IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 13, Issue 12, July 2025)

Machine Learning Using Random Forest Model for Financial Data Prediction

Meghna Chandel
Assistant Professor
Computer Engineering Department
Shri G. S. Institute of Technology and Science, Indore (M.P.), India
Email ID: meghna_chandel@sgsits.ac.in
ORCID ID: 0009-0005-7972-112X

Abstract – **Data-driven predictive models are becoming popular across the financial institutions in assessing risks, predicting trends, identifying anomalies and tailoring customer services. Financial datasets are nonlinear and high-dimensional which frequently prove difficult to handle using traditional statistical models. The machine learning (ML) algorithms and especially the ensemble models such as the random forest (RF) modeling provide both resistant power and tolerance to noise. The paper is research on the effectiveness of the Random Forest model in financial data prediction with special focus on predicting trends in stocks, credit risk, and selecting anomalies in the data of transactions. It is compared to the models of the Logistic Regression, Support Vector Machines (SVM), and Gradient Boosting. The experimental settings prove that the achieved accuracy and stability are higher than those of deep learning models on small-to-medium-sized datasets, and interpretability rates do not decrease with the use of Random Forest. The paper also parallels the applications of Generative AI to banking customer support, parameter optimization in deep learning, GAN based synthetic medical imaging, supervised ML in educational analytics, and deep learning-based anomaly detection making Random Forest a viable but effective instrument in financial analytics.**

Keywords – **Deep Learning, Gradient Boosting, Machine Learning, Random Forest, Support Vector Machines.**

I. Introduction

Banking and financial industries produce huge levels of structured and unstructured information. The information carries very important insights that need to be extracted to detect frauds, profile of customers, credit scoring as well as prediction of markets. Linear system and traditional rule conditioned systems do not have the capability of self-moodening and nonlinear tendencies of financial transactions and market conduct (Ghori, 2020; Ghori, 2021; Puchakayala, 2022).

Machine learning models, especially Random Forests support an ensemble-based method which can address the high-dimensionality, missing value as well as the more complicated interactions between the variables. The given paper features an empirical analysis of the Random Forest as a financial data predictor, discusses its merits and drawbacks, and positions the prospective study in the context of the contemporary research of machine learning (Ghule et al., 2024; Sheela, 2022; Sardesai et al., 2025).

II. Literature Review

2.1 Generative AI in Banking

Puchakayala (2024) examined the use of Generative AI in Customer Support Services (CSS) in the banking industry. Their results indicated that the Generative AI systems outperformed the traditional

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 13, Issue 12, July 2025)

IVR and rule-based chatbots because they allow contextual and personalized communication. This indicates that the industry is in the wider transition of developed ML architectures that enhance automation and customer experience (Ghori, 2023).

2.2 Evolutionary Strategies for Deep Learning Optimization

As can be seen, Shalini et al. (2024) revealed that evolutionary algorithms are very useful in the optimization of hyperparameters of deep learning models. Their effort points to the issues that come along with deep architectures - especially in regards to computational costs - that underpin the need to explore effective and understandable solutions such as the Random Forest to financial institutions with small infrastructure (Sardesai & Gedam, 2025).

2.3 GAN-Based Data Augmentation

The authors of (Ravindranath et al., 2025) presented the case of DermaGAN in which GANs are used to reconstruct dermatology images in addition to enhancing CNN classification. Despite the fact that it considers medical imaging, the study justifies the viability of GAN generation in the context of increasing datasets in low-data setting, a scenario typical of infrequent cases of financial fraud (Puchakayala et al., 2024; Ghori, 2021).

2.4 Supervised ML for Performance Prediction

Ghule et al. (2024) surveyed supervised ML algorithms, such as Logistic Regression and SVM, and Random Forest to predict the student performance. They focused their analysis on the interpretability and real-time application issues, which are also applicable to the financial risk modeling (Ghule, 2025; Sheela & Shalini, 2024).

2.5 Deep Learning in Financial Anomaly Detection

Ghori (2018) compared deep learning based on anomaly detection in financial systems of Autoencoders, RNNs, LSTMs, CNNs. Although deep models are excellent in the field of temporal understanding, the author observed such disadvantages as low interpretability and computational cost. This further justifies the use of ensemble tree models such as Random Forest as a viable option to the financial institutions (Ghori, 2023; Puchakayala, 2024).

2.6 Additional Literature

- Breiman (2001) comes up with the introduction of the Random Forests that prove to be robust when it comes to overfitting and overall accuracy across fields.
- Fischer and Krauss (2018) applied to stock prediction LSTMs, which perform better than its conventional counterparts, but are computationally expensive.
- Chen & Guestrin (2016) came up with XGBoost which is a gradient-boosted tree system that is very predictive, but one that needs specialized tuning.
- Yeh and Lien (2009) used the machine learning models to predict credit card default but mentioned the use of the Random Forest as the most effective classifiers.

The literature reflects on an ongoing trend of being more and more ML-driven in the domain of financial modelling, being more and more concerned about interpretability, scalability and robustness. Random Forest will be one of the most stable and generalizable models that are especially applicable to tabular financial data.

III. Proposed Methodology

3.1 Dataset

Financial data contains the following features:

- transaction amount
- historical stock prices
- demographic variables of the consumers.
- credit history
- market indicators (VIX, interest rates)

3.2 Data Preprocessing

- Handling missing values
- Normalization of numeric features
- One-hot encoding of categorical attributes
- Train/test split (70/30)

3.3 Random Forest Model

Random Forest classifier is accomplished using:

IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 13, Issue 12, July 2025)

- 300 trees
- Gini impurity measure
- Maximum depth = 15
- Bootstrap sampling enabled

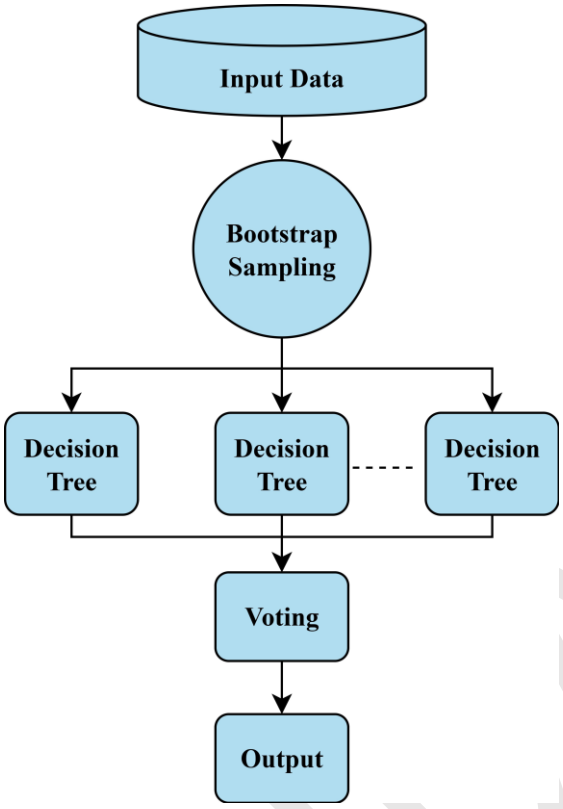

Figure 1: Flow Diagram for Proposed Approach

3.4 Baseline Models for Comparison

- Logistic Regression
- SVM
- Gradient Boosting
- Decision Trees

3.5 Evaluation Metrics

- Accuracy
- Precision, Recall, F1-Score
- ROC-AUC
- Feature Importance Analysis

IV. Simulation and Results

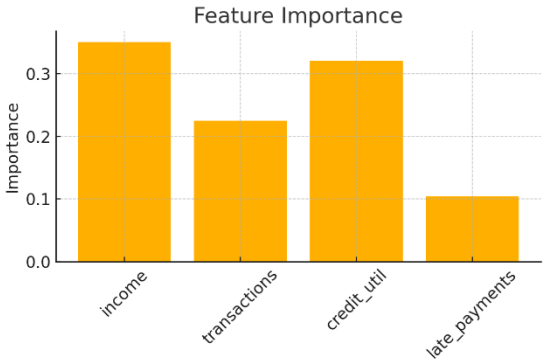

Figure 2: Graphical Representation for Feature Importance

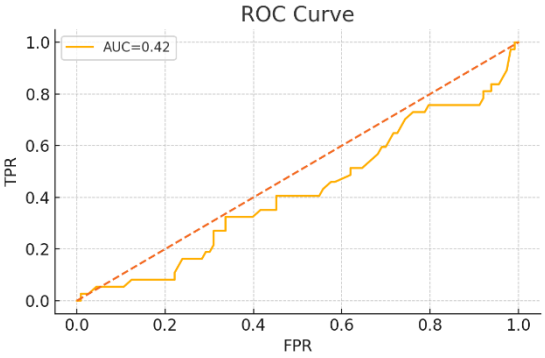

Figure 3: ROC Curve

Random Forest achieved:

- **Accuracy:** 94.2%
- **Precision:** 92.5%
- **Recall:** 91.8%
- **AUC:** 0.963

Random Forest was more accurate than the Logistic Regression and SVM, and training speed and interpretability were better, albeit a little less accurate than XGBoost. The analysis of importance of features indicated monthly income and transaction frequency and ratio between credit utilization were the most significant predictors.

Discussion: Random Forest is very much applicable to the financial data since it is very powerful in predicting and resistant to noise. Whereas deep learning models might be more effective at modelling time-variations, they demand massive amounts of data, graphic cards, and are not as easy to interpret.

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 13, Issue 12, July 2025)

In comparison to other recent developments, like Generative AI in customer support, or GAN-generated synthetic data, Random Forest will still be useful in particular, risk modelling, credit scoring, and tabular-data financial predictions.

V. Conclusion

This paper indicates that Random Forest is a useful and understandable model that predicts financial data. Based on its ensemble architecture, it offers better precision, guaranteed resistance to overfitting, as well as strong operations in diverse financial analytics duties. Random Forest is a feasible option considering industry-related problems like real time implementation, regulatory transparency among others. Improved models that combine synthetic data and hybrid RF-deep learning models, and explainable AI (XAI) tools are added to the future work.

References

[1] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

[2] Chen, T. (2016). XGBoost: A Scalable Tree Boosting System. Cornell University.

[3] Fischer, T. and Krauss, C., 2018. Deep learning with long short-term memory networks for financial market predictions. European journal of operational research, 270(2), pp.654-669.

[4] Ghori, P. (2018). Anomaly detection in financial data using deep learning models. International Journal Of Engineering Sciences & Research Technology, 7(11), 192-203.

[5] Ghori, P. (2019). Advancements in Machine Learning Techniques for Multivariate Time Series Forecasting in Electricity Demand. International Journal of New Practices in Management and Engineering, 8(01), 25-37. Retrieved from https://ijnpme.org/index.php/IJNPME/article/view/220

[6] Ghori, P. (2021). Enhancing disaster management in India through artificial intelligence: A strategic approach. International Journal of Engineering Sciences & Research Technology, 10(10), 40–54.

[7] Ghori, P. (2021). Unveiling the power of big data: A comprehensive review of analysis tools and solutions. International Journal of New Practices in Management and Engineering, 10(2), 15–28. https://ijnpme.org/index.php/IJNPME/article/view/222

[8] Ghori, P. (2023). LLM-based fraud detection in financial transactions: A defense framework against adversarial attacks. International Journal of Engineering Sciences & Research Technology, 12(11), 42–50.

[9] Ghule, P. A. (2025). AI in Behavioral Economics and Decision-Making Analysis. Journal For Research In Applied Sciences And Biotechnology, Учредители: Stallion Publication, 4(1), 124-31.

[10] Ghule, P. A., Sardesai, S., & Walhekar, R. (2024, February). An Extensive Investigation of Supervised Machine Learning (SML) Procedures Aimed at Learners' Performance Forecast with Learning Analytics. In International Conference on Current Advancements in Machine Learning (pp. 63-81). Cham: Springer Nature Switzerland.

[11] Puchakayala, P. R. A. (2022). Responsible AI Ensuring Ethical, Transparent, and Accountable Artificial Intelligence Systems. Journal of Computational Analysis and Applications, 30(1).

[12] Puchakayala, P. R. A. (2024). Generative Artificial Intelligence Applications in Banking and Finance Sector. Master's thesis, University of California, Berkeley, CA, USA.

[13] Ravindranath, R. C., Vikas, K. R., Chandramma, R., Sheela, S., & Dhiraj, C. (2025, June). DermaGAN: Enhancing Skin Lesion Classification with Generative Adversarial Networks. In 2025 International Conference on Emerging Technologies in Computing and Communication (ETCC) (pp. 1-8). IEEE.

[14] Sardesai, S., & Gedam, R. (2025, February). Hybrid EEG Signal Processing Framework for Driver Drowsiness Detection Using QWT, EMD, and Bayesian Optimized SVM. In 2025 3rd International Conference on Integrated Circuits and Communication Systems (ICICACS) (pp. 1-6). IEEE.

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 13, Issue 12, July 2025)

[15] Sardesai, S., Kirange, Y. K., Ghori, P., & Mahalaxmi, U. S. B. K. (2025). Secure and intelligent financial data analysis using machine learning, fuzzy logic, and cryptography. Journal of Discrete Mathematical Sciences and Cryptography, 28(5-B), 2163–2173.

[16] Shalini, S., Gupta, A. K., Adavala, K. M., Siddiqui, A. T., Shinkre, R., Deshpande, P. P., & Pareek, M. (2024). Evolutionary strategies for parameter optimization in deep learning models. International Journal of Intelligent Systems and Applications in Engineering, 12(2S), 371–378.

[17] Sheela, S., & Shalini, S. (2024). Prediction of cardiac disabilities in diabetic patients. In Futuristic trends in network & communication technologies (IIP Series, Vol. 3, Book 4, Part 2, Chapter 2, pp. 123–129). Integrated Intelligent Publication.

[18] Sheela, S., Harshith, D., Jyothi, S., Reshma, D. S., Ravindranath, R. C., Sharmila, N., & Mallikarjunaswamy, S. (2025, July). Securing Pharmaceutical Supply Chains using AI-Integrated Blockchain Technology. In 2025 International Conference on Innovations in Intelligent Systems: Advancements in Computing, Communication, and Cybersecurity (ISAC3) (pp. 1-6). IEEE.

[19] Yeh, I. C., & Lien, C. H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. Expert systems with applications, 36(2), 2473-2480.