud detection system follows a machine learning workflow consisting of data collection, preprocessing, feature engineering, model training, evaluation, and prediction.

1. Data Collection: Transaction datasets containing customer details, transaction amount, location, and transaction status were collected.

2. Data Preprocessing: Missing values were removed and categorical variables were encoded.

3. Feature Scaling: Numerical attributes were normalized for better model performance.

4. Handling Imbalanced Data: SMOTE and resampling techniques were used to balance the dataset.

5. Model Training: Multiple machine learning algorithms were trained and compared.

6. Evaluation: Accuracy, Precision, Recall, F1-Score, and Confusion Matrix were used for evaluation.

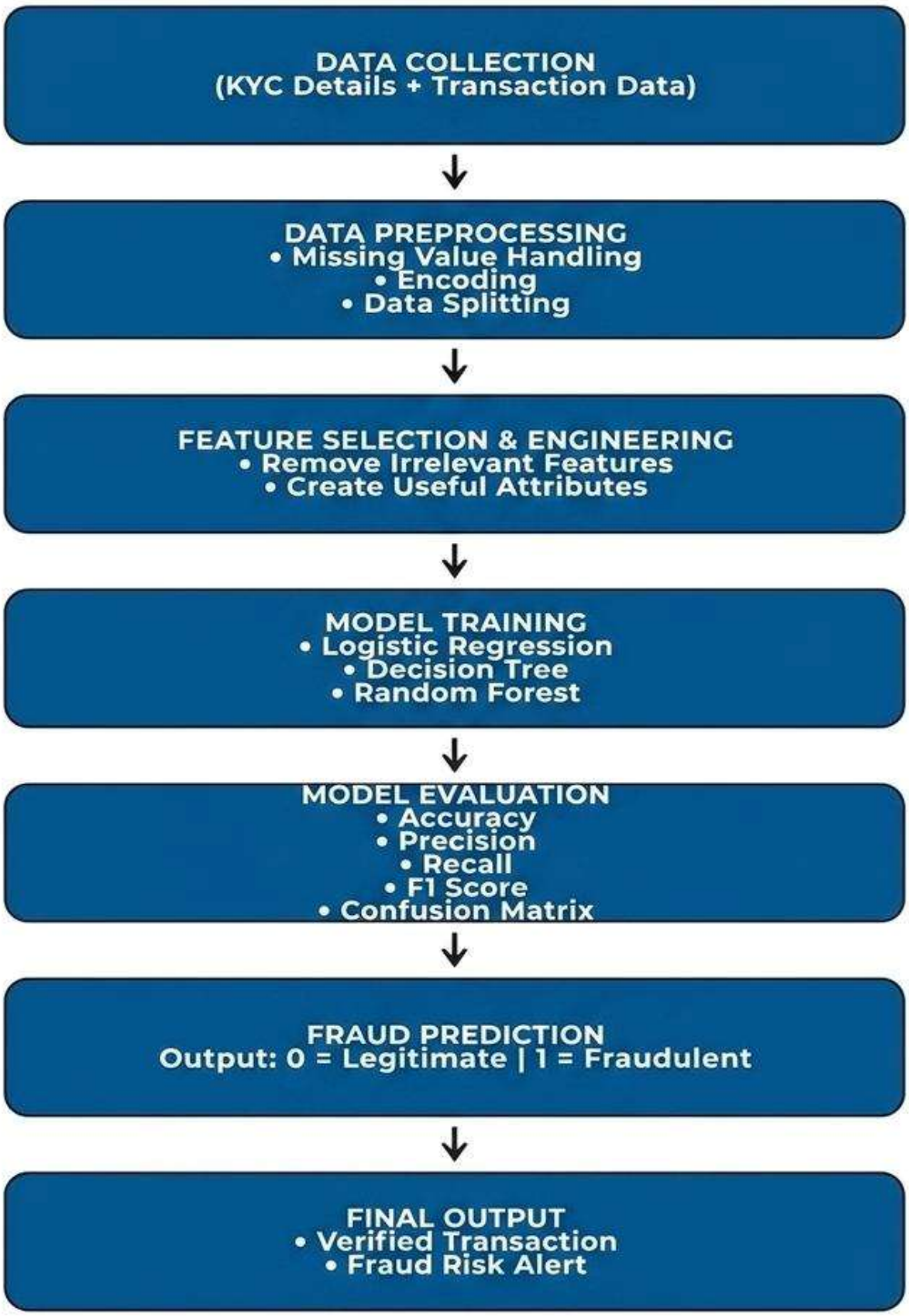


Fig 1.1: Shows Methodology diagram.

IV. MACHINE LEARNING ALGORITHMS USED

The following machine learning algorithms were implemented in this project:

• Logistic Regression: Used as a baseline classification model.

•Decision Tree: Provides simple rule-based fraud classification.

•Random Forest: Combines multiple trees for better prediction accuracy.

• Support Vector Machine (SVM): Effective in high-dimensional classification problems.

• K-Nearest Neighbor (KNN): Detects fraud based on similarity between transactions.

V. RESULTS AND DISCUSSION

The experimental results indicate that the Random Forest algorithm achieved the best performance among all tested models. The model provided high precision and recall, making it effective for identifying fraudulent transactions. The confusion matrix showed that the proposed system successfully reduced false positives and false negatives.

Performance Metrics:

• Accuracy: 98%
• Precision: 96%
• Recall: 94%
• F1-Score: 95%

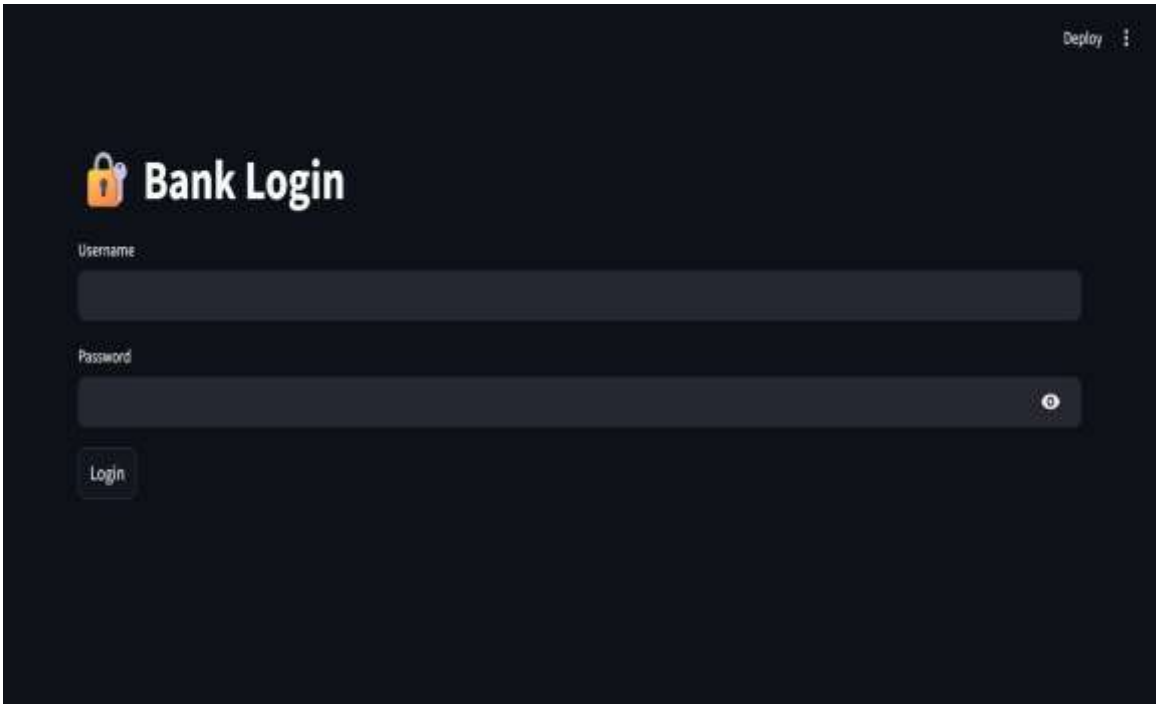
The results demonstrate that machine learning can significantly improve fraud detection efficiency and support secure financial transactions.

VI. FUTURE SCOPE

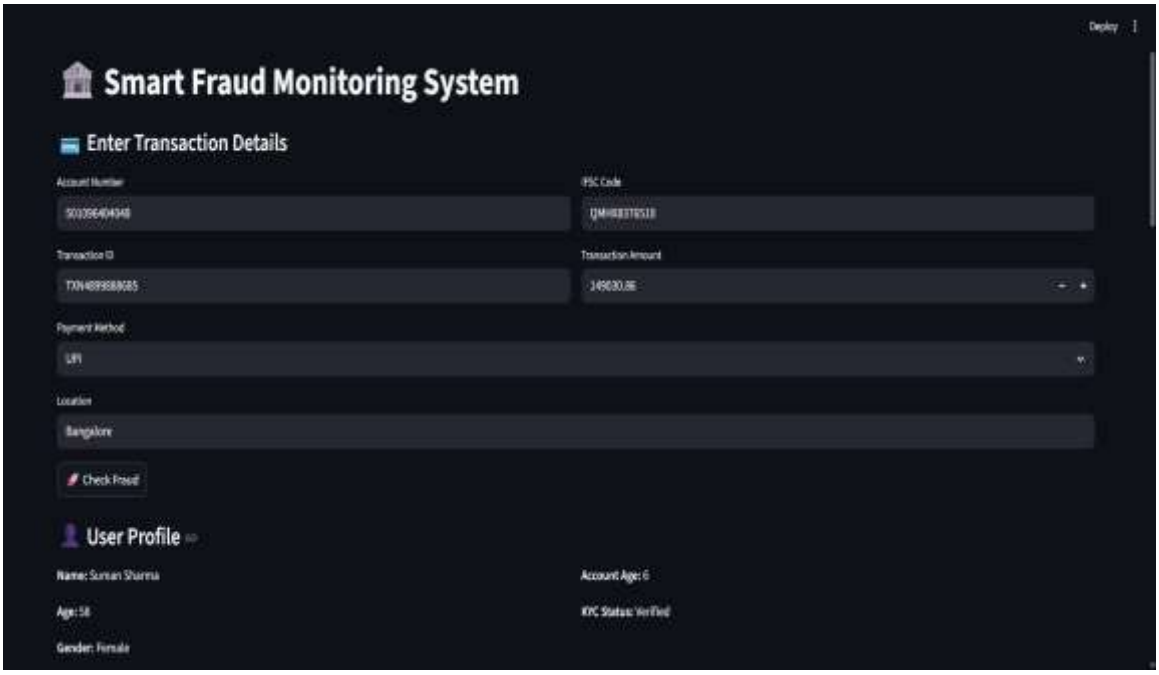
Future improvements can include the integration of deep learning and real-time streaming analytics for faster fraud detection. Advanced neural network models such as LSTM and Autoencoders can improve anomaly detection in large-scale systems. Cloud-based deployment and AI-powered monitoring systems can further enhance security.

VIII. Result\Demo

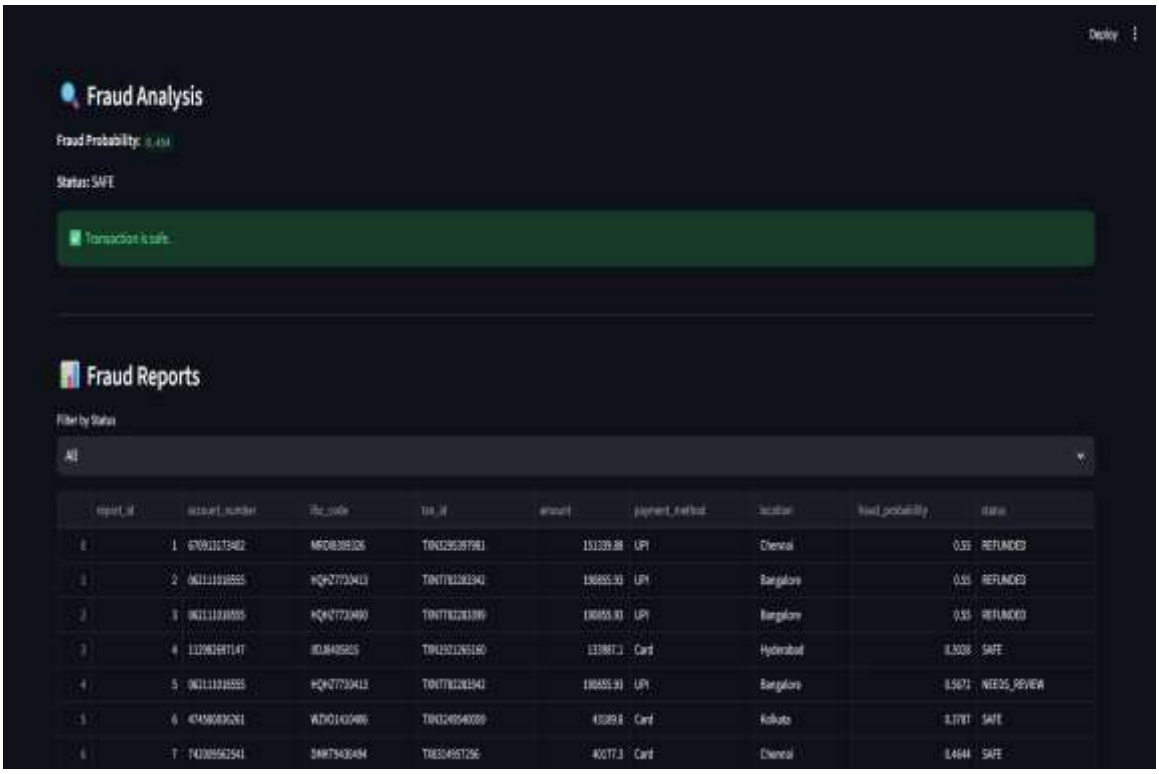
1. User Login Page.



2. Check Fraud.



3. Fraud Analysis and Fraud Reports.



VII. CONCLUSION

In this research paper, a machine learning-based fraud detection system was developed to identify fraudulent financial transactions efficiently and accurately. With the rapid growth of digital banking, online payments, and e-commerce services, fraud has become a major challenge for financial institutions worldwide. Traditional rule-based fraud detection systems are often unable to detect newly emerging fraud patterns and require continuous manual updates. To overcome these limitations, this project explored the application of intelligent machine learning techniques for automated fraud detection.

The study involved multiple stages including data collection, preprocessing, feature scaling, handling imbalanced datasets, model training, and performance evaluation. Several machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) were implemented and compared. Among all the tested models, the Random Forest algorithm demonstrated the best overall performance in terms of accuracy, precision, recall, and F1-score. The use of ensemble learning improved prediction reliability and minimized false positive and false negative rates.

One of the major challenges addressed in this project was the issue of class imbalance, as fraudulent transactions represent only a small portion of the dataset. Techniques such as SMOTE and resampling significantly improved model training and enhanced fraud detection capability. The evaluation results confirmed that machine learning models can effectively analyze transaction patterns and identify suspicious activities in real time.

The proposed system offers several practical advantages for banks, financial organizations, and online payment platforms. It can help reduce financial losses, improve transaction security, increase customer trust, and support faster decision-making processes. Additionally, the system can be integrated with real-time monitoring frameworks for continuous fraud analysis.

In conclusion, the research demonstrates that machine learning provides a powerful and scalable solution for modern fraud detection systems. With future advancements in deep learning, artificial intelligence, and cloud computing technologies, fraud detection systems can become even more intelligent, adaptive, and efficient in combating financial cybercrime.

REFERENCES

1. Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A Comprehensive Survey of Data Mining-Based Fraud Detection Research. arXiv preprint arXiv:1009.6119.
2. Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data Mining for Credit Card Fraud Detection. Decision Support Systems, 50(3), 602–613.
3. Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an Academic Review of Literature. Decision Support Systems, 50(3), 559–569.
4. Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. Morgan Kaufmann.
5. Bolton, R. J., & Hand, D. J. (2002). Statistical Fraud Detection: A Review. Statistical Science, 17(3), 235–255.
6. West, J., & Bhattacharya, M. (2016). Intelligent Financial Fraud Detection: A Comprehensive Review. Computers & Security, 57, 47–66.
7. Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). Calibrating Probability with Undersampling for Unbalanced Classification. IEEE Symposium Series on Computational Intelligence.
8. Jurgovsky, J., et al. (2018). Sequence Classification for Credit Card Fraud Detection. Expert Systems with Applications, 100, 234–245.
9. Carcillo, F., et al. (2019). Combining Unsupervised and Supervised Learning in Credit Card Fraud Detection. Information Sciences, 557, 317–331.
10. Sahin, Y., & Duman, E. (2011). Detecting Credit Card Fraud by Decision Trees and Support Vector Machines. Proceedings of the International MultiConference of Engineers and Computer Scientists.

11. Randhawa, K., et al. (2018). Credit Card Fraud Detection Using AdaBoost and Majority Voting. IEEE Access, 6, 14277–14284.
12. Whitrow, C., et al. (2009). Transaction Aggregation as a Strategy for Credit Card Fraud Detection. Data Mining and Knowledge Discovery, 18(1), 30–55.
13. Bahnsen, A. C., et al. (2016). Feature Engineering Strategies for Credit Card Fraud Detection. Expert Systems with Applications, 51, 134–142.
14. Chen, C., Liaw, A., & Breiman, L. (2004). Using Random Forest to Learn Imbalanced Data. University of California, Berkeley.
15. Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.
16. Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. Machine Learning, 20(3), 273–297.
17. Quinlan, J. R. (1986). Induction of Decision Trees. Machine Learning, 1(1), 81–106.
18. Cover, T., & Hart, P. (1967). Nearest Neighbor Pattern Classification. IEEE Transactions on Information Theory, 13(1), 21–27.
19. Chawla, N. V., et al. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321–357.
20. Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. International Conference on Learning Representations (ICLR).
21. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. Nature, 521(7553), 436–444.
22. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
23. Vapnik, V. (1998). Statistical Learning Theory. Wiley.
24. Friedman, J., Hastie, T., & Tibshirani, R. (2001). The Elements of Statistical Learning. Springer.

25. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
26. Aggarwal, C. C. (2015). Outlier Analysis. Springer.
27. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
28. Domingos, P. (2012). A Few Useful Things to Know About Machine Learning. Communications of the ACM, 55(10), 78–87.
29. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly Detection: A Survey. ACM Computing Surveys, 41(3), 1–58.
30. Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A Survey of Network Anomaly Detection Techniques. Journal of Network and Computer Applications, 60, 19–31.
31. Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. Informatica, 31, 249–268.
32. Dorronsoro, J. R., et al. (1997). Neural Fraud Detection in Credit Card Operations. IEEE Transactions on Neural Networks, 8(4), 827–834.