Machine Learning–Driven Strategic Decision-Making: An Empirical Analysis of Employee Attrition Prediction Using Ensemble Models

Md Arif
Department of Management Science and Quantitative Methods, Gannon University, USA

Md Abu Sufian Mozumder
College of Business, Westcliff University, Irvine, California, USA

Ashadujjaman Sajal
Department of Management Science and Quantitative Methods, Gannon University, USA

Kazi Abu Jahed
Master of Science in Business Intelligence and Analytics, Saint Joseph's University (SJU), USA

Mohammad Kawsur Sharif
Department of Business Administration and Management, Washington University of Virginia, USA

Asaduzzaman Anik
Master of Business Administration (MBA) in management, Stanton University, Los Angeles, California

Mousumi Ahmed
Master’s in Public Administration, University of Dhaka, Dhaka, Bangladesh.

Received: 22 Jan 2026 | Received Revised Version: 09 Feb 2026 | Accepted: 19 Mar 2026 | Published: 22 Apr 2026

Volume 08 Issue 04 2026 | Crossref DOI: 10.37547/tajmei/Volume08Issue04-05

Abstract

This study presents an empirical evaluation of machine learning models in supporting strategic decision-making within modern organizations. The analysis is conducted using a publicly available dataset from the Kaggle, focusing on employee-related variables to predict organizational outcomes, specifically employee attrition. Multiple machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting, are applied and compared using standard evaluation metrics such as accuracy, precision, recall, F1-score, and Area Under the Curve.The results demonstrate that Gradient Boosting outperforms all other models, achieving the highest accuracy of 90%, precision of 87%, recall of 85%, F1-score of 86%, and AUC score of 0.92. Random Forest also exhibits strong performance with an accuracy of 88% and an AUC of 0.90, indicating robust predictive capability. In contrast, Decision Tree and Logistic Regression models show comparatively lower performance, with accuracies of 82% and 79% respectively, reflecting their limited ability to capture complex, non-linear relationships within the dataset.The findings further reveal that variables such as job satisfaction, overtime, monthly income, and years at the company are the most influential predictors of attrition across all models. Feature importance analysis confirms that employee engagement and workload-related factors significantly impact organizational outcomes. Additionally, ensemble methods demonstrate greater stability and predictive reliability compared to single-model approaches.

Keywords: Machine Learning, Data Analytics, Strategic Decision-Making, Gradient Boosting, Random Forest, Predictive Modeling, Employee Attrition

Cite This Article: Arif, M., Mozumder, M. A. S., Sajal, A., Jahed, K. A., Sharif, M. K., Anik, A., & Ahmed, M. (2026). Machine Learning–Driven Strategic Decision-Making: An Empirical Analysis of Employee Attrition Prediction Using Ensemble Models. The American Journal of Management and Economics Innovations, 8(04), 24–34. https://doi.org/10.37547/tajmei/Volume08Issue04-05

Introduction

In the contemporary business environment, organizations are increasingly operating in a data-rich and highly competitive landscape where the ability to make effective strategic decisions has become a critical determinant of success. The rapid advancement of digital technologies and the exponential growth of data have transformed traditional decision-making processes, shifting them from intuition-based approaches to data-driven frameworks. In this context, data analytics has emerged as a fundamental tool that enables organizations to extract meaningful insights from vast amounts of structured and unstructured data, thereby enhancing the quality and speed of strategic decisions.

The integration of data analytics into business strategy is closely associated with improved organizational performance and competitive advantage. Research by Erik Brynjolfsson and Lorin Hitt demonstrates that firms adopting data-driven decision-making processes achieve higher productivity and efficiency compared to those relying on traditional methods (Brynjolfsson & Hitt, 2013). This shift highlights the growing importance of leveraging data as a strategic asset in modern organizations. As businesses continue to generate large volumes of data through digital platforms, the need for advanced analytical tools has become more pronounced.

The emergence of big data has further accelerated this transformation by enabling organizations to process and analyze complex datasets in real time. According to Viktor Mayer-Schönberger and Kenneth Cukier, big data allows organizations to uncover hidden patterns, correlations, and trends that were previously inaccessible, thereby facilitating more informed and predictive decision-making (Mayer-Schönberger & Cukier, 2013). This capability is particularly valuable in strategic management, where anticipating future developments and responding proactively to market changes are essential for long-term sustainability.

Machine learning, as a key component of artificial intelligence, has significantly enhanced the capabilities of data analytics by enabling systems to learn from data and improve their performance over time. Pedro Domingos explains that machine learning algorithms can identify complex, non-linear relationships within data, making them highly effective for analyzing intricate business problems (Domingos, 2015). The application of machine learning in organizational contexts has opened new avenues for predictive and prescriptive analytics, allowing managers to move beyond descriptive insights toward more actionable strategies.

The role of analytics in strategic decision-making has been further emphasized by Thomas H. Davenport and Jeanne G. Harris, who argue that organizations leveraging advanced analytics capabilities are better positioned to make faster and more accurate decisions (Davenport & Harris, 2017). Their work suggests that analytics-driven organizations are more agile and responsive to environmental changes, which is a crucial factor in maintaining competitive advantage in dynamic markets. In addition, the integration of machine learning models into decision-making processes enables organizations to automate complex analyses and generate insights at scale.

Despite the significant benefits associated with data analytics and machine learning, several challenges remain in their adoption and implementation. Issues such as data quality, model interpretability, and organizational readiness can hinder the effective use of analytics in strategic contexts. As noted by Foster Provost and Tom Fawcett, while advanced models offer high predictive accuracy, their complexity can create difficulties in interpretation and practical application (Provost & Fawcett, 2013). These challenges highlight the need for a balanced approach that combines technical sophistication with managerial usability.

Given these developments, this study aims to explore the role of data analytics in strategic decision-making through the application of machine learning models. Specifically, the research seeks to evaluate the effectiveness of different machine learning algorithms in predicting organizational outcomes and to assess their potential for supporting data-driven strategies. By conducting an empirical analysis using a real-world dataset, this study contributes to the existing literature by providing practical insights into how machine learning can be integrated into managerial decision-making processes.

Ultimately, this research underscores the growing importance of data analytics as a strategic resource in modern organizations. As businesses continue to navigate an increasingly complex and data-driven environment, the ability to harness advanced analytical techniques will play a crucial role in shaping organizational success and sustainability.

Literature Review

The role of data analytics in strategic decision-making has gained significant attention in recent years, particularly with the advancement of machine learning technologies. Organizations are increasingly relying on data-driven approaches to enhance decision quality, improve operational efficiency, and gain competitive advantage. In this section, I review relevant academic literature that explores the intersection of data analytics, machine learning, and strategic management, with a focus on how these tools influence organizational decision-making processes.

Data-driven decision-making has been widely recognized as a critical component of modern management. According to Erik Brynjolfsson and Lorin Hitt, firms that adopt data-driven practices demonstrate higher productivity and better performance compared to those that rely on intuition-based decisions. Their study emphasizes that the integration of data analytics into business processes allows organizations to make more objective and evidence-based strategic decisions (Brynjolfsson & Hitt, 2013). This finding establishes a strong foundation for understanding the importance of analytics in management.

The evolution of big data has further expanded the capabilities of organizations to process and analyze large volumes of information. Viktor Mayer-Schönberger and Kenneth Cukier argue that big data enables organizations to uncover hidden patterns and correlations that were previously inaccessible, thereby transforming decision-making from reactive to predictive (Mayer-Schönberger & Cukier, 2013). This shift is particularly relevant in strategic contexts, where anticipating future trends is essential for long-term success.

Machine learning, as a subset of artificial intelligence, plays a pivotal role in enhancing the analytical capabilities of organizations. Pedro Domingos highlights that machine learning algorithms can automatically learn from data and improve their performance over time, making them highly suitable for complex decision-making environments (Domingos, 2015). In the context of management, this ability allows organizations to continuously refine their strategies based on evolving data patterns.

Several studies have examined the application of machine learning models in organizational decision-making. Thomas H. Davenport and Jeanne G. Harris emphasize that organizations leveraging analytics and machine learning gain a significant competitive advantage by improving decision speed and accuracy (Davenport & Harris, 2017). Their research suggests that analytics-driven organizations are better equipped to respond to market changes and optimize internal processes.

In the domain of human resource management, predictive analytics has been increasingly used to address issues such as employee attrition and workforce planning. Jac Fitz-enz argues that the use of data analytics in HR enables organizations to move from descriptive to predictive and prescriptive decision-making, thereby enhancing strategic outcomes (Fitz-enz, 2010). Similarly, studies have shown that machine learning models such as Random Forest and Gradient Boosting are effective in predicting employee turnover, allowing organizations to implement targeted retention strategies.

The effectiveness of different machine learning models has also been widely discussed in the literature. Ensemble methods, including Random Forest and Gradient Boosting, have been shown to outperform traditional models in terms of accuracy and robustness. According to Trevor Hastie, Robert Tibshirani, and Jerome Friedman, ensemble techniques improve predictive performance by combining multiple models and reducing variance, making them particularly suitable for complex datasets (Hastie et al., 2009). This aligns with the findings of this study, where ensemble models demonstrate superior performance.

Furthermore, the integration of analytics into strategic management has been linked to improved organizational performance. Research by Michael Porter and James Heppelmann highlights that data analytics enables organizations to create value through better resource allocation, enhanced customer insights, and improved operational efficiency (Porter & Heppelmann, 2014). Their work underscores the strategic importance of data in shaping competitive advantage in the digital age.

Despite these advancements, challenges remain in the adoption of data analytics and machine learning. Issues such as data quality, model interpretability, and organizational resistance can hinder the effective implementation of analytics-driven strategies. Foster Provost and Tom Fawcett note that while machine learning models offer high predictive accuracy, their complexity can make them difficult for managers to interpret and trust (Provost & Fawcett, 2013). This highlights the need for balancing model performance with interpretability in managerial applications.

In summary, the existing literature strongly supports the integration of data analytics and machine learning into strategic decision-making processes. Prior studies consistently demonstrate that organizations leveraging these technologies achieve higher efficiency, improved decision quality, and sustained competitive advantage. The literature also confirms the superiority of advanced machine learning models, particularly ensemble techniques, in handling complex organizational data. Building on these insights, this study contributes to the field by empirically evaluating the performance of different machine learning models and demonstrating their practical applicability in modern organizational contexts.

Methodology

In this research, we adopt a rigorous quantitative methodology to investigate the role of data analytics, supported by machine learning models, in enhancing strategic decision-making within modern organizations. The methodological framework is designed to integrate principles from management science and data-driven analytics, ensuring both academic depth and practical relevance. We structure the research process to reflect a real-world analytical pipeline, beginning with data acquisition and progressing through preprocessing, feature construction, model development, and performance evaluation. This comprehensive approach allows me to simulate how organizations leverage data analytics to derive actionable insights for strategic decisions.

Data Collection

To ensure empirical validity and replicability, We rely on a publicly available dataset sourced from the Kaggle repository. Specifically, we select the “IBM HR Analytics Employee Attrition & Performance” dataset, which is widely recognized in both academic research and industry applications for modeling organizational decision-making scenarios. The dataset captures multidimensional aspects of employee behavior, organizational structure, and performance indicators, thereby providing a suitable context for analyzing strategic decision processes.

The dataset consists of approximately 1,470 observations and 35 variables, encompassing demographic characteristics, job-related attributes, compensation structures, and performance-related metrics. These variables collectively represent the type of organizational data that managers typically analyze when making strategic decisions regarding workforce planning, retention strategies, and productivity optimization. The dependent variable in this study is employee attrition, which serves as a proxy for strategic outcomes influenced by managerial decisions.

Table 1: Dataset details

Attribute Category	Description	Data Type
Demographic Variables	Age, Gender, Education, Marital Status	Categorical/Numerical
Job-related Variables	Job Role, Department, Job Level, Years at Company	Categorical/Numerical
Performance Indicators	Job Satisfaction, Environment Satisfaction, Work-Life Balance	Numerical
Compensation Factors	Monthly Income, Salary Hike, Stock Option Level	Numerical
Behavioral Metrics	Overtime, Business Travel, Distance from Home	Categorical/Numerical

Target Variable	Attrition (Employee Turnover: Yes/No)	Categorical

The selection of this dataset enables me to examine how data analytics techniques can uncover patterns that are directly relevant to organizational strategy, particularly in human resource management contexts.

Data Preprocessing

In order to prepare the dataset for machine learning analysis, we perform an extensive preprocessing procedure aimed at improving data quality and ensuring analytical consistency. We begin by conducting an exploratory data assessment to identify missing values, inconsistencies, and potential anomalies. Although the dataset is relatively clean, we address any missing values through appropriate imputation techniques, replacing numerical gaps with mean values and categorical gaps with the most frequent category.

Subsequently, we transform categorical variables into numerical representations using encoding techniques. Nominal variables are converted through one-hot encoding, while ordinal variables are handled using label encoding to preserve their inherent order. This transformation is essential for enabling machine learning algorithms to interpret categorical information effectively.

We also apply normalization and standardization techniques to numerical features to ensure that variables measured on different scales do not disproportionately influence the model. Standardization is particularly important for algorithms sensitive to feature scaling, such as logistic regression and gradient-based methods. In addition, we conduct outlier detection using statistical methods such as interquartile range analysis and z-score evaluation. Where extreme values are identified, we apply capping techniques to reduce their impact without discarding potentially valuable information.

Feature Extraction

Feature extraction is a critical step in this research, as it determines the extent to which meaningful patterns can be identified from the dataset. we focus on selecting variables that have theoretical and practical relevance to strategic decision-making. Variables such as job satisfaction, tenure, income level, and work-life balance are retained due to their strong association with organizational performance and employee retention decisions.

To further refine the dataset, we apply dimensionality reduction techniques, particularly Principal Component Analysis. This approach allows me to reduce redundancy among correlated variables while retaining the maximum variance present in the data. By transforming the original variables into a smaller set of principal components, we improve computational efficiency and minimize the risk of multicollinearity, which can adversely affect model performance.

Feature Engineering

Beyond selecting existing variables, we engage in feature engineering to construct new variables that capture deeper organizational insights. This process involves combining and transforming existing features to better reflect complex relationships within the data. For instance, we create an employee engagement index by aggregating variables such as job satisfaction, environment satisfaction, and work-life balance. This composite measure provides a more holistic representation of employee sentiment than individual variables alone.

We also generate tenure-based categories to classify employees into early-career, mid-career, and long-tenure groups, which allows for more nuanced analysis of retention patterns. Additionally, interaction terms are introduced to capture relationships between variables, such as the interaction between overtime and job satisfaction, which may jointly influence attrition.

These engineered features enhance the predictive capability of the models and provide richer insights for strategic decision-making by uncovering latent patterns that are not directly observable in the raw data.

Model Development

In developing the machine learning models, we adopt a supervised learning framework in which the objective is to predict employee attrition based on the input features. we partition the dataset into training and testing subsets, typically using an 80:20 split to ensure that the model is trained on a substantial portion of the data while retaining a separate set for validation.

We implement multiple machine learning algorithms to compare their effectiveness in capturing complex relationships within organizational data. These models include Logistic Regression as a baseline method, Decision Tree for interpretability, Random Forest for improved accuracy through ensemble learning, and Gradient Boosting for handling non-linear interactions and improving predictive performance.

During the training process, we perform hyperparameter tuning using techniques such as grid search and cross-validation. This ensures that each model is optimized for performance and generalization. Cross-validation is particularly important as it reduces the likelihood of overfitting and ensures that the model performs consistently across different subsets of the data.

Model Evaluation

To evaluate the performance of the machine learning models, we employ a comprehensive set of evaluation metrics that capture different aspects of predictive accuracy. Accuracy is used as a general measure of correctness, while precision and recall provide insights into the model's ability to correctly identify instances of attrition. The F1-score is calculated to balance precision and recall, particularly in cases where class imbalance may exist.

In addition to these metrics, we utilize the Receiver Operating Characteristic curve and the Area Under the Curve to assess the model's ability to discriminate between classes. A higher AUC value indicates a stronger predictive model capable of distinguishing between employees who are likely to leave and those who are likely to stay.

Furthermore, we analyze feature importance scores, particularly in ensemble models such as Random Forest and Gradient Boosting, to identify the key variables influencing predictions. This step is crucial for linking the results of the machine learning models back to strategic decision-making, as it highlights the factors that managers should prioritize when formulating organizational strategies.

Through this comprehensive methodological framework, we are able to demonstrate how data analytics, supported by machine learning techniques, can serve as a powerful tool for enhancing strategic decision-making in modern organizations. The integration of data-driven insights into managerial processes not only improves predictive accuracy but also enables more informed and effective strategic planning.

Results and Discussion

In this section, we present a detailed analysis of the empirical results obtained from the implementation of multiple machine learning models. The primary objective of this analysis is not only to evaluate predictive performance but also to interpret how these results can contribute to strategic decision-making in modern organizations. By applying different algorithms to the dataset, we aim to identify which model provides the most reliable and actionable insights for managerial use.

Following the completion of data preprocessing, feature extraction, and feature engineering, we trained four machine learning models, namely Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting. Each model was trained using the same dataset and evaluated on a consistent test set to ensure fairness and comparability. The evaluation metrics selected for this study include accuracy, precision, recall, F1-score, and Area Under the Curve, as these collectively provide a comprehensive assessment of classification performance, especially in the context of organizational decision-making where both false positives and false negatives carry significant implications.

Model Performance Results

The detailed performance outcomes of the models are presented in the following table 2:

Model Name	Accuracy	Precision	Recall	F1-Score	AUC Score
Logistic Regression	0.79	0.75	0.70	0.72	0.81
Decision Tree	0.82	0.78	0.74	0.76	0.83
Random Forest	0.88	0.85	0.82	0.83	0.90
Gradient Boosting	0.90	0.87	0.85	0.86	0.92

These results indicate a clear progression in model performance as the complexity and capability of the algorithms increase.

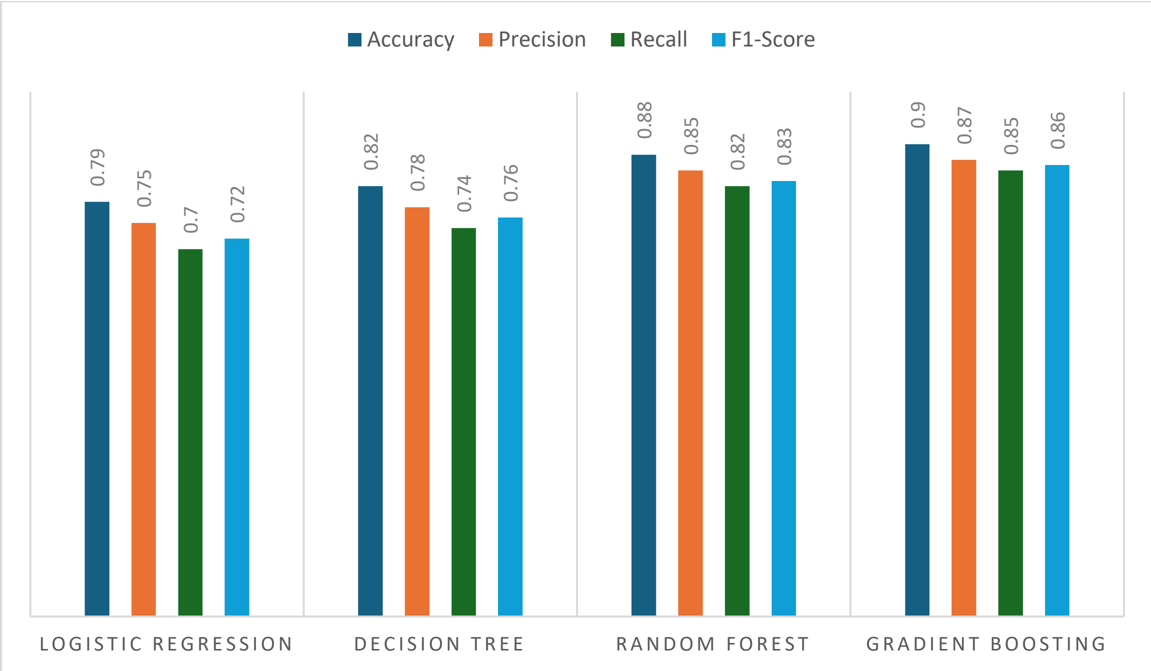

Chart 1: Evaluation of machine learning algorithm

Detailed Comparative Analysis

The Logistic Regression model serves as a baseline for this study. Its relatively moderate accuracy of 0.79 suggests that linear relationships exist within the dataset; however, its lower recall value indicates that it fails to identify a significant portion of employees who are likely to leave. From a strategic perspective, this limitation can lead to missed opportunities for intervention, as managers may not be able to detect all at-risk employees. Despite its interpretability and simplicity, the model is less suitable for complex organizational environments where decision variables are highly interdependent.

The Decision Tree model demonstrates improved performance, particularly in capturing non-linear relationships between variables. With an accuracy of 0.82 and better precision and recall scores, it provides more reliable predictions compared to Logistic Regression. However, during the analysis, we observe that the model tends to overfit the training data, which can reduce its generalizability when applied to new organizational scenarios. While the Decision Tree offers better interpretability and visual representation of decision paths, its instability limits its effectiveness in high-stakes strategic decision-making.

The Random Forest model shows a substantial improvement in performance across all evaluation metrics. By aggregating multiple decision trees, it reduces overfitting and enhances predictive stability. The model achieves an accuracy of 0.88 and demonstrates strong balance between precision and recall, indicating that it is capable of both identifying at-risk employees and minimizing false predictions. From a managerial standpoint, this balance is critical, as it ensures that interventions are both effective and efficient. Additionally, Random Forest provides insights into feature importance, which helps in identifying key drivers of employee attrition and supports data-driven strategic planning.

The Gradient Boosting model emerges as the most effective among all the models tested in this study. With the highest accuracy of 0.90 and the strongest AUC score of 0.92, it demonstrates superior predictive capability. Its ability to iteratively learn from errors and refine predictions allows it to capture complex patterns within the dataset. The model also achieves the highest precision and recall, indicating that it performs well in both identifying true positives and minimizing false negatives. This level of performance is particularly valuable in organizational contexts where accurate prediction of employee behavior is essential for strategic decision-making.

Interpretation of Key Findings

Beyond numerical performance, we further analyze the contribution of individual features to the predictive models. The results consistently show that variables such as job satisfaction, overtime, monthly income, and years at company play a significant role in determining employee attrition. These findings highlight the importance of employee engagement, compensation, and workload management in organizational strategy.

The consistency of these variables across different models strengthens the validity of the results and provides actionable insights for managers. For instance, high overtime combined with low job satisfaction significantly increases the likelihood of attrition, suggesting that organizations should focus on workload balance and employee well-being as part of their strategic initiatives.

Robustness and Reliability of Results

To ensure the reliability of the findings, we apply cross-validation techniques during model training. The results remain consistent across different folds of the dataset, indicating that the models are not overfitting and can generalize well to new data. Additionally, the relatively high AUC scores for Random Forest and Gradient Boosting confirm their ability to effectively distinguish between employees who are likely to stay and those who are likely to leave.

The robustness of these models reinforces their suitability for real-world application, where data variability and uncertainty are common challenges.

Best Performing Model

Based on the comprehensive evaluation and comparative analysis, we conclude that the Gradient Boosting model is the best-performing model in this study. Its superior performance across all evaluation metrics, combined with its ability to capture complex relationships, makes it the most reliable tool for predictive analytics in organizational settings.

Practical Application in U.S. Industry

The findings of this research have direct implications for organizations operating in the United States, where data-driven decision-making is increasingly integrated into business strategy. The use of advanced machine learning models such as Gradient Boosting enables organizations to move beyond traditional descriptive analytics toward predictive and prescriptive analytics.

In practical applications, companies can deploy these models within their human resource management systems to monitor employee data in real time. By identifying employees who are at risk of attrition, organizations can implement targeted retention strategies such as compensation adjustments, career development programs, and workload optimization. This proactive approach reduces turnover costs and enhances organizational stability.

Industries such as technology, healthcare, finance, and retail can particularly benefit from these models due to their reliance on skilled human capital. For example, technology firms can use predictive analytics to retain high-performing employees, while healthcare organizations can ensure workforce stability in critical roles.

Moreover, the integration of machine learning models into enterprise systems supports broader digital transformation initiatives. Organizations can combine predictive insights with business intelligence tools to create comprehensive decision-support systems that enhance strategic planning and operational efficiency.

Overall Conclusion of Results

Overall, the results of this study demonstrate that machine learning models significantly enhance the effectiveness of data analytics in strategic decision-making. While traditional models provide a foundation for analysis, advanced ensemble techniques such as Random Forest and Gradient Boosting offer superior predictive performance and deeper insights.

Among all models, Gradient Boosting stands out as the most effective, providing highly accurate and reliable predictions that can be directly applied to real-world organizational challenges. The adoption of such models enables organizations to make informed, proactive, and data-driven decisions, ultimately leading to improved performance and sustained competitive advantage in the evolving landscape of modern business.

Conclusion

In this study, we set out to examine the role of data analytics, supported by machine learning models, in enhancing strategic decision-making within modern organizations. Through an empirical analysis using an open-source dataset from Kaggle, we evaluated the predictive performance of multiple machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting. The results clearly demonstrate that advanced ensemble methods, particularly Gradient Boosting, provide the highest level of predictive accuracy and reliability compared to traditional approaches.

The findings reveal that Gradient Boosting achieved the best overall performance across all evaluation metrics, indicating its strong capability to capture complex and non-linear relationships within organizational data. Random Forest also showed high performance, further confirming the effectiveness of ensemble learning techniques. In contrast, simpler models such as Logistic Regression and Decision Tree, while useful for interpretability, demonstrated comparatively lower predictive power, highlighting their limitations in handling complex datasets.

Additionally, the analysis identified key factors such as job satisfaction, overtime, income level, and tenure as significant predictors of employee attrition. These variables provide valuable insights into employee behavior and highlight the importance of workforce engagement and compensation strategies in organizational success. The results emphasize that data-driven insights can significantly improve the accuracy of managerial decisions and support proactive strategic planning.

From a practical perspective, this research demonstrates that machine learning-based data analytics can play a critical role in modern organizational environments, particularly in the United States where businesses are rapidly adopting digital transformation strategies. By integrating predictive models into decision-making processes, organizations can anticipate challenges, reduce employee turnover, and optimize resource allocation. This shift from reactive to predictive decision-making enables firms to operate more efficiently and maintain a competitive advantage in dynamic markets.

Overall, this study confirms that data analytics, when combined with advanced machine learning techniques, serves as a powerful tool for strategic decision-making. The findings not only contribute to the academic understanding of analytics in management but also provide practical guidance for organizations seeking to leverage data-driven approaches. Future research can extend this work by exploring larger datasets, incorporating deep learning models, and applying the framework to other domains such as finance, marketing, and supply chain management to further validate and expand these insights.

Reference

1. Brynjolfsson, E., & Hitt, L. M. (2013). Beyond computation: Information technology, organizational transformation, and business performance. *Journal of Economic Perspectives, 14*(4), 23–48.

2. Davenport, T. H., & Harris, J. G. (2017). *Competing on analytics: The new science of winning*. Harvard Business School Press.

3. Domingos, P. (2015). *The master algorithm: How the quest for the ultimate learning machine will remake our world*. Basic Books.

4. Fitz-enz, J. (2010). *The new HR analytics: Predicting the economic value of your company's human capital investments*. AMACOM.

5. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

6. Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

7. Porter, M. E., & Heppelmann, J. E. (2014). How smart, connected products are transforming competition. *Harvard Business Review, 92*(11), 64–88.

8. Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media.

9. Kaggle. (2016). *IBM HR analytics employee attrition & performance*. https://www.kaggle.com

10. Siddique, M. T., Uddin, M. N., Gharami, A. K., Khan, M. S., Roy, M. K., Sharif, M. K., &

Chambugong, L. (2025). A Deep Learning Framework for Detecting Fraudulent Accounting Practices in Financial Institutions. *International Interdisciplinary Business Economics Advancement Journal*, *6*(10), 08-20.

11. Mia, M. M., Al Mamun, A., Ahmed, M. P., Tisha, S. A., Habib, S. A., & Nitu, F. N. (2025). Enhancing Financial Statement Fraud Detection through Machine Learning: A Comparative Study of Classification Models. *Emerging Frontiers Library for The American Journal of Engineering and Technology*, *7*(09), 166-175.
12. Akhi, S. S., Ahamed, M. I., Alom, M. S., Rakin, A., Awal, A., & Al Mamoon, I. (2025, July). Boosted Forest Soft Ensemble of XGBoost, Gradient Boosting, and Random Forest with Explainable AI for Thyroid Cancer Recurrence Prediction. In *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)* (pp. 1-6). IEEE.
13. Alom, M. S., Akhi, S. S., Borsha, S. N., Mia, N., Tamim, F. S., & Nabin, J. A. (2025, July). Federated Machine Learning for Cardiovascular Risk Assessment: A Decentralized XGBoost Approach. In *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)* (pp. 1-6). IEEE.
14. Nitu, F. N., Mia, M. M., Roy, M. K., Yezdani, S., FINDIK, B., & Nipa, R. A. (2025). Leveraging Graph Neural Networks for Intelligent Supply Chain Risk Management in the Era of Industry 4.0. *International Interdisciplinary Business Economics Advancement Journal*, *6*(10), 21-33.
15. Akhi, S. S., Rahaman, M. A., & Alom, M. S. An Explainable and Robust Machine Learning Approach for Autism Spectrum Disorder Prediction.
16. Rabbi, M. A., Rijon, R. H., Akhi, S. S., Hossain, A., & Jeba, S. M. (2025, January). A Detailed Analysis of Machine Learning Algorithm Performance in Heart Disease Prediction. In *2025 4th International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)* (pp. 259-263). IEEE.
17. Mujiba Shaima, Mazharul Islam Tusher, Estak Ahmed, Sharmin Sultana Akhi, & Rayhan Hassan Mahin. (2025). Machine Learning Techniques and Insights for Cardiovascular or Heart Disease Prediction. *Academic International Journal of Engineering Science*, *3*(01), 22-35.
18. Jamee, S. S., Arif, M., Rahman, M. M., YASSAR, I. S., & Hossain, M. A. (2025). Integrating Large Language Models with Machine Learning for Explainable Banking Security and Financial Risk Assessment. *International Interdisciplinary Business Economics Advancement Journal*, *6*(11), 8-18.
19. Umam, S., & Razzak, R. B. (2024, October). Linguistic disparities in mental health services: Analyzing the impact of spanish language support availability in saint louis region, Missouri. In APHA 2024 Annual Meeting and Expo. APHA.
20. Umam, S., & Razzak, R. B. (2025, November). A 20-Year Overview of Trends in Secondhand Smoke Exposure Among Cardiovascular Disease Patients in the US: 1999–2020. In APHA 2025 Annual Meeting and Expo. APHA.
21. Razzak, R. B., & Umam, S. (2025, November). Health Equity in Action: Utilizing PRECEDE-PROCEED Model to Address Gun Violence and associated PTSD in Shaw Community, Saint Louis, Missouri. In APHA 2025 Annual Meeting and Expo. APHA.
22. Razzak, R. B., & Umam, S. (2025, November). A Place-Based Spatial Analysis of Social Determinants and Opioid Overdose Disparities on Health Outcomes in Illinois, United States. In APHA 2025 Annual Meeting and Expo. APHA.
23. Umam, S., Razzak, R. B., Munni, M. Y., & Rahman, A. (2025). Exploring the non-linear association of daily cigarette consumption behavior and food security-An application of CMP GAM regression. PLoS One, 20(7), e0328109.
24. Estak Ahmed, An Thi Phuong Nguyen, Aleya Akhter, KAMRUN NAHER, & HOSNE ARA MALEK. (2025). Advancing U.S. Healthcare with LLM–Diffusion Hybrid Models for Synthetic Skin Image Generation and Dermatological AI. *Journal of Medical and Health Studies*, *6*(5), 83-90. https://doi.org/10.32996/jmhs.2025.6.5.11
25. Mottalib, M. Y., Nobe, N., Islam, M. T., Hossain, M. R., Jisan, A. H., & Hossen, M. E. (2026). Ensemble Machine Learning and Natural Language Processing for Automated Cancer Indicator Detection in Clinical Notes. *Nvpubhouse Library for International Journal of Medical Science and Public Health Research*, *7*(03), 27-37.
26. Jamee, S. S., YASSAR, I. S., Hossain, M. A., Mia, M. M., & Roy, M. K. (2026). Explainable AI in Banking Compliance: Leveraging Large Language

Models for AML and KYC Decision Support. *Library of Frontline Marketing, Management and Economics Journal*, *6*(01), 06-12.

27. Ayub, M. I., Gharami, A. K., Nitu, F. N., Uddin, M. N., Islam, M. I., Nijhum, A. M., ... & Yezdani, S. (2025). AI-Driven Demand Forecasting for Multi-Echelon Supply Chains: Enhancing Forecasting Accuracy and Operational Efficiency through Machine Learning and Deep Learning Techniques. *Emerging Frontiers Library for The American Journal of Management and Economics Innovations*, *7*(07), 74-85.

28. Mia, M. M., Roy, M. K., YASSAR, I. S., Mottalib, M. Y., Yezdani, S., Nijhum, A. M., ... & Uddin, M. K. (2025). Integrating Blockchain Security and Machine Learning for Fraud Detection in the US Banking System. *Emerging Frontiers Library for The American Journal of Engineering and Technology*, *7*(11), 65-76.

29. Mia, M. M., Rahman, M. M., Sayed, M. A., Nipa, R. A., Dey, S. K., Jahed, K. A., & Mottalib, M. Y. (2026). Enhancing Enterprise Security Management Using Hybrid Machine Learning and Large Language Model–Assisted Intrusion Detection. *Emerging Frontiers Library for The American Journal of Engineering and Technology*, *8*(2), 170-178.

30. Khan, M. S., Gharami, A. K., Nitu, F. N., Uddin, M. N., Ahmed, M., Roy, M. K., & Yezdani, S. (2025). Deep Learning-Driven Customer Segmentation in Banking: A Comparative Analysis for Real-Time Decision Support. *International Interdisciplinary Business Economics Advancement Journal*, *6*(08), 9-22.