SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 4, (2026), pp. 9718-9728
Open Access. Online & Print

DOI: 10.5281/zenodo.20160179

A MACHINE LEARNING–DRIVEN ANALYTICAL FRAMEWORK FOR INTELLIGENT DECISION SUPPORT IN COMMERCIAL MANAGEMENT SYSTEMS

[1*]Dr. Aman Kumar, [2]Sandeep Rajani, [3]Dr. Suddapalli Suman Babu, [4]Dr. Sobiya, [5]Dr. Sumesh Ramankutty, [6]Shankar Prasad Mitra

*[1]*Associate professor, Apex institute of management, Chandigarh University , Mohali-140413, Punjab, India, Email: athmcch@gmail.com, Orcid Id: http://orcid.org/0009-0001-3468-4856*
[2]Research Scholar, Department of Computer Science, Sai Nath University, Ranchi, Jharkhand, India, Email Id: s_rajani_in@yahoo.com, Orcid Id: https://orcid.org/0000-0002-7050-3979
[3]Professor, School of Business, Specialization in OB & HR, Woxsen University, Hyderabad - 502345, Telangana, India, Email Id: suman626@gmail.com, Orcid Id: https://orcid.org/0000-0001-6420-7455
[4]Assistant Professor, Department of commerce, Kalasalingam Academy of Research and Education, Krishnankoil, Tamil Nadu, India, Email Id: sobiya16oct@gmail.com
[5]Independent Researcher, Specialisatiion in Digital Marketing, VIT Alumini, VIT University, Maharashtra, Mumbai, India, Email Id: ramansumesh15@gmail.com, Orcid Id:0000-0002-2886-7736
[6]Assistant Professor, Department of Computer Science & Engineering, Specialization in AI, Brainware University, Barasat, Kolkata – 125, West Bengal-700125, India, Email Id: spmitra2016@gmail.com, Orcid Id: 0009-0003-5474-5534

***Received:* 29/12/2025**
***Accepted:* 26/03/2026**

Corresponding Author: Aman Kumar (athmcch@gmail.com)

Abstract

The rapid growth of digital commerce has generated large volumes of transactional and customer behavioural data, creating opportunities for intelligent decision-making through advanced analytical techniques. This study proposes a machine learning–driven analytical framework for intelligent decision support in commercial management systems with the objective of predicting transaction values and customer retention behaviour. A commercial transaction dataset containing customer demographic information, purchasing patterns, and interaction features was utilized for model development and evaluation. Data preprocessing techniques, including data cleaning, feature transformation, one-hot encoding, and feature scaling, were applied to prepare the dataset for machine learning analysis. The dataset was divided into training and testing sets using an 80:20 split to ensure reliable model evaluation. Both regression and classification models were implemented, including baseline models, traditional statistical models, and a proposed hybrid ensemble approach that integrates Random Forest and Gradient Boosting algorithms. Model performance was assessed using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2) for regression analysis, while classification performance was evaluated using accuracy, precision, recall, F1-score, and confusion matrix analysis. The experimental results demonstrate that the proposed hybrid ensemble model significantly improves regression performance by reducing prediction errors and achieving higher explanatory power compared to baseline and traditional models. However, the classification results indicate similar performance across all models due to the presence of class imbalance within the dataset. Overall, the findings highlight the effectiveness of machine learning–based analytical frameworks in enhancing predictive analytics and supporting data-driven decision-making in commercial management environments.

Keywords: *Machine Learning, Decision Support Systems, Commercial Analytics, Hybrid Ensemble Learning, Predictive Modelling*

1. Introduction

The invention of digital technologies, statistical proliferation of data-driven systems have significantly transformed the contemporary management activity of commercial organizations. Companies are increasingly relying on an enormous amount of transactional, behavioural and operational data in order to support the decision making process and to enhance business performance. Conventional decision-support systems are likely to utilize rule based or manual analysis, which might not be effective in addressing complex data and dynamic business environments. As a result, the evolution of machine learning and analytic data framework have emerged as a potent instrument for extracting insights out of large datasets and supporting smart decision making in business systems.

The machine learning methods may assist organizations to examine trends in complicated datasets and develop predictive models that could be utilized to predict patterns, anticipate customer behaviour and optimize operational methods. Recent research has shown the effectiveness of machine learning-based analytical frameworks in enhancing the decision making process in various domains. For example, Ni et al. have proposed an analytical framework for using machine learning for modelling livestock behaviour and health monitoring using wearable sensor systems and the role of data-driven methods for automated decision support systems [1]. Similarly, Pourramezani et al. came up with a data-driven framework for the management of cyber-physical-social distribution networks, where the importance of sophisticated data analytics in complex infrastructure systems [2] was highlighted. These studies illustrate the increasing role that machine learning plays in the design of intelligent systems that are capable of processing large-scale amounts of data and justifying real-time decisions.

Artificial intelligence and machine learning technologies have been widely used in energy management and smart grid systems as well. Banad et al. discussed the role of artificial intelligence and machine learning in smart grid applications and the integration of digital twin technologies and intelligent data analytics to boost the decision-making capabilities [3]. Furthermore, Al-Khamees et al. reviewed intelligent systems and their applications in the decision support environment and highlighted the fact that the use of advanced machine learning models for tackling complex decision-making challenges in different industries is growing [4]. This trend indicates that the intelligence of analytical procedures in decision support systems should be integrated to make the decision systems more efficient in operations.

Machine learning models have been broadly used in industrial and technological environments to predict and monitor performance in the field of predictive monitoring. Bhoi et al. suggested a data-driven power electronics system monitoring based on edge-cloud computing and machine learning algorithms to enhance the reliability of the system and predictive maintenance [5]. In the same way, Murugan proposed a machine-learning-based framework of financial risk management and analysis that demonstrates the way predictive analytics can be applied to enhance decision making in financial systems [6]. Agarwal et al. also presented the neural network methods of the financial risk prediction procedure, advantages of using machine learning methods in processing financial complex data [7].

Sustainability and urban management applications have also been combined in machine learning. Li et al. [8] created a framework of machine learning-based and remote sensing-based sustainable urban development and environmental monitoring. In the energy sector, digital twin technologies in combination with machine learning algorithms have been applied to enhance the optimization of energy systems and infrastructure management. Das et al. worked on the use of digital twins and machine learning in the optimization of renewable energy systems and the smart grid [9], while Li et al. proposed an analytical framework for digital twin-based engineering asset management and supporting predictive maintenance and infrastructure planning [10].

In addition to infrastructure and energy systems, machine learning frameworks have been used to provide supply chain management and industrial operations. Badakhshan et al. suggested a hybrid framework with simulation and machine learning techniques for better decision-making in supply chain systems [11]. Similarly, an IoT-enabled smart energy management system that relies on forecasting algorithms and optimization strategies for improving renewable energy generation was proposed by Rao et al. [12]. Furthermore, a review of machine learning and deep learning applications to load forecasting in smart grid system (Biswal et al) [13] presented the significance of predictive analytics in the improvement of energy management strategies.

Despite the fact that we are witnessing the growing use of machine learning in various sectors, machine learning application in commercial management systems still needs to be further explored. Commercial datasets are often complex, and can have many interdependent relationships between customer behavior, transaction patterns, and operational variables that cannot be captured as well by traditional statistical models. Therefore, advanced analytical frameworks, which combine multiple machine learning techniques, are needed to improve the predictive performance to support intelligent decision-making in the commercial scenarios.

To mitigate the above challenges, a machine learning based analytical framework for intelligent decision support for commercial management systems is proposed in this research work. The suggested framework combines the regression and classification models to analyze the commercial transaction data and derive the predictive insights concerning the revenue prediction and customer retention behavior. The framework will improve the predictive accuracy and will help in making decisions which are data based in the existing commercial systems since the procedure will incorporate the ensemble learning techniques together with the large scale data preprocessing and evaluation schemes.

Objective of the study

1. To develop a machine learning–based analytical framework for intelligent decision support in commercial management systems.
2. To predict transaction value and customer retention using regression and classification models.
3. To compare baseline, traditional, and hybrid ensemble models using performance evaluation metrics.

2. Methodology

This part describes the methodological framework that was employed in the construction and testing of the proposed machine learning powered analytical framework to support intelligent decision making in business management systems [14]. The methodology will be dedicated to predicting the outcomes of commercial transactions and retention behaviour of customers on the basis of machine learning methods. The proposed framework is a combination of data preprocessing, model development, training and testing processes, performance evaluation mechanisms of analyzing commercial transaction data and data-driven decision making processes.

2.1 Dataset Description

The experiments have been performed with a commercial transaction dataset that includes customer behaviour, transaction and interaction data. The dataset contains a number of attributes relating to customer demographics and purchase characteristics, as well as patterns of website interaction. Such features give valuable insights to predict transaction values as well as customer retention behaviour in commercial systems. The attributes that are present in the dataset are customer age, gender, city, product category, unit price, purchase quantity, discount amount, payment method, device type, session duration, page’s view, delivery time, and customer rating. In addition to these attributes, the dataset has the total transaction amount and an indicator indicating whether a customer is a returning customer. These variables collectively represent the demographic information, purchasing patterns, browsing behaviour and service-related factors which influence commercial transactions. The dataset contains both numerical and categorical variables, thus making it suitable for machine learning-based predictive analytics. Such features make it possible to develop intelligent decision support models that could detect the patterns in customer transactions and predict future behaviour.

2.2 Data Preprocessing

Before we applied machine learning models, the dataset was subjected to several preprocessing steps to improve the quality of the data and ensure that it is compatible with machine learning algorithms. Initially, it was found that identifier attributes such as Order ID and Customer_ID were removed from the modelling process, as these variables don't add meaningful predictive information and can add noise to the learning process.

The Date attribute was converted to structured temporal features in order to capture possible seasonal patterns for customer purchasing behaviour. Handling Missing values was an important preprocessing step, and missing numerical values were treated using median-based imputation, and categorical variables were filled using the most frequent category.

Since the dataset has some categorical variables like gender, city, product category, payment method, and device type, these variables were converted into numerical format using the one-hot encoding technique. This change enables the machine learning algorithms to treat categorical data. Moreover, numerical characteristics were also normalized using feature scaling to make sure that not all of the variables that have varying scopes do not disproportionately affect the process of learning.

2.3 Experimental Design

To verify the predictive ability of the suggested analytical framework, the dataset was divided into a training and a testing data set. The machine learning models were trained in the training part and their predictive capabilities on unknown data on known data were tested in the testing part. The current study used an 80:20 train-test split to make sure that there is a balance between the training of the model and testing the performance of the model.

This experimental design ensures that the models are under testing in a real life setting and also gives a valid indication of how the models would fare in new commercial transaction data.

2.4 Model Development

The analytical framework as suggested incorporates two predictive modules which are to be applied during the decision-making processes in commercial systems. The first module is the regression analysis for predicting the total transaction value, and the second module uses classification analysis for predicting customer retention behaviour.

For the regression task, three models were implemented for the purpose of evaluating predictive performance. The baseline model is a mean predictor that estimates transaction values using an average of the target variable. The traditional model uses linear regression, which is a frequently used statistical technique in predictive analytics. The proposed model makes use of a hybrid ensemble regression model that combines the Random Forest algorithm and the Gradient Boosting algorithm using an ensemble voting mechanism. This approach of hybridisation allows us to take advantage of the best of both worlds in order to capture complex nonlinear relationships within commercial datasets for better predictive accuracy.

For the classification task, the same comparative modelling strategy was taken. The baseline model predicts the majority class in the dataset and is the minimal benchmark for the classification performance. The traditional model of classification uses logistic regression and is widely used in the field of binary classification problems. The classification model proposed is based on a hybrid ensemble classifier incorporating Random Forest and Gradient Boosting algorithm to enhance the robustness of the classification and increase the detection of customer retention patterns.

2.5 Performance Evaluation Metrics

In order to test the efficacy of the proposed analytical framework, several performance evaluation metrics were employed in regression and classification tasks.

For regression analysis, the model performance was evaluated in terms of Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R2). MAE is a measure of the average magnitude of the prediction errors, while RMSE is a measure of the square root of the average square Prediction errors, giving more weight to the larger errors. R2 is an indicator of the percentage of the variation in the target variable that can be described by the model and also shows how well such a model is.

To analyze the models in terms of accuracy, precision, recall and F1-score, the models were evaluated. The ratio of the number of correct instances to the number of instances predicted is the accuracy. Precision: The proportion of the correct positive examples that are so predicted to the number of positive examples that are predicted; recall: The proportion of the correct positive examples that are actually found to the number of actual positive examples. The harmonic mean of the precision and the recall measure is the F1-score, a balanced score of the performance of classification. Besides these numerical measures, the confusion matrix analysis was conducted to verify the performance of the models in classification. The confusion matrix is a comprehensive table of results of the classification since it exhibits the true positive, true negative, false positive, and false negative. Such analysis allows identifying the misclassification patterns and provides a greater understanding of the performance of each model on the different classes. Confusion matrices of the standard classifier, the traditional logistic regression model and the proposed hybrid ensemble classifier are obtained.

2.6 Experimental Implementation

All experiments were carried out in the Python programming language in a computational environment over the cloud. Data preprocessing and analysis were carried out using the Pandas and NumPy libraries, while machine learning models were implemented using the Scikit-learn framework. Visualization of the result of the experiment, such as a regression comparison plot, actual vs. predicted plot, residual plot and confusion matrix has been performed using the Matplotlib library. This computational configuration offers a flexible and reproducible environment for implementing machine learning algorithms and assessing the performance of these algorithms in decision-support applications where commercial use is concerned.

3. Results

This section presents the experimental results obtained from the proposed machine learning driven analytical framework developed for intelligent decision support in commercial management systems. The evaluation is based on two predictive tasks: revenue prediction by regression analysis and customer retention prediction by classification analysis. The performance of the proposed hybrid ensemble framework is compared with the performance of baseline and traditional machine learning models using multiple evaluation metrics.

3.1 Regression Model Performance

The regression experiment is aimed at predicting the total transaction amount (Total_Amount), which is extremely critical in the commercial decision-making process, such as revenue forecasting, pricing and demand analysis. To examine the applicability of the proposed analytical framework, three models were taken into consideration:

- Baseline Model: Mean predictor
- Traditional Model: Linear Regression
- Proposed Model: Hybrid Ensemble combining Random Forest and Gradient Boosting

The performance of these models was evaluated using three widely used regression metrics:

- Mean Absolute Error (MAE)
- Root Mean Square Error (RMSE)
- Coefficient of Determination (R^2)

The comparative performance of the three regression models is illustrated in Figure 1. As shown in Figure 1, the proposed hybrid ensemble model is much better than the baseline and traditional models. The hybrid model has the lowest values of MAE and RMSE, which indicates that the prediction errors between actual and predicted amounts for transactions are significantly reduced. In contrast, the baseline model has the highest values of error, reflecting the low predictive capability of the model. Furthermore, the R2 value of the hybrid ensemble model is very close to 1.0, which shows that the proposed framework is successful in capturing a large part of the variance of the target variable. As we can see, the linear regression model has moderate performance in predicting, and the baseline model cannot capture meaningful relationships within the dataset. These results show that the use of multiple machine learning algorithms organized in an ensemble framework can improve predictive accuracy and can give a more reliable analytical model for commercial decision-support systems.

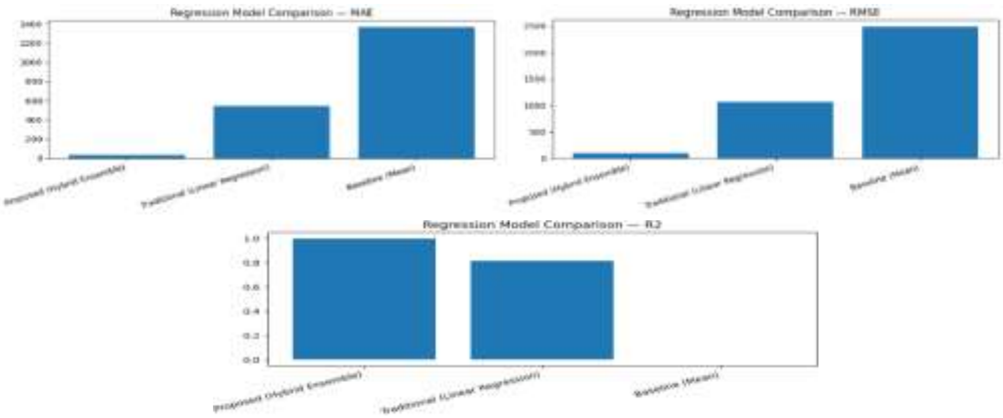

Figure 1. Regression model comparison based on MAE, RMSE, and R^2 metrics for the baseline model, traditional linear regression model, and the proposed hybrid ensemble model.

3.2 Actual vs Predicted Value Analysis

In order to further assess the predictive power of the proposed regression model, the relationship between the real and the predicted values of the transaction was examined. The scatter plot shown in Figure 2 shows the relationship between the actual number of transactions and the predicted values produced by the proposed hybrid ensemble model. As seen from the figure, most of the data points are near the diagonal reference line, which means that the predicted values match well with the actual values. This good correspondence gives the high power of prediction of the proposed model. The results further prove that the hybrid ensemble approach is effective in capturing complex nonlinear relationships in the commercial dataset.

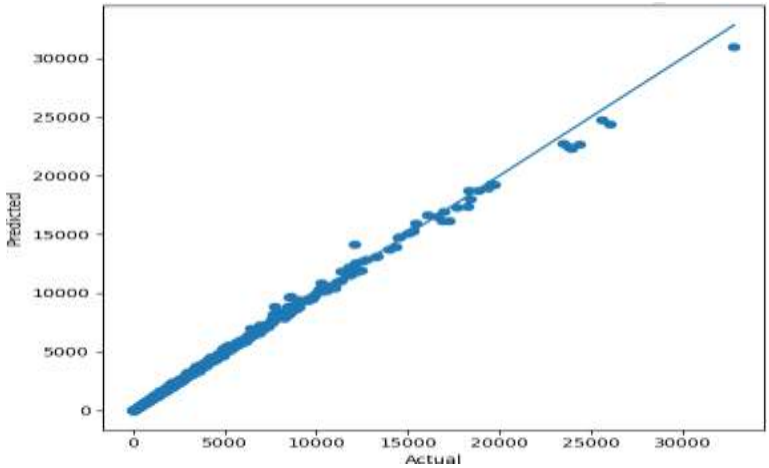

Figure 2. Actual versus predicted transaction amounts (Total_Amount) using the proposed hybrid ensemble model.

3.3 Residual Analysis

The remaining analysis was made to investigate how the prediction errors are distributed and to examine whether the proposed regression model is stable. The distribution of the residuals against the predicted values is given in Figure 3. The residual plot shows that most of the residual values are clustered around zero with no pronounced systematic structure. This implies that there is no significant bias in the proposed model and the model shows stable predictions with various transaction values. The more or less homogenous spread of the residuals is another argument that supports the robustness of the proposed analytical framework.

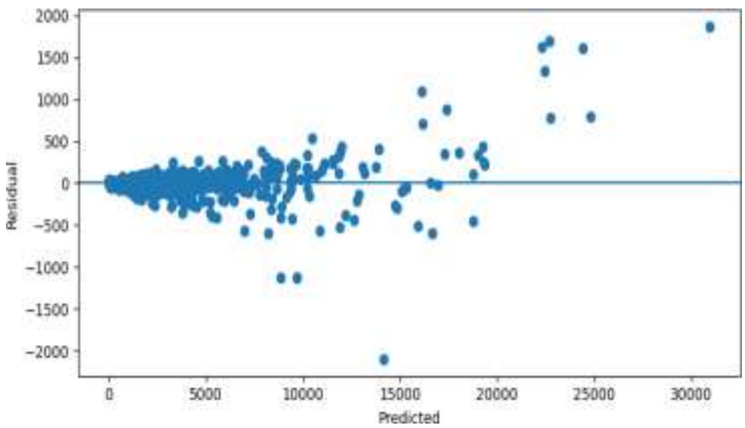

Figure 3. Residual distribution of the proposed hybrid ensemble regression model.

3.4 Classification Performance Evaluation

Apart from revenue prediction, the proposed framework was also tested for customer retention prediction. The classification task is to predict if a customer will return (Is Returning Customer), which is an important factor in commercial decision-making processes such as marketing campaign optimization and customer loyalty management.

Three classification models were evaluated:

- Baseline Model: Majority class predictor
- Traditional Model: Logistic Regression
- Proposed Model: Hybrid Ensemble classifier combining Random Forest and Gradient Boosting

The classification performance was evaluated using the following metrics:

- Accuracy
- Precision
- Recall
- F1-score

The results of the classification are summarized in Table 1. The results show that all models reached the same classification performance in all evaluation measures. This result implies a dominant class distribution in the dataset, which has a major impact on the predictive behaviour of the models.

Table 1. Classification performance comparison of baseline, traditional, and proposed hybrid ensemble models.

Model	Accuracy	Precision	Recall	F1-score
Baseline (Majority)	0.882111	0.882111	1.000000	0.937364
Traditional (Logistic Regression)	0.882111	0.882111	1.000000	0.937364
Proposed (Hybrid Ensemble)	0.882111	0.882111	1.000000	0.937364

3.5 Confusion Matrix Analysis

In order to assess the classification performance of the models, the confusion matrices are generated for the baseline model, the logistic regression model (traditional model), and the proposed hybrid ensemble model. The confusion matrices of these models are shown in Figures 5, 6 and 7, respectively.

As shown in Figure 5, the baseline majority classifier predicts all instances as being the dominant class. From the confusion matrix, we can see that 3008 instances that belong to the positive class are correctly classified, and 402 instances that belong to the negative class are incorrectly predicted as a part of the positive class. This behaviour is related to the drawback of the majority classifier, as it doesn't actually learn any meaningful decision boundaries but it just predicts the most frequent class.

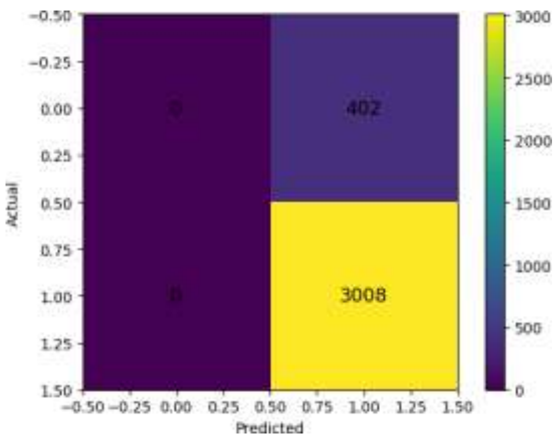

Figure 5. Confusion matrix for the baseline majority classifier.

Similarly, the confusion matrix of the traditional logistic regression model is shown in Figure 6, and the above-mentioned classification pattern is found. The model accurately recognizes 3008 positive cases but incorrectly recognizes all 402 negative cases as positive. This means that the logistic regression model is also highly influenced by the presence of class imbalance in the dataset.

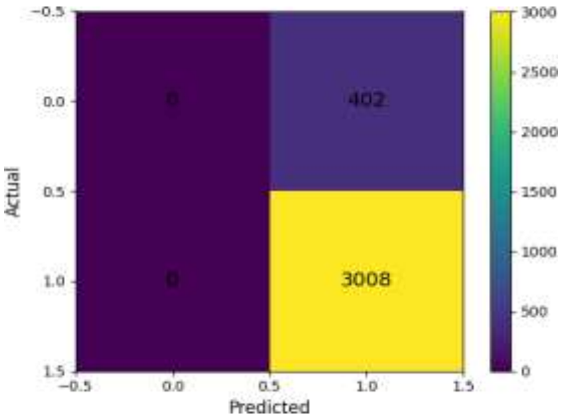

Figure 6. Confusion matrix for the traditional logistic regression classifier.

The confusion matrix of the proposed hybrid ensemble model can be seen in Figure 7. As we can see, the ensemble model has the same classification distribution as the baseline model and the logistic regression model. The model is able to predict 3008 positive instances and misclassifies 402 negative instances. This result further affirms that the classification performance is determined mostly by the distribution of the dominant class in the dataset not the model architecture. Overall, the confusion matrix analysis proves that all three models have the same prediction results because of the imbalanced nature of the dataset. While the proposed hybrid ensemble framework is shown to have a powerful predictive capability in the regression task, the classification results indicate the importance of balanced datasets when evaluating machine learning models for customer retention prediction.

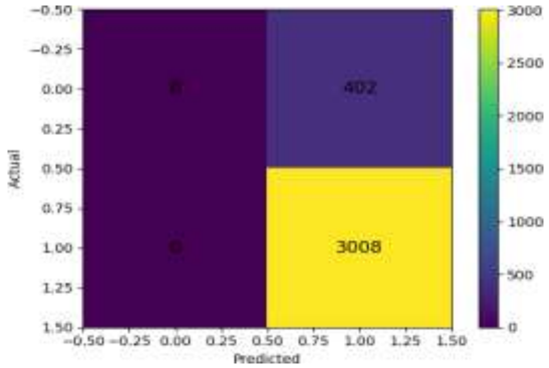

Figure 7. Confusion matrix for the proposed hybrid ensemble classifier.

3.6 Classification Model Comparison

In order to make a visual comparison of the performance of the classification, a graphical analysis was performed in terms of the F1-score metric, which represents a balanced measure of precision and recall. The comparison of classification models using F1-score metric is given in Figure 4. The figure of the visualization proves the hypothesis that the baseline, traditional, and proposed hybrid ensemble models produce an equivalent level of performance. The findings shown in Table 1 are supported by this observation as well.

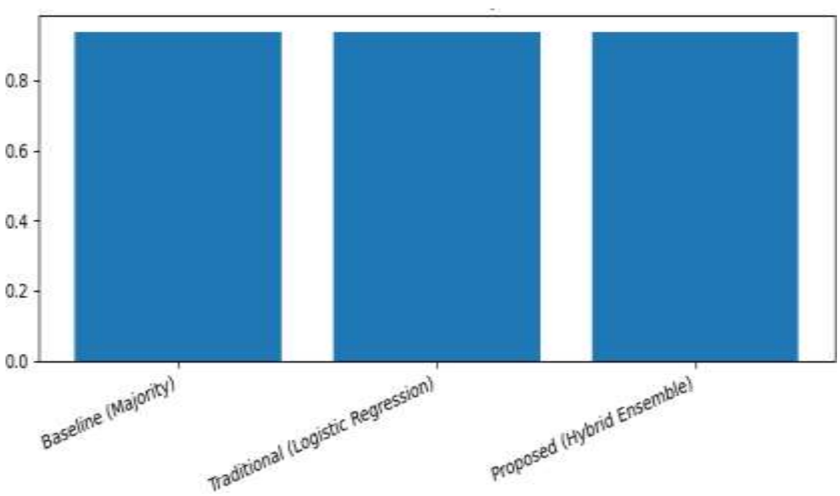

Figure 4. Comparison of classification models using the F1-score metric.

4. Discussion

The research findings of the experiment conducted in this paper demonstrate the usefulness of the suggested machine learning-based analytical framework in the context of intelligent decision support in commercial management systems. The framework is a combination of regression and classification models to conduct the analysis of commercial transaction data and forecast the values of transactions and customer retention behaviour. The results of the regression analysis indicate that the proposed hybrid ensemble model is significantly better at the task of prediction than the baseline and traditional model. The ensemble approach, which blends Random Forest and Gradient Boosting approaches, is effective in representing complex nonlinear relationships among variables such as product category, pricing, quantity of purchase and customer interaction behavior. This improvement in the predictive accuracy is the benefit of ensemble-based machine learning frameworks for commercial analytics application.

The results of the regression analysis show that the suggested hybrid ensemble model has less error values and higher coefficient of determination than traditional linear regression. The better performance indicates that ensemble learning methods can model the dynamic and multidimensional nature of commercial transaction datasets more effectively than the single model approach. Similar results have been found in studies involving the application of machine learning frameworks to complex decision-support systems applied in industrial and infrastructure fields. For example, the recent trend of sustainability assessment and intelligent analytical framework shows that the combination of machine learning approach with analytical modelling can improve the predictive capability and decision making processes in complex systems [15]. The results of the present study support this observation by showing that the integration of several machine learning algorithms can enhance the reliability of predictive analytics in commercial management systems.

In addition to regression analysis, the classification task was also carried out to predict whether a customer is likely to return to the platform. The accuracy, precision, recall and F1-score performances of the models are seen from the classification results and are similar for all the models. The confusion matrix analysis also suggests that the classification results are closely dependent on the distribution of dominant class in the dataset. This phenomenon is common in many real-world datasets where class imbalance has an impact on the ability of machine learning models to distinguish between minority and majority classes. Despite this limitation, the proposed framework does offer valuable insights into customer behavior and shows the potential of machine learning models to support retention-related decision-making processes.

The application of smart Decision support system in different fields has gained increasingly more attention over the past years. Specifically, machine learning constructs have found widespread applications in the area of environmental management, streamlining of energy systems and infrastructure. To illustrate, intelligent decision-support models have been utilised successfully in the management of ground water supply, and infrastructures control system to maximise the operational efficiency and the allocation of resources [16]. Equally, artificial intelligence along with digital twins has been shown to be effective in assisting the intelligent decision-making process in the complex cyber-physical environment [17]. The given developments underscore the growing significance of machine learning frameworks in the context of strategic decision-making in different domains.

The other important revelation that can be drawn through this study is the possible application of machine learning models to manage resources sustainably and intelligently. The recent research has concentrated on utilizing the artificial intelligence and machine learning in order to optimize the use of energy and improve sustainability in building management systems [18]. When applied in the business systems context, the predictive analytics solution can be employed to streamline the pricing strategies, discover valuable customers as well as improve customer interactions. Besides, models of the circular economy and customer-focused service platforms have also been optimized with machine learning-based decision making platforms to enhance the operational efficiency and support innovation-driven business models [19]. The findings of the present study correspond to these trends and are used to define the potential of machine learning methods in smarter and more data-oriented commercial decision support systems.

The applications of machine learning frameworks in the resource and agricultural management systems have also been increasingly employed in order to improve operational efficiency and sustainability. As an example, artificial intelligence-based irrigation systems to maximize the use of water resources and help achieve sustainable farms have been suggested [20]. On the same note, we have seen that physics-guided machine learning techniques are employed in the evaluation of the performance and economic feasibility of renewable energy systems [21]. These studies show that machine learning models can be widely applicable to solving complex optimization and decision support problems across the spectrum of different areas.

To conclude, the interaction of machine learning and digital twins has led to new possibilities of advanced data analytics and optimization of the system. According to recent studies, using text analytics in combination with digital twins and machine learning can assist in providing additional information about environmental performance and infrastructure control [22]. These structures indicate that a spectrum of analytical methods should be adopted in the quest to enhance the predictive modelling and decision support functions. Similarly, the hybrid ensemble framework that is suggested in this paper demonstrates that one can integrate machine learning models to create effective analytical instruments to serve commercial decision support services.

Generally, the findings of this paper justify the application of machine learning-based analytical constructs in enhancing predictive modelling and in facilitating intelligent decision making in business management systems. The combination of ensemble learning methods with proper data preprocessing and evaluation methods is a scalable strategy to analyze complex commercial datasets and derive actionable insights for business decision-makers.

5. Conclusion

This study showed analytical machine learning for intelligent decision support in commercial management systems by combining regression and classification analyses to analyze transaction and customer behaviour data. The proposed framework implemented a hybrid ensemble approach that used Random Forest and Gradient Boosting algorithms to improve the predictive performance. The performance of the regression results showed that the hybrid ensemble model has considerably outperformed the baseline and traditional linear regression models in terms of the prediction errors and the coefficient of determination, which revealed the hybrid ensemble model's capacity to successfully capture the complex relations contained within commercial datasets. Furthermore, the actual versus predicted analysis and the residual plots proved the stability and reliability of the proposed model in order to predict transaction values. In addition to revenue prediction, the framework was also tested for customer retention prediction using classification techniques. Although the baseline, logistic regression, and hybrid ensemble classifiers showed comparable classification results because of the imbalance of classes in the dataset, the experimental analysis showed the importance of the characteristics of the dataset in model performance. In general, the findings affirm that machine-learning-related analytic models can prove quite helpful towards expanding predictive analytics and data-driven decisions in the business world. The proposed framework offers a scalable and flexible way of analyzing commercial transaction data and can help organizations to better their operational strategies, customer engagement, and revenue forecasting.

References

[1] G. Ni, Y. Jia, Z. Shi, F. Chang, J. Miao, J. Wang, and W. Tang, "A full end-to-end analytical framework for livestock behavior modeling and health assessment using wearable electronic recording system and machine learning," *Smart Agricultural Technology*, p. 101686, 2025.

[2] K. Pourramezani, H. R. Baghaee, and G. B. Gharehpetian, "Data-driven management for cyber-physical-social distribution networks: A comprehensive review and complex system-of-systems data analytics framework," *Energy Reports*, vol. 14, pp. 2170–2190, 2025.

[3] Y. M. Banad, S. S. Sharif, and Z. Rezaei, "Artificial intelligence and machine learning for smart grids: From foundational paradigms to emerging technologies with digital twin and large language model-driven intelligence," *Energy Conversion and Management: X*, vol. 28, p. 101329, 2025.

[4] H. A. Al-Khamees, A. A. Smadi, M. K. Alsmadi, A. A. Alkannad, A. Abugabah, L. A. Almusfar, and B. Althani, "Advancing decision-making: A comprehensive review of intelligent systems, applications, and challenges," *Intelligent Systems with Applications*, p. 200631, 2026.

[5] S. K. Bhoi, S. Chakraborty, B. Verbrugge, S. Helsen, S. Robyns, M. El Baghdadi, and O. Hegazy, "Intelligent data-driven condition monitoring of power electronics systems using smart edge–cloud framework," *Internet of Things*, vol. 26, p. 101158, 2024.

[6] M. S. Murugan, "Large-scale data-driven financial risk management & analysis using machine learning strategies," *Measurement: Sensors*, vol. 27, p. 100756, 2023.

[7] S. Agarwal, R. Kumar, S. Yadav, C. Soundarya, P. Gottumukkala, and N. Gandhewar, "Large-scale data-driven financial risk management & analysis using back propagation neural network technique," in *Hybrid and Advanced Technologies*. Boca Raton, FL, USA: CRC Press, 2025, pp. 167–172.

[8] F. Li, T. Yigitcanlar, M. Nepal, K. Nguyen, and F. Dur, "Machine learning and remote sensing integration for leveraging urban sustainability: A review and framework," *Sustainable Cities and Society*, vol. 96, p. 104653, 2023.

[9] O. Das, M. H. Zafar, F. Sanfilippo, S. Rudra, and M. L. Kolhe, "Advancements in digital twin technology and machine learning for energy systems: A comprehensive review of applications in smart grids, renewable energy, and electric vehicle optimisation," *Energy Conversion and Management: X*, vol. 24, p. 100715, 2024.

[10] Y. Li, Q. Wang, X. Pan, J. Zuo, J. Xu, and Y. Han, "Digital twins for engineering asset management: Synthesis, analytical framework, and future directions," *Engineering*, vol. 41, pp. 261–275, 2024.

[11] E. Badakhshan, N. Mustafee, and R. Bahadori, "Application of simulation and machine learning in supply chain management: A synthesis of the literature using the Sim-ML literature classification framework," *Computers & Industrial Engineering*, vol. 198, p. 110649, 2024.

[12] C. K. Rao, S. K. Sahoo, and F. F. Yanine, "IoT enabled intelligent energy management system employing advanced forecasting algorithms and load optimization strategies to enhance renewable energy generation," *Unconventional Resources*, vol. 4, p. 100101, 2024.

[13] B. Biswal, S. Deb, S. Datta, T. S. Ustun, and U. Cali, "Review on smart grid load forecasting for smart energy management using machine learning and deep learning techniques," *Energy Reports*, vol. 12, pp. 3654–3670, 2024.

[14] U. Uygurr, "E-commerce Customer Behavior and Sales Analysis Dataset," *Kaggle*, 2024. [Online]. Available: https://www.kaggle.com/datasets/umuttuygurr/e-commerce-customer-behavior-and-sales-analysis-tr.

[15] L. Chen, L. Huang, X. Li, Z. Chen, K. S. Woon, P. S. Yap, and J. Chen, "Sustainability assessment on green construction using a novel analytical framework integrating machine learning and emergy analysis," *Developments in the Built Environment*, p. 100797, 2025.

[16] P. Ataei, A. Takhtravan, M. Gheibi, B. Chahkandi, M. G. Faramarz, S. Wacławek, and K. Behzadian, "An intelligent decision support system for groundwater supply management and electromechanical infrastructure controls," *Heliyon*, vol. 10, no. 3, 2024.

[17] A. Hasan and D. T. Nguyen, "Integrating agentic AI and digital twins for intelligent decision-making systems," *Array*, p. 100721, 2026.

[18] N. Ahmed, Y. Ding, G. Canti, and N. Anglani, "Sustainable building's energy management with artificial intelligence and machine learning: A decadal scoping review (2014–2024)," *Energy Reports*, vol. 15, p. 109082, 2026.

[19] T. Wicaksono, M. Marhadi, A. F. Wijaya, V. Anatasia, and K. Taralik, "Customer-centric circular economy as-a-service decision-making: Machine learning-driven open innovation in food service," *Cleaner Environmental Systems*, p. 100302, 2025.

[20] G. E. Oğuztürk, "AI-driven irrigation systems for sustainable water management: A systematic review and meta-analytical insights," *Smart Agricultural Technology*, vol. 11, p. 100982, 2025.

[21] U. Boonbumroong, N. Ditnin, and A. Jansasithorn, "Physics-guided machine learning framework for performance and economic assessment of a pico-hydroelectric system," *Results in Engineering*, p. 108582, 2025.

[22] S. Karatzas, G. Papageorgiou, V. Lazari, S. Bersimis, A. Fousteris, P. Economou, and A. Chassiakos, "A text analytic framework for gaining insights on the integration of digital twins and machine learning for optimizing indoor building environmental performance," *Developments in the Built Environment*, vol. 18, p. 100386, 2024.