A Comparative Study of XGBoost and Random Forest for Predicting Agile Training Sales Success Using Explainable Machine Learning Models

Khalid Ali Alzahrani[1,*], Adel Ali Alzahrani[2]

[1,2]Department of Information Science - College of Arts and Humanities - King Abdulaziz University- Jeddah-Saudi Arabia

ABSTRACT

This study investigates the application of machine learning to predict the success of Agile training sales outcomes, aiming to enhance decision-making and performance management in Agile organizations. Using historical sales and monthly performance data, two predictive algorithms Random Forest and XGBoost were developed and evaluated to classify transactions based on their likelihood of success, represented by the “Won” status. The models incorporated key performance indicators such as Target Achievement Ratio, Sales Target Difference, and Sales Value, combined with categorical attributes like course type and sales category. The results show that both models performed effectively, with XGBoost achieving higher accuracy (87.9 percent) and recall (0.97), demonstrating its superior capability in identifying successful transactions. To enhance interpretability, SHAP (SHapley Additive exPlanations) analysis was used to determine the most influential features affecting predictions. The findings reveal that goal-related performance metrics, particularly Target Achievement Ratio and Sales Target Difference, were the strongest determinants of success, while categorical factors contributed less significantly. These results indicate that consistent goal attainment and effective performance monitoring are central to sales success in Agile training environments. The study contributes to Agile management literature by demonstrating how predictive and explainable artificial intelligence can improve forecasting accuracy, support data-driven decision-making, and promote continuous improvement within Agile business frameworks.

Keywords Agile Management, Machine Learning, Predictive Analytics, Sales Forecasting, Explainable Artificial Intelligence (XAI)

INTRODUCTION

Agile management has evolved into a critical approach for organizations striving to remain competitive in highly dynamic and uncertain business environments [1]. Originating from the Agile Manifesto in software development, the methodology emphasizes adaptability, collaboration, customer-centricity, and continuous improvement. Over the past decade, Agile principles have transcended their technological origins and are now widely applied in sectors such as project management, operations, marketing, and education [2]. As organizations continue to embrace Agile transformation, the need for effective Agile training programs has grown substantially. These programs aim to equip professionals with the knowledge and mindset required to implement Agile frameworks such as Scrum, Kanban, and SAFe. Despite their increasing adoption, managing and predicting the success of Agile training sales remains a significant challenge. Sales performance is often influenced by diverse factors such as target-setting accuracy, course relevance, seasonal trends, customer

Submitted: 1 March 2025
Accepted: 15 April 2025
Published: 1 August 2025

Corresponding author
Khalid Ali Alzahrani,
k.alzahrani@kau.edu.sa

Additional Information and Declarations can be found on page 223

How to cite this article: K. A. Alzahrani and A. A. Alzahrani, “A Comparative Study of XGBoost and Random Forest for Predicting Agile Training Sales Success Using Explainable Machine Learning Models,” *Agile Manag.*, vol. 1, no. 3, pp. 212-225, 2025.

engagement, and the effectiveness of sales strategies. Relying solely on intuition or traditional statistical methods is insufficient to capture the nonlinear and multidimensional patterns that characterize sales dynamics in Agile environments.

Recent advancements in Artificial Intelligence (AI) and data analytics have provided organizations with powerful tools to enhance predictive accuracy and decision-making capabilities [3]. Machine Learning (ML), in particular, has emerged as a reliable method for analyzing complex business data and forecasting future trends. Studies have demonstrated that ML algorithms can outperform traditional statistical techniques by identifying hidden correlations and adapting to changing market behaviors [4]. Ensemble learning methods, such as Random Forest and XGBoost, have gained considerable attention due to their robustness, scalability, and ability to handle nonlinear relationships between predictors and outcomes. These algorithms have been successfully applied in various business domains, including customer churn prediction, financial risk analysis, and demand forecasting [5]. However, while the potential of ML for improving business performance is well established, its application in Agile management contexts remains relatively underexplored. This lack of focused research limits the understanding of how predictive analytics can enhance Agile sales forecasting and decision-making processes.

The state of the art in predictive analytics highlights a growing trend toward integrating AI-driven forecasting within managerial and organizational practices. Studies in sales forecasting have shown that hybrid and ensemble-based models can significantly improve accuracy and provide insights into customer behavior and purchasing patterns. Moreover, the emergence of XAI frameworks, such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME), has advanced the interpretability of machine learning models. These developments have made AI systems more transparent and actionable for managerial decision-making. Nonetheless, most existing research focuses on general business or marketing applications rather than on Agile training environments, where performance metrics are tied closely to iterative goal-setting and adaptive learning. Consequently, there remains a critical gap in understanding how predictive and interpretable AI models can support Agile organizations in optimizing their sales operations and aligning performance outcomes with Agile principles of transparency, adaptability, and continuous improvement.

Addressing this research gap, the present study applies two ensemble machine learning algorithms, Random Forest and XGBoost, to predict the likelihood of success in Agile training sales transactions. The analysis uses historical sales and performance data that include key variables such as Target Achievement Ratio, Sales Target Difference, and Sales Value. These indicators reflect the degree of alignment between sales outcomes and organizational goals, offering valuable insights into factors influencing success. In addition to achieving predictive accuracy, this study incorporates SHAP analysis to explain the reasoning behind model outputs and identify which variables contribute most to successful sales predictions. This dual focus on prediction and interpretability enhances the applicability of AI models in real business settings, ensuring that results are both reliable and explainable to decision-makers.

The contribution of this research is twofold. First, it advances the empirical understanding of how machine learning can be utilized within Agile management to predict and analyze sales success. By comparing Random

Forest and XGBoost models, the study provides evidence of the superior predictive performance of gradient-boosting techniques in handling complex Agile data structures. Second, the integration of explainable AI through SHAP enhances the transparency of model predictions, making them more accessible for practitioners and managers. The findings of this study contribute to the state of the art by bridging the gap between Agile management and AI-based predictive analytics. Furthermore, they provide actionable insights for managers seeking to implement data-driven forecasting systems, improve performance monitoring, and optimize resource allocation. Ultimately, this research demonstrates how the fusion of machine learning and Agile methodologies can enable organizations to achieve greater adaptability, empirical decision-making, and sustained business performance in rapidly evolving markets.

Literature Review and Related Works

Predictive analytics has become an essential component of modern business intelligence, enabling organizations to transform large and complex datasets into actionable insights. Traditional forecasting techniques, including regression and moving averages, have been widely used to estimate sales performance, but these methods often fail to capture nonlinear patterns and multidimensional relationships in dynamic markets [6]. The emergence of ML has addressed many of these limitations by offering flexible models that can adapt to complex data distributions and evolving organizational contexts [7]. ML algorithms have been successfully applied in numerous domains such as demand forecasting, marketing analytics, and financial prediction, where they have consistently demonstrated superior accuracy compared to traditional statistical models [8].

Recent advancements in ensemble learning have further enhanced predictive capabilities. Techniques such as Random Forest and XGBoost combine the strengths of multiple base learners to reduce variance and improve generalization [9]. Studies have shown that ensemble models outperform single classifiers in various sales forecasting scenarios, particularly when dealing with high-dimensional and noisy data [10]. XGBoost, in particular, has gained prominence for its efficiency, scalability, and strong performance in handling structured tabular data, making it suitable for business and financial prediction tasks [11]. Comparative analyses between ensemble models and deep learning architectures have also indicated that, for small to medium-sized structured datasets, tree-based models often achieve better interpretability and performance [12].

While the predictive performance of ML models has been well established, the lack of interpretability remains a significant barrier to their widespread adoption in management and decision-making contexts [13]. The concept of XAI has emerged to address this challenge by providing transparency into how complex models make predictions. Techniques such as SHAP and LIME (Local Interpretable Model-Agnostic Explanations) allow researchers and practitioners to understand the contribution of each feature to the model's output [14]. These methods have been increasingly applied in domains such as finance, healthcare, and marketing, where understanding model rationale is critical for trust and accountability [15]. Among these, SHAP has gained particular attention due to its solid theoretical foundation and ability to deliver consistent and locally accurate explanations [16].

Several studies have demonstrated the value of combining predictive modeling with explainability for business intelligence. Integrating XAI techniques with ML models enables organizations to identify not only which factors drive outcomes but also how these factors interact under different conditions [17]. This is particularly relevant in sales forecasting, where success depends on numerous interacting variables, such as pricing strategy, customer behavior, seasonality, and sales targets. Existing research has shown that interpretable models support better managerial decision-making by aligning algorithmic insights with business intuition and strategy [18]. However, most studies have focused on general sales or marketing applications and have not examined predictive modeling within the Agile management domain. Agile organizations operate under unique principles of iterative planning, rapid adaptation, and empirical feedback, which require predictive tools capable of handling fast-changing performance data [19].

Despite increasing interest in predictive analytics, the integration of machine learning and explainable AI within Agile management contexts remains limited. Current literature tends to emphasize process optimization, sprint efficiency, and team collaboration, with relatively few studies addressing predictive performance measurement or sales forecasting in Agile environments [20]. This research seeks to fill that gap by applying ensemble machine learning models, namely Random Forest and XGBoost, in combination with SHAP analysis, to predict and interpret Agile training sales success. Through this approach, the study contributes to the state of the art by demonstrating how predictive and interpretable AI models can enhance data-driven decision-making, performance monitoring, and continuous improvement within Agile organizations.

Methodology

This study adopted a quantitative research design employing supervised machine learning algorithms to predict the likelihood of successful Agile training sales transactions. The methodological framework consisted of six sequential stages: data acquisition, data preprocessing, feature engineering, model development, model evaluation, and explainability analysis. Each step was systematically designed to ensure data integrity, predictive accuracy, and interpretability. The overall workflow of the research is illustrated in figure 1, which outlines the process from raw data collection to model interpretation. The entire methodology was implemented using Python with the Pandas, Scikit-learn, XGBoost, and SHAP libraries, ensuring computational efficiency and reproducibility in all experiments.

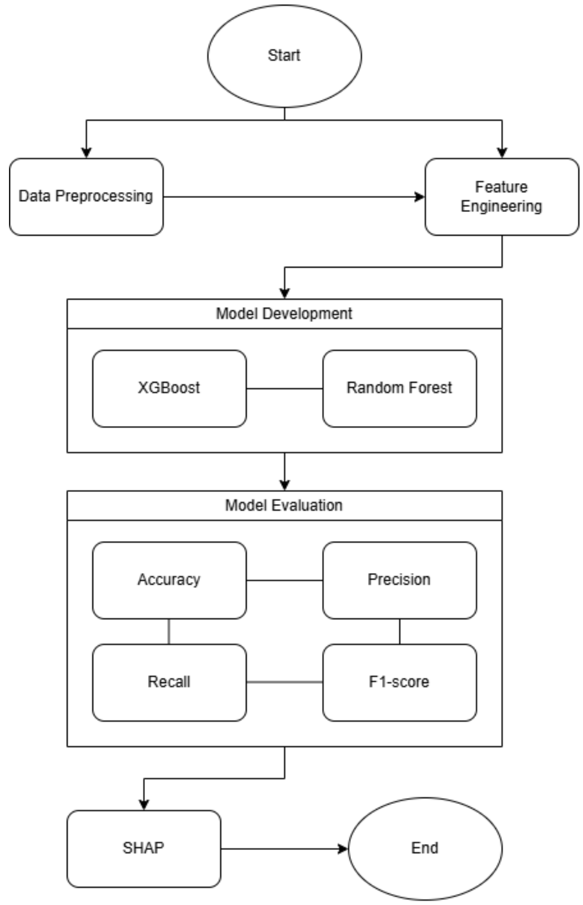

Figure 1 **Research Steps**

The dataset used in this research originated from an Agile training organization that offers professional certification programs such as Certified Scrum Master (CSM), Certified Scrum Product Owner (CSPO), and PMI-ACP. The data were compiled in a multi-sheet Excel file titled Agile Sales Project Dataset, containing transaction records, performance metrics, and sales targets. The Sales Data sheet included detailed attributes such as transaction date, course name, batch number, sales status, number of tickets sold, transaction amount, and category. The Monthly Goals sheet included aggregated indicators such as monthly targets, actual achievements, total payments, and differences between expected and achieved sales. These data provided both transactional (micro-level) and performance (macro-level) views of the organization’s sales operations. The dependent variable (Status) represented the sales outcome, encoded as 1 for successful transactions (“Won”) and 0 for unsuccessful transactions (“Lost”). The independent variables included both numerical attributes (e.g., sales value, target achievement) and categorical features (e.g., course type, category).

Before model training, the raw data underwent several preprocessing steps to ensure quality and consistency. Column names were standardized to lowercase and stripped of whitespace to prevent mismatched references during coding. Irrelevant columns such as comments and order identifiers were excluded. Missing categorical values were replaced with the mode, while missing numerical values were imputed using mean substitution. The date column was converted to a datetime format, from which month and year attributes were extracted to enable temporal analysis. The Sales Data and Monthly Goals sheets were merged using the month attribute as a relational key to align each transaction with its respective monthly performance data. Categorical variables such as Course Type and Category were numerically encoded using the LabelEncoder technique, while continuous features such as Amount were standardized using the z-score normalization method. Outliers were detected using the InterQuartile Range (IQR) approach and adjusted to reduce model distortion.

Feature engineering was conducted to generate derived predictors that enhance the model's ability to identify sales success patterns. Two critical indicators were computed from the Monthly Goals sheet: the Target Achievement Ratio (TAR) and the Sales Target Difference (STD). These were mathematically defined as:

$$
\begin{aligned}
TAR &= \frac{A_{month}}{T_{month}} \\
STD &= A_{month} - T_{month}
\end{aligned}
\tag{1}
$$

A_{month}represents the achieved monthly sales and T_{month}represents the corresponding sales target. The Target Achievement Ratio indicates the extent to which a goal was achieved, while the Sales Target Difference captures the surplus or deficit relative to the target. Together with variables such as Sales Value (in Local Currency), Agile Course Type, and Sales Category, these indicators formed the primary input features for the predictive models. All continuous features were scaled using the StandardScaler method to improve numerical stability and facilitate faster convergence.

The predictive modeling stage involved two supervised ensemble learning algorithms: Random Forest (RF) and Extreme Gradient Boosting (XGBoost). The Random Forest model was configured with 200 decision trees; each trained on bootstrap samples to reduce variance and prevent overfitting. The XGBoost model was trained with 300 estimators, a learning rate of 0.05, and a maximum depth of 5. The dataset was split into training and testing sets using a 75–25 stratified ratio to maintain proportional representation of both "Won" and "Lost" classes. Model training was conducted using the Scikit-learn and XGBoost libraries. Each algorithm produced binary classification outputs representing the probability of a transaction being successful.

Finally, explainability analysis was performed using SHAP to interpret feature influence on the model's decisions. The SHAP values were computed using the TreeExplainer function to quantify each feature's contribution to the predicted outcome. The results showed that Sales Target Difference and Target Achievement Ratio were the most influential predictors of success, followed by Sales Value (in Local Currency). Higher target achievement and positive target differences increased the probability of a "Won" classification, indicating that goal attainment strongly drives sales success in Agile training contexts. Conversely, Course Type and Category were less impactful, suggesting that operational performance metrics outweigh product segmentation in determining outcomes. The inclusion of SHAP ensured that the predictive models were not only accurate but also interpretable, thereby aligning the results with Agile management principles of transparency, empirical decision-making, and continuous learning.

Algorithm 1: Predictive Modeling Framework for Agile Training Sales Success

Input: D_{sales}(Sales transaction data), D_{goals}(Monthly goals data)

Output:

Start

Load $D_{sales} = \{x_i, y_i\}_{i=1}^{N}$and $D_{goals} = \{t_m, a_m\}_{m=1}^{M}$.

Merge both datasets on the month key:

$$D = D_{sales} \cup D_{goals}$$

Handle missing data:

If numeric → replace with mean; If categorical → replace with mode.

Standardize continuous features:

$$x_{ij}^{*} = \frac{x_{ij} - \mu_j}{\sigma_j}$$

Compute engineered features: TAR and STD.

$$TAR = \frac{A_{month}}{T_{month}}$$
$$STD = A_{month} - T_{month}$$

Split dataset into training and testing subsets:

$$(X_{train}, Y_{train}), (X_{test}, Y_{test}) = Split(X, Y, 0.75)$$

Train models:

Random Forest:

$$RF.fit(X_{train}, Y_{train})$$

XGBoost:

$$XGB.fit(X_{train}, Y_{train})$$

Predict outcomes:

$$\hat{Y}_{RF} = RF.predict(X_{test}), \hat{Y}_{XGB} = XGB.predict(X_{test})$$

Evaluate performance metrics:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
$$Precision = \frac{TP}{TP + FP}$$
$$Recall = \frac{TP}{TP + FN}$$
$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Select best model:

$$M^{*} = \arg\max\,(Accuracy(M))$$

Compute explainability using SHAP:

$$S_{SHAP} = SHAP(M^{*}, X_{test})$$

Identify top contributing features:

$TopFeatures =$ {Sales Target Difference,Target Achievement Ratio,Sales Value}

End

Result

The predictive models developed in this study were designed to assess the probability of achieving successful Agile training sales outcomes, which were defined by the “Won” status in the transaction records. Two machine learning algorithms, Random Forest and XGBoost, were applied to the dataset using input features derived from historical sales transactions and monthly goal performance metrics. The selected features included key indicators such as the Target Achievement Ratio and Sales Target Difference, which reflect both the degree of goal attainment and the deviation between actual and expected performance. These variables were combined with other sales characteristics, such as transaction amount, course type, and category, to create a comprehensive dataset that captures different aspects of Agile training sales performance. This integration of quantitative and categorical data allowed the models to learn complex patterns underlying successful transactions and provided a data-driven approach to understanding the factors that contribute to sales success in Agile training contexts.

The evaluation results indicate that both algorithms performed effectively in predicting successful sales transactions. The Random Forest model achieved an accuracy of 86.2 percent, while the XGBoost model demonstrated slightly higher performance with an accuracy of 87.9 percent. In terms of classification metrics, XGBoost recorded the highest recall of 0.97 and an F1-score of 0.93 for identifying successful outcomes. These values suggest that the model was able to correctly identify the majority of “Won” cases while maintaining a strong balance between precision and recall. Although both models exhibited relatively lower performance in predicting the minority “Lost” class, this can be attributed to the class imbalance present in the dataset, where successful transactions significantly outnumbered unsuccessful ones. Overall, the findings demonstrate that both models are reliable tools for forecasting Agile training sales performance, with XGBoost showing superior sensitivity and predictive stability. Table 1 presents a comparison of the primary performance metrics for both algorithms.

Table 1 Model Performance Comparison

Model	Accuracy	Precision (Class 1)	Recall (Class 1)	F1-Score (Class 1)
Random Forest	0.862	0.90	0.94	0.92
XGBoost	0.879	0.90	0.97	0.93

The confusion matrices presented in figure 2 and figure 3 provide a more detailed illustration of how each model classified the sales transactions. The Random Forest model achieved strong overall accuracy, correctly identifying 343 of the transactions that were successfully closed as “Won.” However, it also misclassified 21 of these successful cases as “Lost,” indicating a small number of false negatives. This misclassification suggests that, while the model effectively captured the majority of successful outcomes, it struggled to identify a limited subset of transactions that shared characteristics with unsuccessful sales. These results imply that certain patterns in the data, such as low sales value or marginal goal achievement, may have contributed to the confusion in the model’s decision-making process.

In contrast, the XGBoost model demonstrated a higher level of sensitivity and precision in identifying successful Agile training sales. It correctly classified 353 successful transactions and reduced the number of false negatives to only 11 cases. This improvement indicates that XGBoost was more capable of distinguishing between the subtle differences in features that separate successful from unsuccessful sales. The algorithm's gradient boosting approach, which focuses on minimizing residual errors in each iteration, likely contributed to this enhanced predictive performance. As a result, XGBoost not only achieved higher accuracy but also showed superior consistency in correctly recognizing "Won" transactions, making it a more dependable tool for sales prediction in Agile training contexts.

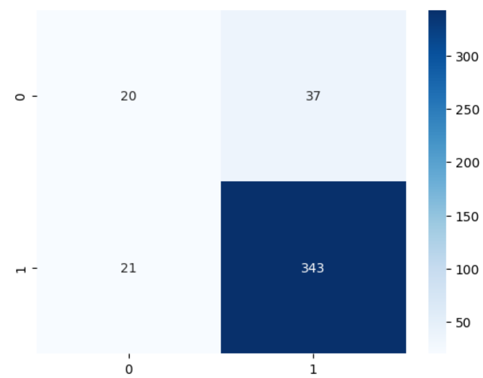

Figure 2 Confusion Matrix Random Forest

Despite their overall strong performance, both models exhibited limited capability in detecting unsuccessful sales, or "Lost" cases. The lower recall and precision values for this class suggest that the models were influenced by the imbalance in the dataset, where successful sales significantly outnumbered unsuccessful ones. This imbalance can lead to a bias in prediction, as the algorithms tend to favor the dominant class. To address this limitation, future research could apply data balancing techniques such as Synthetic Minority Oversampling (SMOTE) or cost-sensitive learning. Additionally, incorporating more granular features—such as lead source quality, customer engagement level, or marketing campaign data—may enhance the models' ability to distinguish potential sales failures more accurately. Improving the prediction of unsuccessful outcomes would provide organizations with more actionable insights for risk mitigation and targeted intervention in Agile sales management.

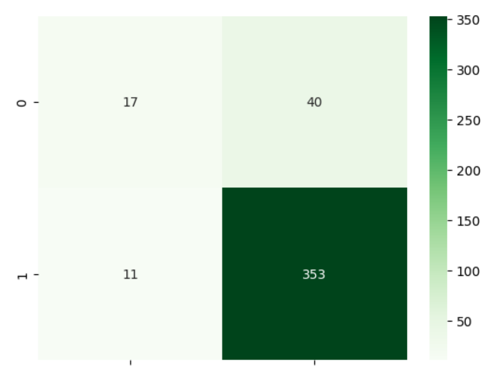

Figure 3 Confusion Matrix XGBoost

To enhance interpretability and understand the underlying mechanisms driving the model's predictions, the XGBoost algorithm was analyzed using SHAP. This method provides insight into how individual features contribute to the model's decision-making process by quantifying their marginal impact on the predicted outcome. The SHAP summary plot, presented in figure 4, highlights the features that exerted the greatest influence on sales success predictions. Among these, Sales Target Difference and Target Achievement Ratio emerged as the two most critical predictors. A larger positive difference between the achieved and expected sales targets was associated with a higher probability of a successful transaction, indicating that performance exceeding targets strongly drives sales success. Similarly, a higher target achievement ratio was consistently linked to a "Won" classification, reflecting that consistent goal attainment is a key determinant of favorable sales outcomes within Agile training contexts.

In addition to the performance-based indicators, the Sales Value (in Local Currency) was also identified as an important contributor to model predictions. This finding suggests that transactions involving higher monetary values are more likely to result in successful outcomes, possibly due to stronger client commitment or higher perceived value of premium Agile training programs. In contrast, categorical variables such as Agile Course Type and Sales Category exhibited lower SHAP values, indicating a relatively limited effect on sales success compared to quantitative goal-oriented features. This suggests that while course type and category may influence sales to some extent, they do not substantially determine the likelihood of achieving a sale. Overall, the SHAP analysis underscores the significance of performance and goal-related metrics as the primary drivers of success, highlighting the importance of maintaining high achievement rates and effective target management strategies within Agile sales operations.

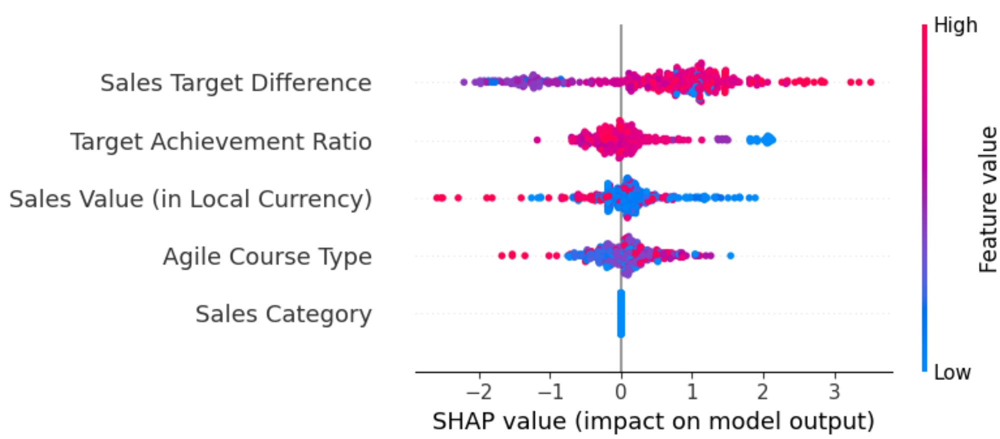

Figure 4 SHAP Summary Plot of Feature Importance

Overall, the results demonstrate that machine learning can effectively predict the likelihood of successful Agile training sales using historical performance data. Among the two algorithms tested, XGBoost provided the most accurate and interpretable predictions. The explainable AI analysis further confirms that organizational performance metrics, such as target achievement and sales difference, are the most critical factors influencing sales outcomes. These findings emphasize the value of adopting data-driven strategies in Agile management and sales planning, enabling organizations to forecast outcomes more precisely and allocate resources more efficiently.

Discussion

The results of this study demonstrate the potential of machine learning techniques to improve predictive decision-making in Agile training sales management. Both the Random Forest and XGBoost models successfully predicted sales outcomes, but the XGBoost model achieved higher accuracy, recall, and F1-scores, indicating a stronger ability to identify successful transactions. This finding is consistent with previous research that emphasizes the superior performance of boosting algorithms in managing complex and nonlinear data structures. The XGBoost model's iterative learning process allows it to adjust to residual errors, which enhances its predictive precision and adaptability. These characteristics are particularly aligned with Agile principles, where continuous learning and iterative improvement are central to achieving performance excellence. The use of such models provides Agile organizations with analytical tools that can transform raw sales data into meaningful insights, enabling more accurate forecasting and evidence-based decision-making. Furthermore, the integration of machine learning into Agile operations supports the concept of empirical process control, as it enables organizations to make informed adjustments based on objective performance indicators rather than subjective judgments.

The interpretability analysis using SHAP revealed that quantitative metrics related to performance and goal attainment play a critical role in determining sales success. The Sales Target Difference and Target Achievement Ratio were identified as the most influential features, emphasizing that organizations that consistently meet or exceed their targets are more likely to achieve positive sales outcomes. The importance of these variables indicates that structured goal management and consistent monitoring are fundamental drivers of success in Agile sales environments. The Sales Value (in Local Currency) was also found to have a positive impact, suggesting that higher-value transactions may be associated with stronger client commitment and better resource alignment. In contrast, qualitative features such as Agile Course Type and Sales Category contributed less to the predictive performance, implying that product differentiation alone is insufficient to ensure sales success. Instead, success is more strongly influenced by how effectively sales teams execute strategies and achieve objectives. From a managerial perspective, these insights highlight the need for Agile organizations to strengthen their data-driven practices by focusing on measurable performance indicators. By leveraging explainable machine learning models such as XGBoost with SHAP analysis, sales managers can identify key success drivers, optimize resource allocation, and continuously refine their strategies in line with Agile values of adaptability, transparency, and empirical improvement.

Conclusion

This study explored the application of machine learning to predict the success of Agile training sales outcomes and demonstrated that data-driven models can significantly enhance decision-making accuracy within Agile management contexts. Using Random Forest and XGBoost algorithms, the research revealed that both models performed effectively, with XGBoost achieving slightly higher accuracy and recall in identifying successful transactions. The results showed that predictive analytics can help organizations anticipate performance trends and allocate resources more efficiently, supporting Agile principles of continuous improvement and empirical process control. The SHAP analysis further enhanced interpretability by identifying the most influential predictors of sales success, namely the Sales Target Difference and Target Achievement Ratio, which underscore the importance of consistent goal attainment and exceeding sales expectations. Additionally, higher sales values were found to positively influence success probabilities, whereas categorical variables such as Agile Course Type and Sales Category had comparatively minor impacts. These insights indicate that performance management and execution effectiveness play a greater role in determining success than product differentiation. From a managerial perspective, the findings highlight the potential of integrating explainable artificial intelligence tools within Agile organizations to foster transparency, optimize target-setting, and refine strategic decision-making. Future research should focus on addressing data imbalance issues and incorporating broader contextual variables, such as customer engagement or marketing effectiveness, to improve model robustness and generalizability. Overall, the study contributes to the growing body of knowledge linking artificial intelligence and Agile methodologies, demonstrating that predictive and interpretable analytics can serve as valuable enablers for achieving adaptability, precision, and sustained performance in dynamic business environments.

Declarations

Author Contributions

Conceptualization: K.A.A. and A.A.A.; Methodology: A.A.A.; Software: K.A.A.; Validation: K.A.A. and A.A.A.; Formal Analysis: K.A.A. and A.A.A.; Investigation: K.A.A.; Resources: A.A.A.; Data Curation: A.A.A.; Writing Original Draft Preparation: K.A.A. and A.A.A.; Writing Review and Editing: A.A.A. and K.A.A.; Visualization: K.A.A.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, vol. 16, no. 8, pp. 785–794, 2016, doi: 10.1145/2939672.2939785

[2] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324

[3] Y. Ensafi, S. H. Amin, G. Zhang, and B. Shah, "Time-series forecasting of seasonal items sales using machine learning–A comparative analysis," *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 1, pp. 1–10, 2022, doi: 10.1016/j.jjimei.2022.100058

[4] A. Mitra, A. Jain, A. Kishore, and P. Kumar, "A comparative study of demand forecasting models for a multi-channel retail company: A novel hybrid machine learning approach," *Oper. Res. Forum*, vol. 3, no. 1, pp. 1–12, 2022, doi: 10.1007/s43069-022-00166-4

[5] M. Zubair, A. Waleed, and A. Rehman, "Machine Learning Insights into Retail Sales Prediction: A Comparative Analysis of Algorithms," in *Horizons of Artificial Intelligence Research*, vol. 2, no. 3, pp. 95–112, 2024, doi: 10.1109/HITE63532.2024.10777132

[6] S. Ganguly and I. Mukherjee, "Enhancing Retail Sales Forecasting with Optimized Machine Learning Models," *arXiv preprint*, vol. 2024, no. October, pp. 1-5, arXiv:2410.13773, Oct. 2024. Doi: 10.48550/arXiv.2410.13773

[7] A. Sharma, A. Singh, and R. Kumar, "Explainable AI for Forensic Analysis: A Comparative Study of SHAP and LIME in Intrusion Detection Models," *Appl. Sci.*, vol. 15, no. 13, pp. 1–18, 2025, Art. no. 7329, doi: 10.3390/app15137329

[8] L. Hobor, M. Brcic, L. Polutnik, and A. Kapetanovic, "Comparative Analysis of Modern Machine Learning Models for Retail Sales Forecasting," *arXiv preprint*, vol. 2025, no. June, pp. 1-20, arXiv:2506.05941, Jun. 2025. Doi: 10.48550/arXiv.2506.05941

[9] D. Gunning and D. W. Aha, "DARPA's Explainable Artificial Intelligence (XAI) Program," *AI Mag.*, vol. 40, no. 2, pp. 44–58, 2019, doi: 10.1609/aimag.v40i2.2850

[10] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 4, pp. e1249.1–e1249.18, 2018, doi: 10.1002/widm.1249.

[11] M. Ibrahim, K. A. Bakar, M. A. Mohamed, and R. Ahmad, "Retail Sales Forecasting Using Machine Learning Algorithms: A Comparative Study," *IEEE Access*, vol. 9, no. 1, pp. 137104–137117, 2021, doi: 10.47672/ajsas.2679.

[12] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A Review of Machine Learning Interpretability Methods," *Entropy*, vol. 23, no. 1, pp. 1–38, Jan. 2021, doi: 10.3390/e23010018.

[13] C. Zhan, X. Zhang, and Z. Li, "E-commerce Sales Forecast Based on Ensemble Learning," in *Proc. IEEE Int. Symp. Product Compliance Engineering - Asia (ISPCE-CN)*, vol. 2020, no. January, pp. 1–5, 2020, doi: 10.1109/ISPCE-CN51288.2020.9321858.

[14] J. Wang, "A Hybrid Machine Learning Model for Sales Prediction," in *Proc. Int. Conf. Intelligent Computing and Human-Computer Interaction (ICHCI)*, 2020, pp. 102–107, doi: 10.1109/ICHCI51889.2020.00083.

[15] M. Gurnani, V. Shah, and K. Raval, "Forecasting Sales Using Fusion of Machine Learning Techniques," in *Proc. Int. Conf. Data Management, Analytics and Innovation (ICDMAI)*, 2017, pp. 45–53, doi: 10.1109/ICDMAI.2017.8073492.

[16] A. Akanksha, S. Rani, and S. Mittal, "Store Sales Forecasting Model to Determine Inventory Stock Levels Using Machine Learning," in *Proc. Int. Conf. Inventive Computation Technologies (ICICT)*, 2022, pp. 882–887, doi: 10.1109/ICICT54344.2022.9850468.

[17] Y. Niu, "Walmart Sales Forecasting using XGBoost Algorithm and Feature Engineering," in *Proc. Int. Conf. Big Data & Artificial Intelligence & Software Engineering (ICBASE)*, 2020, pp. 328–334, doi: 10.1109/ICBASE51474.2020.00103.

[18] F. Almalki and M. Masud, "Financial Fraud Detection Using Explainable AI and Stacking Ensemble Methods," *arXiv preprint*, vol. 2025, no. May, pp. 1-30, arXiv:2505.10050, May 2025. Doi: 10.48550/arXiv.2505.10050

[19] R. Velmurugan, D. P. Muthukumaran, and A. Kannan, "Evaluating Explainable Methods for Predictive Process Analytics: A Functionally-Grounded Approach," *arXiv preprint*, vol. 2020, no. December, pp.1-15, arXiv:2012.04218, Dec. 2020. Doi: 10.48550/arXiv.2012.04218

[20] A. K. Sharma, R. Singh, and M. A. Khan, "A Machine Learning Approach for Sales Forecasting in Retail Supply Chain," *International Journal of Information Management Data Insights*, vol. 2, no. 2, pp. 100095, 2022, doi: 10.1016/j.jjimei.2022.100095.