International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group| Volume 6 Issue 4 PP 137-145, 2025
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.IJERET-V6I4P117

Original Article

Integration of AI in Customer Relationship Management (CRM) for Improved Sales Outcomes

Chetankumar Patel
Independent Researcher

Received On: 23/09/2025 ***Revised On: 06/09/2025*** ***Accepted On: 28/10/2025*** ***Published On: 19/11/2025***

Abstract *- Customer Relationship Management (CRM) systems especially are highly significant to businesses in the sales forecasting procedure and utilizing them as a key instrument in their data-driven decision-making method. In this regard, the authors introduce a GBM-SVM hybrid model that unifies the Gradient Boosting Machine and the Support Vector Machine in order to predict the sales of the Indian supermarket dataset available on Kaggle. The research methodology consists of a number of preprocessing steps, including the following: missing value imputation, error correction, outlier removal, one-hot encoding, and normalization. Feature engineering, too, contributes the creation of temporal attributes such as Month, Year, and Season to sales pattern recognition being more efficient. The dataset is divided for training and testing purposes using stratified sampling to guarantee equal class representation. The proposed GBM-SVM hybrid links the iterative boosting of GBM and the robust optimization of SVM to reduce prediction errors and increase forecasting accuracy. The results from the experiments show an excellent performance with R^2 of 99.9%, MAE of 5.12214, and MSE of 63.51036 which surpass that of classical models such as Random Forest (RF), Logistic Regression (LR), and ensemble techniques like XGBoost. The model's dependability for CRM is shown by the high connection between the anticipated and actual sales. Thus, smart inventory control, revenue forecasting, and strategic retail decision-making may all benefit from such a model.*

Keywords *- Customer Relationship Management (CRM), Artificial Intelligence, Predictive Analytics, Machine Learning, sales dataset, hybrid model (GBM-SVM).*

1. Introduction

Customer Relationship Management (CRM) is a vital feature of every company today as it allows not only to create but also to keep customer relationships that are strong and durable. The phrase CRM encompasses a whole range of approaches, methods, and technologies to control and explore customer communications throughout their lives [1][2]. Classic CRM systems mainly dealt with storing data, segmenting customers, and monitoring interactions; nevertheless, they often lacked generating practical findings for enhancing customer engagement and sales performance, thus making the engagement of potential customers a matter of few as the best strategy of organizations' sales facilitation [3]. The main reliance of the CRM systems was, ultimately, on the storage and organization of customer data. Nevertheless, with the unprecedented increase in the scale, diversity, and speed of customer data, conventional CRM practices are considerably limited and require more advanced analytical techniques to discover insights that can be acted upon.

Modern CRM systems have evolved from being simple software to become multi-channel and smart platforms that integrate cutting-edge technologies and customer experience (CX) ideas [4][5][6]. They handle all the main functions such as Marketing, Sales, and Service, allowing any company to have a unified view of its customers [7][8][9]. Besides, they usually find the right way to customer engagement by maintaining contact with the customers through multiple channels taking into account the different platforms, devices, and services and doing this in the most effective, efficient, and transparent way [10][11][12]. The companies that manage to keep up with the customers' engagement usually do so through their ability to understand customers, be flexible with changes and give support when most needed.

AI-based CRM solutions introduce the use of ML, NLP and predictive modelling to process and derive meaningful conclusions based on massive amounts of customer data. Using informative systems, it is very possible to determine such elements of purchase, customer feelings, and the possibility of customer churning fairly accurately. The work is done through AI by replacing the activities that are time-consuming and prone to mistakes: lead scoring, sales forecasting, and customer service responses, therefore, reducing human error and operation inefficiencies [13][14][15]. Continuous interaction solutions ensure quicker solutions with further engaging customer experience. The most sophisticated CRM systems powered by AI can even pick up on the emotions and intentions of the customers through their communications allowing the latter to be empathetic and thus, reinforcing the relationship over a long period [16][17][18]. AI-assisted CRM shortens the sales process, widens the possibilities of generating and qualifying leads, and the execution of super personalized marketing, which in turn, increases customer service quality, retention rates, and sales growth.

1.1. Motivation and Contribution

The primary motivation for the current research is the necessity for companies to adopt data-driven insights as the main tool for their sales performance and CRM. The fast development of AI and ML has resulted in an era where companies can easily conduct analysis over an extensive amount of CRM data to disclose hidden patterns as well as predict future sales trends. Nevertheless, the choice of the best predictive model is still a major issue. This study has its main goal to come up with a refined machine learning technique focusing on the GBM that yields extremely precise and dependable sales forecasts. The study is aimed at demonstrating the capabilities of ensemble learning methods to such a degree that they might be recognized as key contributors to the overall enhancement of decision-making, customer engagement, and sales performance, thus leading to a comparison of GBM with Random Forest, Logistic Regression, and XGBoost. This is a scientific research that has a number of critical contributions as discussed below:

- The data used in sales came in form of Kaggles where it contains transaction data of an Indian supermarket.
- Applied data cleaning, outlier removal and coding and normalization to enhance the quality of the dataset and model performance.
- Provided an effective, evidence-based system capable of facilitating decision-making and optimization of strategies in the CRM-based sales management.
- The combination of methods (SVM-GBM) ought to be applied as the optimization of the sales performances on the CRM data.
- An increase in R^2 indicates predictive competence, where the best-fit line can make more accurate predictions.

1.2. Justification and Novelty

This research has a powerful motivation in the growing need of data-driven methods of sales optimization and customer interaction in CRM systems. The traditional methods of analysis are unable to reveal non-linear, obscure relationships within sales data which leads to poor predictions. The analysis recommends that GBM model which combines various weak learners to create an extremely precise and strong prediction system be used. This work is novel as it introduces the use of GBM in CRM-based sales forecasting and performance comparison of the model with other models such as RF, LR and XGBoost. This innovative approach does not only enhance the degree of sales forecasting but also indicates the utilization of advanced ensemble learning methodology as a significant contribution towards enhancing decision making, profitability and customer relationship management in real business considerations.

1.3. Organization of the Paper

The structure of the document is as follows: Section II is related to machine learning processes using CRM and sales forecasting. In Section III, the sales data, data cleaning techniques, and model-building process are claimed. The next Section IV is dedicated to showing and analyzing the experimental results through performance evaluation and model comparisons in machine learning. The last Section V wraps up the research with directions for further studies. It's up to researchers to decide how far they want to go.

2. Literature Review

The present research consists of a comprehensive literature survey on the deployment of Customer Relationship Management (CRM) to facilitate better sales outcomes. The review focuses on different types of ML and DL techniques for the improvement of CRM quality and for sales prediction. Table I summarizes key studies, highlighting proposed models, key findings, and limitations.

Mazur et al. (2025) purpose of this research is to create classification models that enable the prediction of a certain Mercedes-Benz offer's likelihood of success in relation to vehicle configuration. The literature on customer relationship management confirms that such a technology makes it possible to allocate resources (salespeople's time, media budgets, production capacity) optimally. This paper evaluates four machine learning models including Random Forest (RF), Gradient Boosting Machine (GBM), eXtreme Gradient Boosting (XGBoost), and Support Vector Machine with RBF kernel (SVM-RBF) basing their efficacy on sales that was represented by the binary variable Success. Random Forest scored 84.3, 0.73 and 0.90 as the accuracy, F1-score and AUC respectively, which was better than the assessed models illustrating a strong ability to distinguish successful and unsuccessful transactions. The results can be used in distribution/production planning, dynamic discounting, lead prioritization and marketing campaign optimization [19].

Tiwari (2024) the results indicate that effective CRM implementation produces a high sales performance. This is because the deployment enables focused marketing strategies, boosts client satisfaction, and promotes sales efficiency. Important success factors for customer relationship management systems include user acceptance, system integration, and data quality. The essay concludes with suggestions for retail managers on how to maximize customer relationship management systems to increase sales and maintain a competitive edge. Learn useful strategies for using customer relationship management (CRM) systems to improve sales effectiveness [20].

Yadav et al. (2023) evaluates and assesses the effectiveness of several machine learning (ML) techniques to address the email marketing subscriber prediction problem. For this work, many analytical machine learning techniques particularly categorization and regressor techniques, that fall under various learning categories are chosen. The dataset of emails with twenty-three features was subjected to models. The results of the experiment show that, with nearly equal accuracy of 95%, Random Forest (RF) and Adaboost surpass all other machine learning techniques. The greatest R^2 scores, 89.8% and 91%, were obtained by KNN and the ensemble method. According to the comparison, the ensemble approach

performs better in terms of accuracy, error value, and R2 score than cutting-edge machine learning techniques [21].

Khumaidi, Nirmala and Herwanto (2022) use exploratory Data Analysis to draw conclusions. An LSTM model's development is influenced by a number of parameters. Data partition, hidden layer count, neuron count, epoch, batch size, and the amount of training data needed for each weight update are all part of this. Each of these variables describes a different aspect of the model. With a hyperparameter batch size of 30, an epoch of 150, three hidden layers, and three dropouts, an RMSE testing of 0.0846 and an RMSE training of 0.0855 were judged to be the best LSTM models [22].

Nagaraju and Vijaya (2021) a way to help that kind of business deal with customer turnover by using ML methods. They use the XYZ Insurance customer churn dataset, which is based in Indonesia, to test out various DL and ML algorithms. ANN classifier. With an accuracy of 91.3111% and a 0.970, DT with forward selection exceeds ANN and NB. For this customer churn dataset and all of XYZ Insurance's CRM resources that aim to improve their automated and vital functions, the Decision Tree technique is recommended [23].

Nguyen et al. (2021) offer a framework for data-driven decision-support systems that model customer sentiment and forecast escalations using a combination of company data and the Internet of Things. A large medical device manufacturer contributes data from thousands of high-end devices to a real-world case study that applies the framework. This industrial benchmark includes intriguing and difficult features; an anonymized version of it is available. Extensive trials show that the proposed framework improves prediction results and makes troubleshooting a viable workflow for end-users by merging IoT and corporate data. They reach a Recall of 50.0% for the task of forecasting customer escalations [24].

Research Gap: Despite significant advancements in using ML and DL models to enhance sales outcomes through CRM, several research gaps remain. Many existing studies focus on limited datasets, lack generalization across industries, or overlook the integration of real-time and unstructured customer data, such as sentiment or behavioral patterns. Additionally, few models emphasize explain ability and scalability for practical CRM applications. Addressing these gaps can lead to more robust, interpretable, and adaptive CRM systems for improved sales performance and customer.

Table 1: Recent Studies on Improved Sales Outcomes Using A Machine Learning Model

Author (Year)	Dataset	Key Findings	Limitations	Future Work
Mazur et al. (2025)	Mercedes-Benz vehicle configuration and sales dataset with binary "Success" variable	Random Forest achieved highest performance (Accuracy 84.3%, F1 = 0.73, AUC = 0.90); ML models effectively predict sales success; results useful for resource allocation, production planning, dynamic discounting, and marketing optimization	Study limited to one automotive brand; binary classification may oversimplify customer decision dynamics; external validity across other markets/products not tested	Extend model to multi-class or probabilistic predictions; evaluate deep learning models; test generalizability across different vehicle brands or industries; integrate real-time customer behavior data
Tiwari (2024)	CRM implementation datasets from retail organizations (qualitative + performance KPIs)	Successful CRM deployment improves sales performance through targeted marketing, higher satisfaction, and enhanced efficiency; highlights success factors like user acceptance, integration quality, and reliable data	Primarily qualitative insights; lacks advanced ML-based analysis; findings may vary across sectors depending on digital maturity and organizational culture	Incorporate analytics-driven CRM evaluation frameworks; assess impact of AI-enabled CRM systems; include cross-industry comparative studies; integrate real-time customer engagement metrics
Yadav et al. (2023)	Email marketing dataset containing 23 features related to subscriber behavior	• Random Forest and AdaBoost achieved the highest accuracy (~95%). • KNN and ensemble models achieved the best R^2 scores (89.8% and 91%). • Ensemble techniques outperformed other ML models in accuracy, error rate, and R^2 score.	• Performance depends on dataset quality and feature distribution. • Limited generalization to datasets from other domains. • Potential overfitting in ensemble models.	• Explore deep learning methods for subscriber prediction. • Incorporate real-time behavioral data. • Improve feature engineering and cross-domain model validation.
Khumaidi, Nirmala & Herwanto (2022)	Customer data analyzed using Exploratory Data Analysis	Created an optimised LSTM model with the following parameters: 30 batches, 150 epochs, RMSE (train) =	Model performance depends on limited parameter tuning; scalability to larger	Extend hyperparameter optimization and apply cross-validation across diverse datasets to

		0.0855 and RMSE (test) = 0.0846 were attained using three hidden layers and 3 dropouts:	datasets not tested	generalize results
Nagaraju & Vijaya (2021)	Insurance Customer Churn Dataset	DT with forward selection achieved highest accuracy (91.31%) and precision (0.970), outperforming ANN and NB	Limited to one insurance dataset; lacks generalization to other CRM sectors	Apply ensemble learning and feature engineering across multi-industry CRM datasets to improve robustness
Nguyen et al. (2021)	IoT and enterprise data from a medical device	Proposed IoT-integrated decision support system achieving 50% Recall in escalation prediction; improved troubleshooting and workflow efficiency	Moderate recall; needs higher precision for industrial applications	Incorporate advanced deep learning (e.g., attention models) and larger IoT datasets to enhance prediction reliability

3. Research Methodology

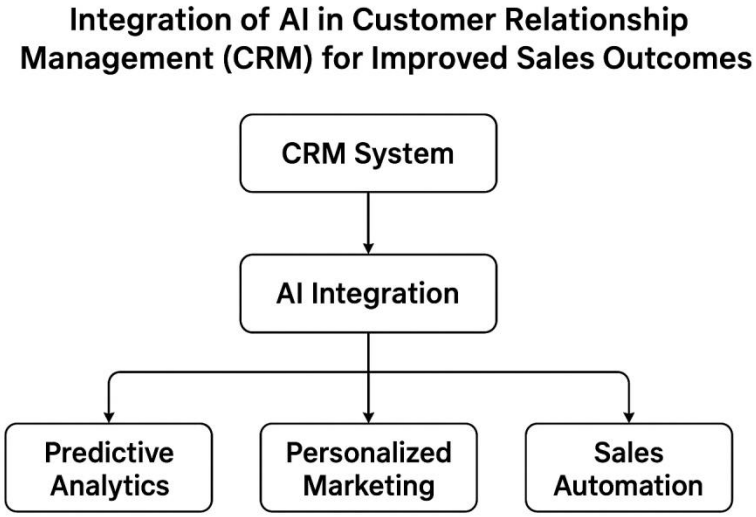

Fig 1: Proposed Flowchart for Improved Sales Outcomes Using Customer Relationship Management (CRM)

The methodology of the study uses a well-organized machine learning technique to enhance sales results relying on CRM data processed through the hybrid (GBM-SVM) model. First, it takes data from the Kaggle Indian supermarket sales dataset and performs analysis on it, then it applies preprocessing which includes treatments like missing values, error correcting, outliers removal, one-hot encoding application, and normalization. Feature engineering created new time-based variables, e.g., Month, Year, and Season, which are more effective in depicting sales trends. After that, the dataset was created by splitting it into training and testing with 80:20 stratified ratio, and thereby keeping class balance. In order to lower prediction errors and enhance the model's performance, weak learners were repeatedly combined to generate the hybrid model. Ultimately, the accuracy of the model was tested using R^2, MAE, and MSE metrics ensuring the predictions were accurate and trustworthy. The proposed methodological flowchart for sales outcomes enhancement through Customer Relationship Management (CRM) is represented in Fig. 1.

The subsequent part of the text comprehensively clarifies each step of the proposed steps, explaining the methods used in data preparation, model building, and performance assessment. Every phase is meticulously planned to give precision, speed, and consistency in reaching the aims of the research:

3.1. Data Gathering and Analysis

The dataset containing sales information that was analyzed in this research came from an Indian supermarket and was downloaded from Kaggle. It consists of 1000 entries and 17 fields providing comprehensive data of the different product lines sold during a continuous four-year span from 2015 to 2018. In order to reveal important insights from the data, drawings, including bar graphs and heatmaps, were applied for elucidating the distribution of sales and correlation among features:

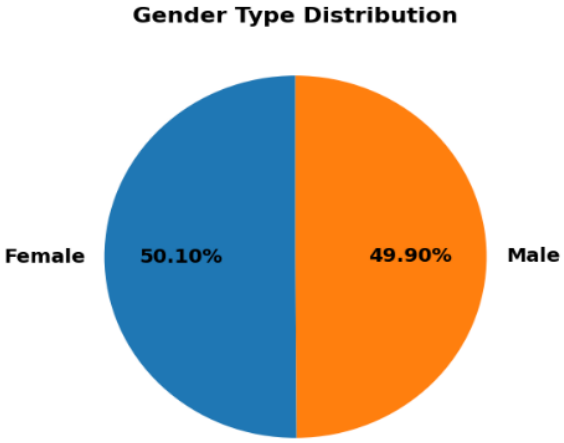

Fig 2: Class Distribution of Sales Dataset

In Fig. 2, the two categories, Female and Male, are shown to be almost equally divided. The information shows that the Female category represents 50.10% of the total distribution which is identified with the yellow-gold half of the chart. The Male category is slightly smaller, accounting for 49.90% of the distribution marked in blue. The visual representation distinctly displays that the dataset or population being evaluated is practically even in terms of gender distribution, with the female group holding a tiny, insignificant majority of 0.20 percentage points.

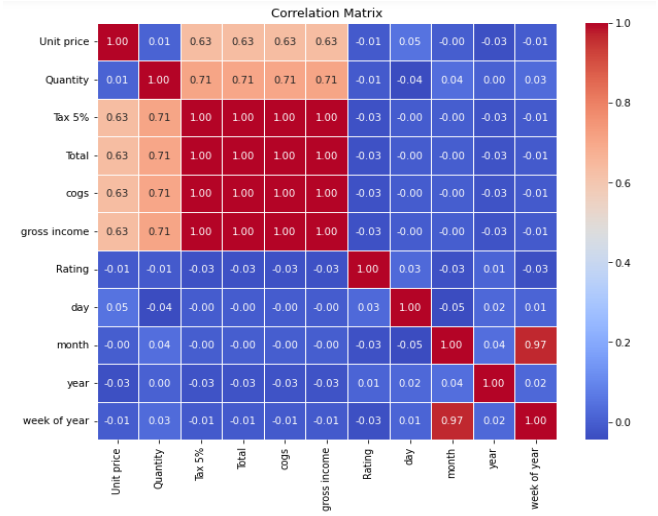

Fig 3: Correlation Matrix Heatmap on Sales Dataset for Improving Sales Outcome

The heatmap of the correlation matrix provides a visual representation of the associations among CRM sales variables, including amount, tax, total, gross income, COGS, rating, and mode of payment. In this presentation, red color indicates high positive correlation (~1.0) between various financial metrics that drive sales performance as depicted in Fig. 3 and blue color indicates weak correlations (~0) between categorical variables and the diagonal line indicates perfect correlations with oneself, hence enabling the use of data to shed light on how to improve customer relationship management and sales performance.

3.2. Data Pre-Processing

A full data preparation procedure that included concatenation, data cleansing and feature engineering was applied to the Sales dataset. The preprocessing stage addressed the problems of missing data, miscellaneous data error correction, deletion of outliers, encoding and normalization of data and moving towards the consistency and accuracy point. The most significant preprocessing steps are summarized and coiled up as follows:

- **Addressing missing values:** The process of handling missing values consisted of identifying and managing the partial data entries that were presented within the dataset. Techniques such as mean or median substitution for numerical variables and mode or categorical replacement for non-numerical ones were applied for missing values imputation.
- **Identifying Errors:** Error detection was about spotting entries in the dataset that were wrong, inconsistent, or duplicated. After detecting these errors, they were either corrected or removed, so data integrity was maintained and quality analysis was guaranteed.
- **Remove outliers:** Outlier removal covered the detection of data points that differed significantly from the normal distribution. The outliers were eliminated so that there would be no skewness in model training and, as a result, the prediction accuracy would improve.
- **One-Hot Encoding:** One-hot encoding method was a very essential encoding technique in machine learning among others that were introduced. The process works by placing a 0 or a 1 in several columns depending on the values presents in one particular column. The binary values represent the relationship of the merged and encoded columns.

3.3. Feature Engineering

The process of creating additional variables from the old ones by enhancing the predictive capacity of the model is known as feature engineering. The data's quality was improved by creating new features that represented the data's underlying patterns and relationships. This technique was essential in increasing the ML model's capability and accuracy. During this stage, new features were generated based on the existing ones; to analyze the sales trend through time, new features that possibly would be relevant were set up. The new features are Day, Month, Year, Month name, Month year, Week of year, and Season. These new features enable us to discover the sales trends over time.

3.4. Min-Max Normalization

Normalization of the records was accomplished by applying the min-max method which confines the values between 0 and 1. The intention behind this was to improve the performance of the classifiers and reduce the influence of outliers. The normalization procedure was mathematically expressed with the use of the following Equation (1):

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (1)$$

Where X represents the original value of the feature, $X'$is the value after normalization, $X_{min}$is the lowest value of the feature, and $X_{max}$is the highest value of the feature.

3.5. Data Splitting

The initial data set was separated using stratified splitting with an 80:20 split into train and test sets. The method used here made certain that the two subsets kept the exact same class distribution as the original dataset, which, in turn, maintained the balance and represented each class during model training and evaluation.

3.6. Classification Equation of (GBM-SVM) Model

A supervised ML model, namely the GBM, was proposed to enhance sales outcomes through the effective use of CRM data. GBM is a potent ensemble learning method that combines many weak learners, usually DT, to create a strong prediction model. GBM, owing to its sequential nature, great predictive performance, and resistance to overfitting when well adjusted, is used broadly in both classification and regression applications. It has the capacity to deal with intricate data patterns as well. The idea behind GBM is to slowly integrate weak classifiers into Equation (2) in order to minimize a differentiable loss function L(y, F(x)) Equation (2):

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

In the Gradient Boosting Machine (GBM) model,$F_m(x)$ represents the model after the m^{th} iteration, which is the cumulative result of all previous weak learners added up to that point. The notation of $F_{m-1}(x)$ indicates the previous iteration of the model, which act as the reference point, at which the current learner trying to improve. The function $h_m(x)$ corresponds to the new weak learner. Finally, γ_m represents the learning rate or step size, a crucial hyperparameter that controls how much each new tree contributes to the overall model. A smaller γ_m value generally slows down learning but enhances model stability and accuracy.

SVMs are a popular supervised learning approach for regression and classification tasks. The fundamental principle behind support vector machines (SVMs), which consist of mathematical descriptions, hyperplanes, and classification margins, is to optimize the space between data points when they are separated into distinct classes. With a set of training data (x1, y1), (x2, y2),., (xn, yn) in Fig. 4, where xi ∈ R d represents a feature vector and yi ∈ {−1, +1} specifies the class label.

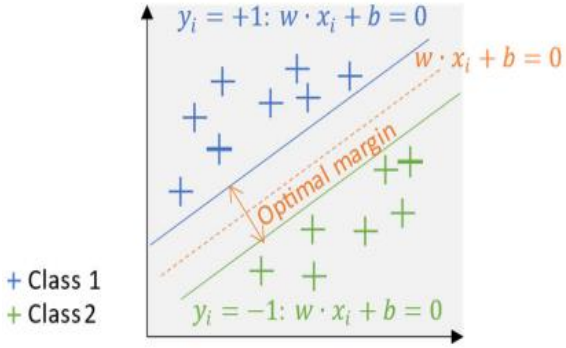

Fig 4: SVM Model in Optimal Hyperplane

The goal of SVMs is to identify the best hyperplane, $w.x_i + b = 0$, for classifying data points.
For all values of i in Equation (3), SVM imposes the following requirement.

$$y_i(w.x_i + b) \geq 1, \quad \forall_i \tag{3}$$

The minimization of the normal vector w's norm is another way to put it. Equation (4) expresses the ensuing optimization problem:

$$minimize\, \frac{1}{2}\|w\|^2 \tag{4}$$

The expression $\frac{1}{2}\|w\|^2$ represents the goal, which is to maximize the margin of separation between the classes by minimizing the norm of the normal vector w.

3.7. Evaluation Metrics

Model evaluation is an essential step in the process of ML projects, as it not only reveals the model's performance but also helps in displaying the results in a straightforward manner. It is the case with regression models, which, instead of predicting preciseness, only aim to approximate the values; therefore, the evaluation was made in terms of how close the predicted outcomes were to the actual values using the main performance metrics which are R², MAE, and MSE.

3.7.1. R-Squared

R² measures the accuracy of the model prediction. A higher R² value indicates that the model fits the data better; R² values can take on values between 0 and 1. An R² score of 1 indicates a perfect prediction of the response data, whereas a model with an R² value of 0 is unable to account for any of the variability surrounding its mean. and may find R² using Equation (5):

$$R^2 = \frac{\sum_{i=1}^{n}(y_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{5}$$

3.7.2. Mean Absolute Error (MAE)

A common metric for assessing the accuracy of a prediction model is the MAE. Without taking the direction of the errors into account, it calculates the average magnitude of the predictions. Performance is improved with a lower MAE number. To determine MAE, use the Equation (6):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(y_i - y_i^P)| \tag{6}$$

Where,
The observed value is denoted by y, the anticipated value by y, and the total number of observations by n.

3.7.3. Mean Squared Error (MSE)

In the domain of predicting Real Estate Prices, when y_i indicates the real price during the period of i and y_i^p Signals the projected price at time, the MSE is derived as per Equation (7) referenced below.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^p)^2 \quad (7)$$

These are the basic metrics. These indicate how well the models are doing in predicting the target variable. In this scenario, they are the basic signs of model's performance.

4. Results and Discussion

The experimental setup is described in this section, and The efficiency and computing capacity of the suggested model are demonstrated through the presentation of its performance assessment throughout both the training and testing stages. The experiments on the enhancement of sales results through the application of the CRM technique were conducted on a top-notch computer. With an NVIDIA RTX 4090 graphics card with 24 GB of VRAM, a 3.0 GHz Intel Core i9-13900K CPU, and 64 GB of DDR5 memory, this PC ran Windows 11 Pro. Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn are some of the Python libraries that were used for data processing, system engineering, and output analysis. The suggested model was constructed using the sales dataset, and its performance was assessed using key performance measures such as R², MAE, and MSE, as shown in Table II. Classification results from the sales dataset show that the suggested model for CRM improved sales performed well. The forecasted and actual scatter plot of the GBM-SVM hybrid model shown in Fig. 5 that exhibits a high linear relationship between forecasted and the actual values of the total sales with the range of 0 to 900. The fact that the blue data points are almost packed in a diagonal which indicates extremely good model performance with nearly no prediction errors is a testament that show indeed that the ensemble method is actually an efficient in CRM analytics in the context of effective sales forecasting.

Table 2: Classification results of the proposed model improved sales outcomes using CRM, using the Sales Dataset

Performance Matrix	Hybrid (GBM-SVM) Model
R^2	99.9
MAE	5.12214
MSE	63.51036

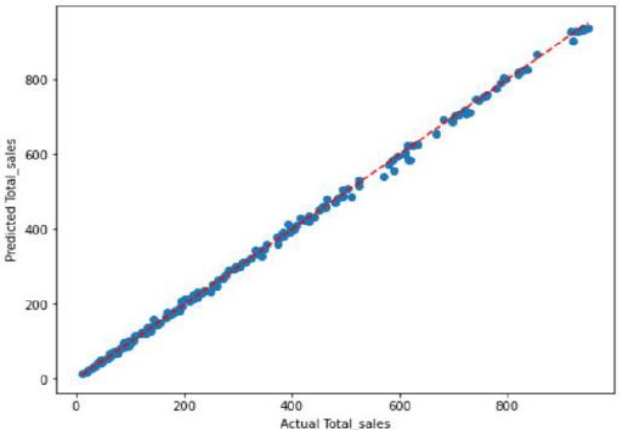

Fig 5: Predicted vs Actual Scatter Plot for the GBM-SVM Model

The forecasted and actual scatter plot of the GBM-SVM hybrid model shown in Fig. 5 that exhibits a high linear relationship between forecasted and the actual values of the total sales with the range of 0 to 900. The fact that the blue data points are almost packed in a diagonal which indicates extremely good model performance with nearly no prediction errors is a testament that show indeed that the ensemble method is actually an efficient in CRM analytics in the context of effective sales forecasting.

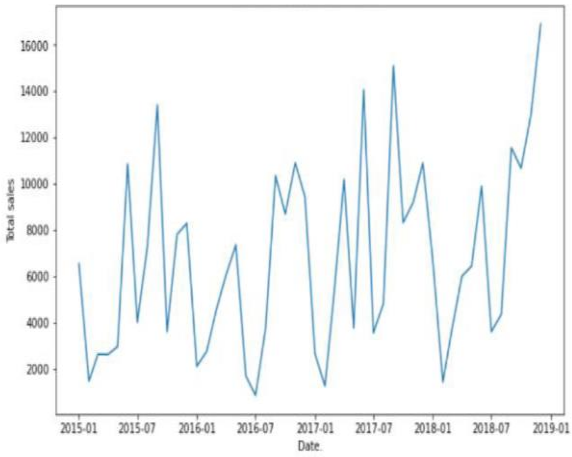

Fig 6: Time Series Plot for the GBM-SVM Model

The time series diagram of the GBM-SVM model, which includes the total sales fluctuations between the years 2015 and 2019 is provided in Fig. 6. The diagram shows substantial fluctuation with sales figures going as low as 2,000 and as high as 16,000, marking seasonal and market impacts through their periodic highs and lows, besides the upward trend of about 2019 which is The clearest sign of the model's capacity to identify temporal relationships.

4.1. Comparative Analysis

The performance of the proposed GRU model was evaluated through a comparative accuracy analysis, which is depicted in Table III. A comparison of ML models for improving sales highlights the supremacy of GBM over other models for any performance measure possible. The RF recorded an R² of 94 along with an MAE of 19.79, whereas

LR and XGBoost were awarded R^2 values of 96 and 97.6, respectively. However, the GBM model surpassed all with an incredible R^2 of 99.90 and the lowest MAE of 5.12214. This proves the exceptional accuracy and dependability of GBM in forecasting sales results using the CRM-based sales dataset.

Table 3: Comparison of Different Machine Learning Models for Improved Sales Outcomes

Model	MAE	R^2
RF[25]	19.79	94
LR[26]	2480.12	96
XGBoost[27]	1317.65	97.6
GBM-SVM	5.12214	99.90

The GBM-SVM hybrid model that was being suggested showed a big benefit by having an unbelievably high R^2 value of 99.9% which resulted in its being recognized as the best predictor of sales outcomes to date. This almost perfect coefficient of determination not only justifies the hybrid model but also indicates the scientific integration of GBM's powerful feature learning capabilities with SVM's strong classification and regression optimization. Hence, the model is able to understand and represent in its predictions very complicated and non-linear relationships between input features and sales performance, so it is very trustworthy for forecasting and decision making in the area of CRM applications.

5. Conclusion and future study

CRM systems are not only the operations for managing customer interactions and relationships but also the ways to identifying the right ones, thus allowing for stratifying customers very precisely. They not only collect data of customers from all possible sources, but also make sales and marketing departments work together and which constitutes the modern way of managing customers. AI added to CRM gives a strong push in the direction of the personalization of customer insights, the taking over of the manually done tasks, and the improvement of sales with their optimization. In SugarCRM, the AI-based integration allows forecasting analytics for sales, AI-oriented marketing, and AI-supported customer service, thus leading to more efficient customer interaction practices and higher customer satisfaction. According to empirical research, among the different types of machine learning models used for sales forecasting, the GBM-SVM hybrid model has proved to be the most accurate one. Although RF yielded an R^2 of 94, LR got 96, and XGBoost was a bit better with 97.6, the GBM model was way ahead of the rest with a remarkable R^2 of 99.90, which guarantees excellent prediction performance. The main theme of the project in taking gradual steps towards CRM systems will be AI, where hybrid and deep learning models that can adjust to evolving customer behaviors can be implemented. The integration of real-time data analytics and XAI will increase the level of transparency, flexibility, and improved accuracy in predictions, and the multi-source integration of data should be extended to cover social media, behavioral, and transactional data to provide more personalized and intelligent customer management.

References

[1] U. Zeynep Ata and A. Toker, "The effect of customer relationship management adoption in business-to-business markets," *J. Bus. Ind. Mark.*, vol. 27, no. 6, pp. 497–507, Jul. 2012, doi: 10.1108/08858621211251497.

[2] E. O. Nogueira and M. Borchardt, "The effects of customer relationship management (CRM) on e-commerce evolution: A systematic review," *Tech. Soc. Sci. J.*, vol. 36, pp. 433–453, Oct. 2022, doi: 10.47577/tssj.v36i1.4124.

[3] V. Shah, "Managing Security and Privacy in Cloud Frameworks: A Risk with Compliance Perspective for Enterprises," *Int. J. Curr. Eng. Technol.*, vol. 12, no. 06, pp. 1–13, 2022, doi: 10.14741/ijcet/v.12.6.16.

[4] Bangar Raju Cherukuri, "Developing Intelligent Chatbots for Real-Time Customer Support in E-Commerce," *Int. J. Sci. Res.*, vol. 11, no. 1, pp. 1709–1719, 2022.

[5] J. Thomas, K. V. Vedi, and S. Gupta, "Enhancing Supply Chain Resilience Through Cloud-Based SCM and Advanced Machine Learning: A Case Study of Logistics," *J. Emerg. Technol. Innov. Res.*, vol. 8, no. 9, 2021.

[6] S. Das and J. Nayak, "Customer Segmentation via Data Mining Techniques: State-of-the-Art Review," 2022, pp. 489–507. doi: 10.1007/978-981-16-9447-9_38.

[7] C. Hildebrand and A. Bergner, "AI-Driven Sales Automation: Using Chatbots to Boost Sales," *NIM Mark. Intell. Rev.*, vol. 11, no. 2, pp. 36–41, Nov. 2019, doi: 10.2478/nimmir-2019-0014.

[8] S. Chatterjee, S. K. Ghosh, R. Chaudhuri, and B. Nguyen, "Are CRM systems ready for AI integration?: A conceptual framework of organizational readiness for effective AI-CRM integration," *Bottom Line*, vol. 32, no. 2, pp. 144–157, 2019, doi: 10.1108/BL-02-2019-0069.

[9] Tanmaykumar Vithaldas Shah, "Leadership in digital transformation: Enhancing customer value through AI-driven innovation in financial services marketing," *Int. J. Sci. Res. Arch.*, vol. 15, no. 3, pp. 618–627, Jun. 2025, doi: 10.30574/ijsra.2025.15.3.1767.

[10] K. Murugandi and R. Seetharaman, "Analysing the Role of Inventory and Warehouse Management in Supply Chain Agility : Insights from Retail and Manufacturing Industries," *Int. J. Curr. Eng. Technol.*, vol. 12, no. 6, pp. 583–590, 2022.

[11] H. Kali, "Redefining Employee Experience: The Role of Workday HCM in Workplace Digitalization," *Int. J. Eng. Sci. Math.*, vol. 11, no. 09, pp. 1–6, 2022.

[12] P. R. Marapatla, "AI-driven donor management: revolutionizing nonprofit fundraising through predictive analytics," *Int. J. Res. Comput. Appl. Inf. Technol.*, vol. 8, no. 1, 2025.

[13] R. Rawat and R. Yadav, "Big Data: Big Data Analysis, Issues and Challenges and Technologies," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1022, no. 1, p. 012014, Jan. 2021, doi: 10.1088/1757-899X/1022/1/012014.

[14] T. A. Mittiga, "Sensing the local charge and strain environments surrounding Nitrogen-Vacancy centers in diamond," University of california, 2020.

[15] A. R. Bilipelli, "End-to-End Predictive Analytics Pipeline of Sales Forecasting in Python for Business Decision Support Systems," *Int. J. Curr. Eng. Technol.*, vol. 12, no. 6, pp. 819–827, 2022.

[16] M. M. Kowsar and M. A. Rahman, "Enterprise Resource Planning and Customer Relationship Management Integration: A Systematic Review of Adoption Models and Organizational Impact," *Rev. Appl. Sci. Technol.*, vol. 01, no. 02, pp. 26–52, Jun. 2022, doi: 10.63125/tn7g2v08.

[17] C. Sezer, *Visibility, democratic public space and socially inclusive cities*, vol. 10, no. 4. 2020. doi: 10.7480/abe.2020.16.4604.

[18] Bailey and W. Brett, "The efficacy of incident management teams and emergent multi-organizational networks in the implementation of the incident command system," *J. Retail. Consum. Serv.*, 2016.

[19] M. Mazur, O. Stopka, M. Stopková, J. Hanzl, A. Borucka, and R. Czerniak, "Assessing the Success of Automotive Sales Transactions Using Selected Machine Learning Algorithms," *Appl. Sci.*, vol. 15, no. 21, p. 11562, Oct. 2025, doi: 10.3390/app152111562.

[20] S. Tiwari, "Customer Relationship Management (CRM) Systems: Their Impact on Sales Performance in the Retail Industry," *Libr. Prog. Int.*, vol. 45, no. 1, pp. 55–60, 2024.

[21] D. Yadav, J. Singh, P. Verma, V. Rajpoot, and G. Chhabra, "A Novel Approach for Enhancing Customer Retention Using Machine Learning Techniques in Email Marketing Application," in *2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS)*, 2023, pp. 1–6. doi: 10.1109/PCEMS58491.2023.10136072.

[22] A. Khumaidi, I. A. Nirmala, and H. Herwanto, "Forecasting of Sales Based on Long Short Term Memory Algorithm with Hyperparameter," in *2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE)*, IEEE, Jan. 2022, pp. 201–206. doi: 10.1109/ISMODE53584.2022.9743079.

[23] J. Nagaraju and J. Vijaya, "Methodologies used for Customer Churn Detection in Customer Relationship Management," in *2021 International Conference on Technological Advancements and Innovations (ICTAI)*, IEEE, Nov. 2021, pp. 333–339. doi: 10.1109/ICTAI53825.2021.9673382.

[24] A. Nguyen *et al.*, "System Design for a Data-Driven and Explainable Customer Sentiment Monitor Using IoT and Enterprise Data," *IEEE Access*, vol. 9, pp. 117140–117152, 2021, doi: 10.1109/ACCESS.2021.3106791.

[25] S. Bauskar, "Predictive Analytics for Sales Forecasting in Enterprise Resource Planning (ERP) Systems Using Machine Learning Technique," *Int. Res. J. Mod. Eng. Technol. Sci.*, vol. 04, no. 06, pp. 2582–5208, 2022, doi: 10.56726/IRJMETS26271.

[26] C. Catal, K. Ece, B. Arslan, and A. Akbulut, "Benchmarking of Regression Algorithms and Time Series Analysis Techniques for Sales Forecasting," *Balk. J. Electr. Comput. Eng.*, vol. 7, no. 1, pp. 20–26, Jan. 2019, doi: 10.17694/bajece.494920.

[27] Y. F. Akande, J. Idowu, A. Misra, S. Misra, O. N. Akande, and R. Ahuja, "Application of XGBoost Algorithm for Sales Forecasting Using Walmart Dataset," in *Lecture Notes in Electrical Engineering*, 2022, pp. 147–159. doi: 10.1007/978-981-19-1111-8_13.