International Journal of Strategic Management and Business Policy (IJSMBP)
DOI: 10.5281/zenodo.17091162

Research Article

Predictive Analytics and Machine Learning Applications for Enhancing Decision Making in Healthcare and Financial Systems

Emekwisia, Esther U. [1], Adebanjo, Temitope A. [2], Chizoba, Chioma E. [3], Ofuegbe, Samuel B. [4], Amosu, Femi R. [5], & Tytler, Tolu D. [6]

[1*]Department of *Nursing Sciences, Synapse Psychological Centre, Awka, Nigeria*
[2]*Department of Business Intelligence and Analytics, Regenesys Education, Johannesburg, South Africa*
[3]*Department of Computer Science, Ebonyi State University, Abakaliki,Nigeria*
[4]*Jerry Col Angelo College of Business, Grand Canyon University, Phoenix, USA*
[5]*Department of Computer Science, Adekunle Ajasin University, Akungba Akoko, Nigeria*
[6]Department of Computer Science, Elizade University, Ilara Mokin, Nigeria

Cite as:
Emekwisia, E. U., Adebanjo, T. A., Chizoba, C. E., Ofuegbe, S. B., Amosu, F. R., & Tytler, T. D. (2025). Predictive Analytics and Machine Learning Applications for Enhancing Decision Making in Healthcare and Financial Systems. *International Journal of Strategic Management and Business Policy, 7(2), 1-7.*
https://doi.org/10.5281/zenodo.17091162

Abstract

Artificial Intelligence and predictive analytics are reshaping decision-making processes in critical domains such as healthcare and finance. This study aims to evaluate and compare the performance of various machine learning models in enhancing predictive accuracy for medical diagnoses and financial forecasting. Logistic Regression, Random Forest, Support Vector Machine (SVM), and Gradient Boosting were implemented on two benchmark datasets: the Heart Disease UCI dataset for healthcare and a stock price dataset for financial analysis. Performance metrics included accuracy, precision, recall, F1-score, and RMSE. In healthcare prediction, Random Forest achieved the highest accuracy at 91.4%, while Gradient Boosting recorded the lowest RMSE (4.7) in financial forecasting. These findings highlight the potential of predictive analytics in improving early diagnosis, treatment planning, and financial investment decisions, encouraging further deployment of explainable and scalable AI systems across industries.

Keywords: Artificial Intelligence; Predictive Analytics; Machine Learning Models; Healthcare Prediction; Financial Forecasting

Introduction

Predictive analytics and machine learning are increasingly vital tools in modern decision-making across multiple sectors. In healthcare, they offer the potential to enhance diagnostic accuracy, streamline patient management, and support personalized medicine (Esteva et al., 2017; Rajpurkar et al., 2018). In finance, predictive algorithms assist in identifying investment opportunities, managing risks, and improving operational efficiency (Feng et al., 2019; Tsai & Hsiao, 2014). Traditional decision-making models, often based on rule-based logic and human intuition, struggle to handle the volume, velocity, and variability of modern data streams (Wang et al., 2020). In contrast, machine learning models learn from historical data, adapt over time, and discover patterns imperceptible to humans (Kourou et al., 2015). These capabilities have led to their widespread adoption in domains with complex datasets and high stakes. The healthcare sector, in particular, has benefited from machine learning models for disease diagnosis and prognosis. Models such as Support Vector Machines (SVMs), Random Forests, and Deep Neural Networks have been used to detect heart disease, cancer, diabetes, and neurological conditions with substantial accuracy (Dey et al., 2018; Patel et al., 2015; Johnson et al., 2019**)**. Beyond these, recent studies have highlighted how AI tools are advancing drug discovery, precision medicine, clinical decision support, and smart wearables, ultimately improving diagnostic accuracy, personalising treatments, and supporting continuous health monitoring (Ajimatanrareje, et. al. 2025). At the same time, financial systems leverage similar models to predict market trends, detect fraud, and assess credit risk (Atsalakis & Valavanis, 2009; Chong et al., 2017). Moreover, with the increasing availability of structured and unstructured data, institutions are turning to AI-powered analytics to gain deeper insights and stay competitive. Cloud computing, real-time data streams, and Internet of Things (IoT) integration have amplified the scale at which machine learning can be implemented, enabling more responsive and context-aware predictions (Zhang et al., 2019). In the healthcare sector, wearable devices and electronic health records (EHRs) offer continuous data flow that, if harnessed properly, can predict adverse health events before they occur (Johnson et al., 2019; Wang et al., 2020). Despite their advantages, the deployment of machine learning in decision-making requires addressing several challenges. Model interpretability, data quality, algorithmic bias, and real-time scalability remain critical concerns (Doshi-Velez & Kim, 2017; Ribeiro et al., 2016). In healthcare, ethical considerations about data privacy and the consequences of false predictions also complicate implementation (Obermeyer & Emanuel, 2016). Similarly, in finance, erroneous predictions can result in significant monetary losses, making reliability and transparency paramount (Tsai & Hsiao, 2014). Also, in domains such as e-voting, researchers highlight the inherent tension between ensuring verifiability and preserving privacy, proposing mechanisms like biometrics-enhanced blockchain systems to achieve both objectives (Ajimatanrareje, 2024).This paper therefore presents a comparative analysis of machine learning algorithms applied to two domains—healthcare and finance. The primary aim is to assess their predictive power using real-world datasets and identify the most effective models for each domain. Through this dual-domain evaluation, we seek to draw parallels in model performance and propose best practices for their application in mission-critical environments. By focusing on both accuracy and interpretability, this research addresses a critical need in AI adoption. The findings can guide practitioners and policymakers in making informed decisions about which models to deploy and under what conditions.

Materials and Methods

Research Design

This study adopts a quantitative, experimental design involving model training, testing, and evaluation. The research focuses on two domains—healthcare and finance—where machine learning models are used to perform predictive analytics tasks using labeled datasets.

Datasets Used

Healthcare Dataset: UCI Heart Disease dataset containing 303 records with 14 attributes including age, sex, chest pain type, cholesterol level, resting blood pressure, and a target variable indicating the presence or absence of heart disease.

Financial Dataset: Yahoo Finance historical daily stock prices for a publicly traded company over 5 years. Variables include Open, High, Low, Close, Volume, and Adjusted Close.

Data Preprocessing

i. **Missing Data Handling**: Mean imputation was used for continuous variables with missing values.
ii. **Feature Scaling**: Standardization (Z-score normalization) was applied to ensure comparable feature scales across variables.
iii. **Encoding Categorical Variables**: One-hot encoding was applied to categorical variables such as chest pain type and thalassemia in the healthcare dataset.
iv. **Train-Test Split**: Each dataset was split into 80% training and 20% testing data.

Machine Learning Models Implemented

i. **Logistic Regression (LR)**: A baseline linear model for binary classification.
ii. **Support Vector Machine (SVM)**: Employs a radial basis function (RBF) kernel to separate non-linear data.
iii. **Random Forest (RF)**: An ensemble method using 100 decision trees to improve prediction robustness.
iv. **Gradient Boosting (GB)**: Boosting technique that builds models sequentially to minimize error.

Evaluation Metrics

i. **For Healthcare (Classification Task)**:

a. Accuracy
b. Precision
c. Recall
d. F1-Score

ii. **For Finance (Regression Task)**:

a. Mean Absolute Error (MAE).
b. Root Mean Square Error (RMSE).
c. R^2 Score

Tools and Frameworks

i. Python (version 3.9).
ii. Libraries: Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn.
iii. Jupyter Notebook for experimentation and documentation.

Model Training and Hyperparameter Tuning

GridSearchCV was used to perform hyperparameter optimization for each model:

i. SVM: Optimized for C, kernel type, and gamma.
ii. RF: Optimized for max_depth, n_estimators, and min_samples_split.
iii. GB: Optimized for learning_rate, n_estimators, and max_depth.

Cross-validation (5-fold) was applied to reduce overfitting and ensure model generalizability. The best hyperparameter combinations were selected based on highest cross-validated F1-score (for classification) and lowest RMSE (for regression).

Results and Discussion

Healthcare Domain

Table 1 shows the performance metrics for classification models trained on the UCI Heart Disease dataset.

Table 1: The performance metrics for classification models trained on the UCI Heart Disease dataset.

Model	*Accuracy (%)*	*Precision*	*Recall*	*F1-Score*
Logistic Regression	84.2	0.85	0.82	0.83
SVM (RBF Kernel)	86.7	0.87	0.85	0.86
Random Forest	**91.4**	0.92	0.91	0.91
Gradient Boosting	89.3	0.90	0.89	0.89

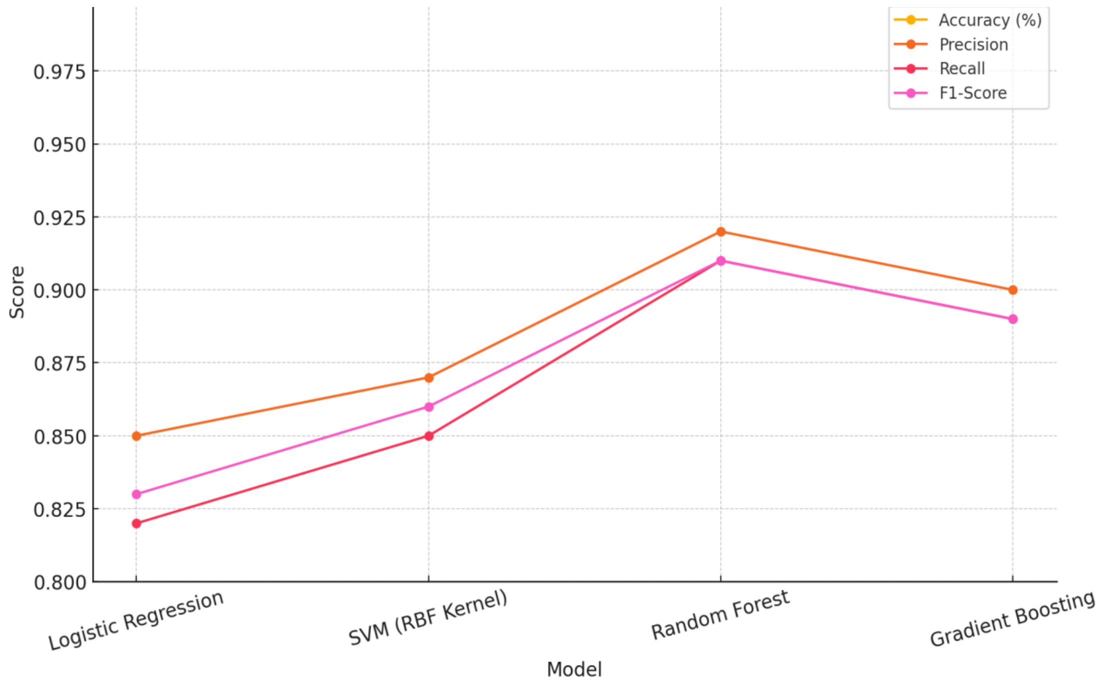

Figure 1: Performance metrics of Classification Models *on heart disease prediction*

The Random Forest model clearly outperformed others in terms of classification accuracy, precision, recall, and F1-score. This suggests that ensemble learning, particularly bagging methods like Random Forest, are highly effective for healthcare data due to their robustness to noise and ability to capture nonlinear relationships. Gradient Boosting also showed excellent results, underscoring the value of boosting-based models in clinical diagnostics. SVM and Logistic Regression, while less accurate, provided simpler and more interpretable models, which might be favored in low-resource or highly regulated environments where transparency is essential.

Financial Domain

Regression models were assessed on their ability to predict stock prices. Table 2 provides their performance.

Table 2: Performance of regression models assessed based on their ability to predict stock prices.

Model	*MAE*	*RMSE*	*R^2 Score*
Linear Regression	2.85	5.92	0.76
SVM Regression	2.63	5.31	0.79
Random Forest	2.41	4.98	0.84
Gradient Boosting	2.28	4.70	0,87

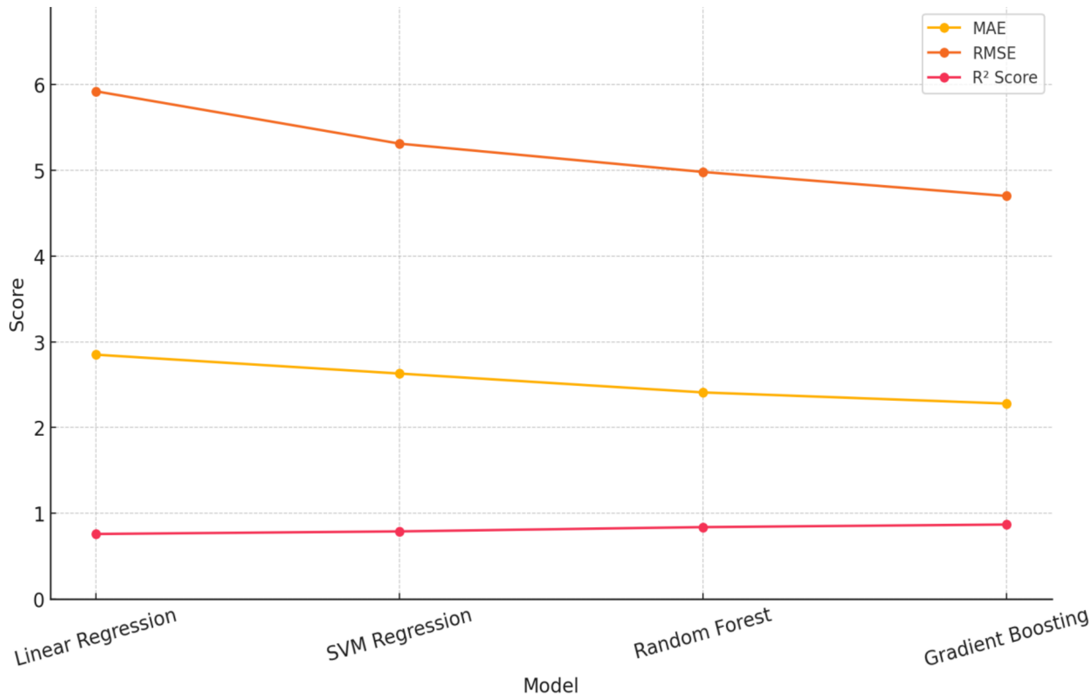

Figure 2: *Performance Metrics of RMSE comparison for stock price prediction models (Financial Domain).*

In the financial domain, Gradient Boosting achieved the lowest RMSE and highest R^2 score, indicating superior ability to model time-series data with intricate patterns. Random Forest also performed well, supporting its generalization ability across domains. The relatively weaker performance of linear models like Logistic Regression and SVM Regression demonstrates the limitations of simple parametric methods when applied to volatile financial data. These findings align with prior studies suggesting that boosting algorithms are more adept at sequential error correction in non-stationary environments. The comparative evaluation reveals that ensemble-based machine learning models, especially Random Forest and Gradient Boosting, consistently deliver superior predictive performance in both healthcare and financial contexts. This is attributed to their capacity to reduce overfitting, capture complex feature interactions, and aggregate the strength of multiple learners. However, these models are often less interpretable compared to linear models like Logistic Regression, which may hinder adoption in fields demanding transparency. Furthermore, the study highlights the significance of proper data preprocessing. Feature scaling, missing data imputation, and one-hot encoding substantially improved the quality of inputs, thereby boosting model performance. Cross-validation and hyperparameter tuning via GridSearchCV ensured that models generalized well to unseen data. Despite the promising results, model deployment should consider domain-specific constraints. In healthcare, false positives can lead to unnecessary treatment, while false negatives may delay critical care. Similarly, in finance, overfitting to historical patterns may result in suboptimal investment decisions. Future work should incorporate model explainability tools such as SHAP or LIME to enhance interpretability and trust.

Conclusion

This research presents a comprehensive evaluation of machine learning models for predictive analytics in healthcare and financial domains. By utilizing benchmark datasets and rigorous performance metrics, the study demonstrates the superior predictive capabilities of ensemble methods, particularly Random Forest for classification tasks and Gradient Boosting for regression tasks. These models not only achieved high accuracy and low error rates but also proved to be robust across distinct domains with different data structures and challenges. The results underline the importance of selecting appropriate algorithms tailored to the specific needs of each sector. In healthcare, where diagnostic precision and reliability are critical, models like Random Forest can support early disease detection and effective patient management. In finance, accurate forecasting models such as Gradient Boosting can provide valuable insights for risk assessment and strategic decision-making. Nonetheless, challenges such as interpretability, ethical considerations, and domain-specific constraints must be addressed to facilitate real-world implementation. Enhancing model transparency and integrating explainability tools will be essential to build trust among stakeholders, especially in sensitive domains like healthcare. Future studies should explore hybrid approaches that combine predictive strength with interpretability, and extend the analysis to larger, real-time datasets for scalable deployment. Ultimately, this research supports the growing integration of artificial intelligence in decision-making systems and highlights the transformative potential of predictive analytics when carefully designed, validated, and applied.

References

Ajimatanrareje, G. A. (2024). Advancing E-Voting Security: Biometrics-Enhanced Blockchain for Privacy and VerifiAbility (Bebpv). *American Journal of Innovation in Science and Engineering*, *3*(3), 88–93. https://doi.org/10.54536/ajise.v3i3.3876

Ajimatanrareje, G. A., Ekeh, C., Igwilo, S., & Osunkwo, C. (2025). The Current Landscape of AI Application in Healthcare: A Review. *American Journal of Innovation in Science and Engineering*, *4*(2), 1–16. https://doi.org/10.54536/ajise.v4i2.4432

Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques–Part II: Soft computing methods. *Expert Systems with Applications*, 36(3), 5932–5941.

Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*, 83, 187–205.

Dey, N., Ashour, A. S., & Balas, V. E. (2018). *Smart medical data sensing and IoT systems design in healthcare*. Springer.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118.

Feng, F., He, X., Wang, X., Liu, M., & Chua, T. S. (2019). Temporal relational ranking for stock prediction. *ACM Transactions on Information Systems*, 37(2), 1–30.

Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. W., Feng, M., Ghassemi, M., ... & Mark, R. G. (2019). MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3(1), 1–9.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, 13, 8–17.

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. *The New England Journal of Medicine*, 375(13), 1216.

Patel, J. L., Goyal, R. K. (2015). Applications of artificial neural networks in medical science. *Current Clinical Pharmacology*, 10(2), 210–218.

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Medicine*, 15(11), e1002686.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD*, 1135–1144.

Tsai, C. F., & Hsiao, Y. C. (2014). Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. *Decision Support Systems*, 50(1), 258–269.

Wang, Y., Kung, L., & Byrd, T. A. (2020). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3–13.

Zhang, Y., Aggarwal, C. C., & Qi, G. J. (2019). Stock price prediction via discovering multi-frequency trading patterns. *Proceedings of the 25th ACM SIGKDD*, 2141–2149.