

Predictive Analytics in Marketing: Evaluating its Effectiveness in Driving Customer Engagement

Khawla Alhasan

Faulty of Business Studies, Arab Open University, Kuwait

ARTICLE INFO

Article history:

Received: 12 January, 2025

Revised: 26 March, 2025

Accepted: 27 March, 2025

Online: 15 June, 2025

Keywords:

Artificial Intelligence

Marketing

Machine Learning

ABSTRACT

Understanding and responding to customer feedback is critical for business success. Customer response data offers valuable insights into preferences, behaviours, and sentiment. By analysing this data, businesses can optimize strategies, enhance customer experiences, and drive growth. Many analysis have been conducted in this field, while the review covers a broad range of AI and ML applications in marketing, all of these analysis were done separately ending with scattered conclusions and recommendations. This study explores a big set of machine learning (ML) techniques to analyse customer response data from a marketing campaign in a retail superstore. The dataset was subjected to thorough pre-processing steps, including handling missing values, outlier detection and feature engineering. Exploratory data analysis (EDA) was conducted to gain awareness of customer behaviour and campaign effectiveness. Various ML models, including Decision Trees, Logistic Regression, Random Forest, K-Nearest Neighbour (KNN), as well as Support Vector Classifier (SVC), were trained and evaluated regarding the balanced dataset. Essentially, our findings propose that ML techniques can effectively analyse customer response and provide valuable information to uncover hidden patterns, make informed decisions, and gain a competitive advantage to assist with optimising marketing strategies in retail environments

1. Introduction

Artificial Intelligence (AI), defined as the simulation of human intelligence by machines, is transforming marketing through its capacity to analyze customer data and predict trends and preferences [1]. The increasing sophistication of machine capabilities, driven by improved algorithms, cost-effectiveness, and the availability of large datasets, empowers businesses to refine their marketing strategies for enhanced customer engagement [2]. Industries such as fashion are leveraging AI-powered trend analysis and predictive analytics [3] to gain a competitive edge. While the integration of AI into marketing presents significant opportunities, the analysis of unstructured data introduces complexities and potential risks [4]. Ultimately, the successful application of AI in marketing holds the promise of improving customer satisfaction and fostering business growth.

Leading companies like Amazon and Domino's Pizza are utilizing AI to optimize delivery logistics and customer service. Similarly, Red Balloon and Macy's are employing AI-driven platforms to enhance customer service and personalize purchasing experiences, while Lexus and Affectiva are exploring AI's potential in designing emotionally resonant advertising campaigns

[5]. Despite the growing adoption of AI in marketing practice, academic research rigorously examining its effective application remains limited. This gap in the literature motivates the present study.

This research addresses the need for empirical investigation into predictive analytics in marketing, with a specific focus on evaluating its effectiveness in driving customer engagement. Marketing, defined as the process of promoting, selling, and distributing products or services, encompasses understanding customer needs, creating value propositions, and effectively communicating these offerings to target audiences [6]. This study innovatively applies a suite of machine-learning techniques to analyze customer response data acquired from a real-world marketing campaign within a retail superstore environment. To ensure data quality, the dataset undergoes meticulous pre-processing, including robust methods for handling missing values, identifying and mitigating outliers, and feature engineering to maximize the informative value of the data.

Machine Learning (ML), a subset of AI, provides the capability for computers to learn from data and improve performance autonomously, without explicit programming [7]. This study leverages the predictive power of several ML algorithms: Decision Trees, Logistic Regression, Random Forest,

*Corresponding Author: Khawla Alhasan, Arab Open University – Kuwait E-mail: k.alhasan@arabou.edu.kw

www.astesj.com

<https://dx.doi.org/10.25046/aj100306>

K-Nearest Neighbour, and Support Vector Classifier. A key methodological contribution is the careful balancing of the dataset to overcome potential biases introduced by class imbalance, a common challenge in customer response data. The performance of the predictive models is rigorously evaluated using a comprehensive set of metrics, including accuracy, precision, and recall, providing a nuanced understanding of each model's strengths and weaknesses.

This research seeks to answer the following key questions:

RQ 1: How effectively can machine learning techniques analyze customer response data derived from a marketing campaign within a retail superstore setting?

RQ 2: Among the machine learning models of Decision Trees, Logistic Regression, Random Forest, K-Nearest Neighbor, and Support Vector Classifier, which demonstrate optimal predictive performance for customer responses to marketing campaigns?

RQ 3: What are the implications of these findings for the optimization of marketing strategies in retail environments, specifically concerning enhancements to customer engagement and the facilitation of business growth? What theoretical contributions does this study offer to the understanding of AI's role in marketing?

The subsequent sections of this paper detail the related literature, the research methodology employed, the interpretation of the developed models, a thorough analysis of the results, and a discussion of the theoretical and practical implications of the findings

2. Literature Review

This section provides a comprehensive literature review of recent studies focusing on the assimilation of artificial intelligence and machine learning techniques in various aspects of marketing and customer experience management. The section also discusses the findings and concludes the aim of this research.

2.1. Previous work

The literature review presents a comprehensive exploration of recent studies focusing on the assimilation of artificial intelligence and machine learning techniques in various aspects of marketing and customer experience management. These studies explore the potential of AI and ML to revolutionise customer interactions, optimise marketing strategies and support business outcomes in general.

In [8], the authors assessed the field of customer experience management (CEM) and investigated how AI and machine learning can autonomously upgrade customer experiences. They examined the challenges associated with implementing AI-driven solutions concerning CEM and proposed strategies to establish critical business drivers using AI and ML techniques. By employing these technologies, businesses can better understand customer needs and preferences, resulting in more personalised and effective customer interactions.

In [9], the authors focused on the application of AI-driven chatbots with Natural Language Processing (NLP) capabilities to improve customer experiences. They demonstrated how these chatbots can engage with customers in natural language conversations, providing timely assistance and support. By harnessing NLP and AI, businesses can automate customer service

processes, reduce response times and improve overall customer satisfaction.

In [10], the authors investigated the role of AI and ML algorithms in processing large volumes of data to make informed marketing decisions. Their study highlighted how AI-powered data analysis can uncover valuable insights into customer behaviour, market trends and competitor strategies. By employing AI-driven data processing, businesses can formulate more accurate and effective marketing strategies, ultimately generating better business outcomes.

The application of AI in marketing is explored by [11], specifically focusing on understanding and analysing customer habits, preferences and behaviours. They discussed how AI-powered analytics can assist businesses to gain further information into customer segments, allowing more targeted and personalised marketing. By exploiting AI, businesses can reinforce customer relationship management roles and tailor marketing strategies to individual customer needs.

In [12], the authors examined the evolution of CRM functions via the adoption of Artificial Intelligence User Interface (AIUI). Their research illustrated how AIUI technologies can simplify customer interactions, improve response times and improve overall customer satisfaction. By integrating AIUI into CRM systems, businesses can provide more efficient and personalised customer support, promoting robust customer relationships and increased loyalty.

In [13], the authors focused on the transformation of traditional retail stores into smart retail stores via the integration of AI and IoT technologies. The authors explored how AI and IoT can enhance customer experiences by providing personalised recommendations, streamlining checkout processes and optimising inventory management. By incorporating smart retail solutions, businesses can produce more engaging and convenient shopping experiences for their customers, generating increased sales and loyalty.

In [14], the researchers investigated advanced AI applications in e-commerce, particularly focusing on AI-supported machines that are capable of tracking human emotions. Their research highlighted how these AI-driven technologies can improve consumer-brand associations and boost product recommendations in e-commerce settings. By exploiting AI to better understand and respond to customer preferences, businesses can establish more tailored and engaging online shopping experiences, ultimately generating higher conversion rates and customer satisfaction.

2.2. Discussion

The previous work, while providing valuable recommendations and findings, lacks a critical analysis of the studies, focusing instead on a broad overview of AI and ML applications in marketing. It overlooks the challenges and limitations of these techniques, particularly in the context of customer response data analysis. Additionally, the review fails to integrate the findings of the individual studies into a cohesive whole, hindering a deeper understanding of the topic.

This research aims to investigate the practical implementation of AI and ML in enhancing customer experiences. While the potential benefits of these technologies are clear, real-world application faces significant challenges. This study seeks to

address these challenges by providing actionable insights and strategies to facilitate the successful integration of AI and ML into business operations. By overcoming these hurdles, businesses can improve customer satisfaction and drive sustainable growth

3. Research Methodology

This section details the methodology employed to investigate customer response prediction. The approach encompasses data collection and preprocessing, data exploration and visualization, model development and evaluation, and subsequent analysis. Each stage is designed to ensure the development of a robust and effective customer response detection model.

3.1. Data Set

The dataset utilized in this research was obtained from Kaggle. It comprises customer attributes relevant to demographics, purchasing behaviors, and interactions within a retail superstore, collected during the previous year's campaign. The dataset includes customer age, education level, marital status, household composition, income, and purchasing history across various product categories over the past two years. Additionally, it incorporates data on customer complaints, enrollment dates, and recent purchases. The application of this dataset and the construction of a predictive model aim to enhance the effectiveness of the superstore's marketing campaign, specifically targeting existing customers, thereby reducing promotional costs while maximizing revenue potential.

3.2. Data Acquisition and Exploration

The research commenced with the importation of essential libraries, including Pandas, NumPy, Matplotlib, Seaborn, and datetime. The dataset, named 'superstore_data.csv', was loaded using Pandas, and its dimensions (number of rows and columns) were printed for initial assessment. Furthermore, the first few rows of the dataset were displayed to gain insight into its structure.

Pandas functions, such as columns, duplicated, isnull, and info, were used to scrutinize the dataset columns, duplicated rows, and null values. Observations indicated that the 'Income' column contained null values, and the 'Dt_Customer' column was of Object type instead of the Date type.

3.3. Data Preparation

Data preparation involved several steps to ensure dataset integrity and suitability for analysis. First, the 'Dt_Customer' column was standardized by converting its format to Date, replacing '/' with '-', and applying the pd.to_datetime function. Missing values in the 'Income' column were imputed using the median value. Outlier detection and management were conducted on 'Year_Birth' and 'Income' by filtering out birth years before 1900 and excessively high-income values beyond a predetermined threshold. Additional outliers were identified and removed from columns associated with product purchase amounts to maintain a distribution representative of realistic consumer behavior.

3.4. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) involved employing various techniques to comprehend the data's distribution and relationships among variables. Pie charts were used to exhibit the distribution of positive and negative responses from the marketing

campaign (Fig. 1). Box plots illustrated the relationships between categorical variables, such as education and marital status, and campaign response. Additionally, box plots underlined the correlations between continuous variables, such as expenditure on different products, and campaign response (Figs. 2 and 3). This analysis provided insights into customer behavior and campaign effectiveness, informing subsequent investigation and modeling.

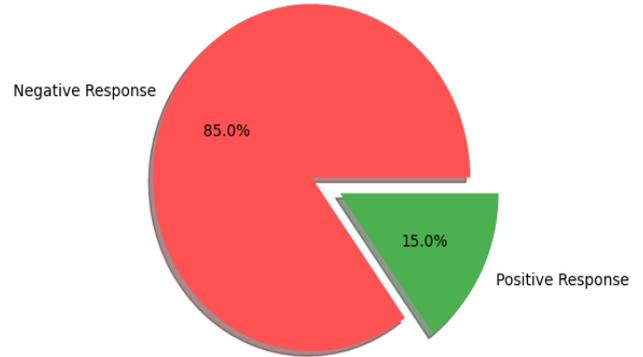


Figure 1: Percentage of negative and positive responses gathered from the marketing campaign

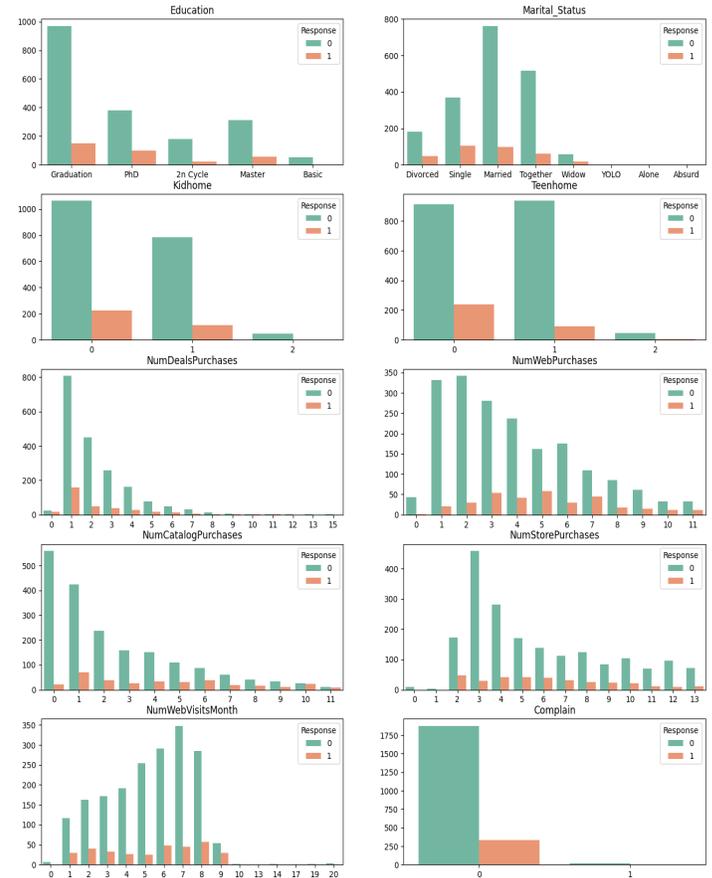


Figure 2: Bar plots illustrating the relationships between the categorical variables and continuous variables with campaign responses

3.5. Preprocessing and Modelling

To prepare the data for effective machine learning, several preprocessing steps were undertaken. Initially, date-type columns were converted to integer representations. This conversion streamlines both the analytical process and the computational

efficiency of subsequent modeling, as machine learning algorithms typically handle numerical data more efficiently. Redundant columns, deemed irrelevant to the predictive task (e.g., 'Id'), were removed to reduce dimensionality and potential noise in the data. This step is crucial for improving model performance by focusing on the most relevant features. Categorical variables were then transformed into a numerical format using one-hot encoding. This technique was chosen to enable machine learning algorithms to effectively process categorical information by creating binary columns for each category.

Prior to model training, the dataset underwent standardization. Standardization scales all features to have a mean of zero and a standard deviation of one. This process ensures that all features contribute equally to the model training process and prevents any single feature from unduly influencing the model due to its scale, which is crucial for algorithms sensitive to feature scaling. To mitigate class imbalance, a common issue in classification tasks, Random Oversampling was employed. This technique was selected to balance the class distribution by duplicating instances from the minority class, providing the model with more training examples for the under-represented class and potentially improving predictive accuracy, particularly for the minority class.

Finally, the preprocessed dataset was partitioned into training and testing sets. This partitioning allows for rigorous evaluation of the model's predictive performance on unseen data, providing an unbiased estimate of its generalization ability and preventing overfitting, thus providing a realistic assessment of its performance in a real-world setting.

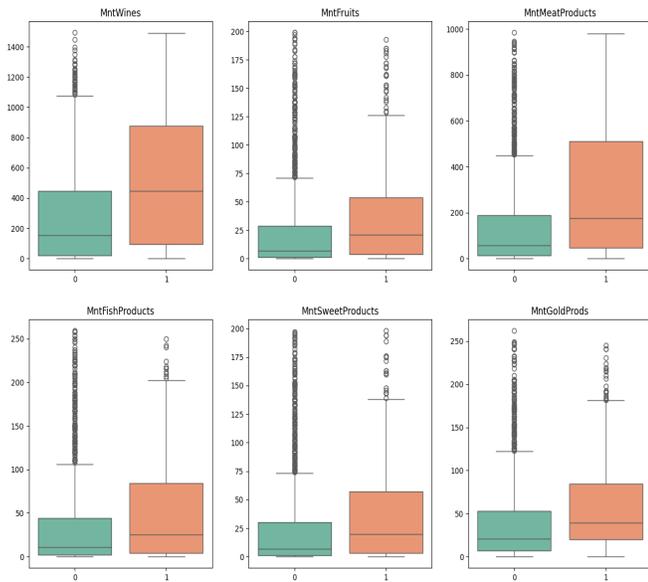


Figure 3: Box Plots illustrating the relationships between the categorical variables and continuous variables with campaign responses

3.6. Model Selection and Evaluation

Classification algorithms, including Decision Tree, Logistic Regression, Random Forest, KNN, and SVC, were applied. Model hyperparameters were optimized using GridSearchCV. Model evaluation metrics, including precision, recall, and accuracy, were calculated using cross-validation and confusion matrices. The Random Forest model demonstrated the best performance in terms of precision, recall, and accuracy.

4. Results Analysis

This section clarifies the results achieved with Random Forest, Decision Trees, Support Vector Classifier, K-Nearest Neighbour and Logistic Regression machine learning models.

4.1. Random Forest

Random Forest exhibited an exceptional performance across each of the evaluation metrics. Fig. 4 shows the confusion matrix of the Random Forest ML technique that report a precision score of 0.952, it accurately identified positive instances, reducing false positives.

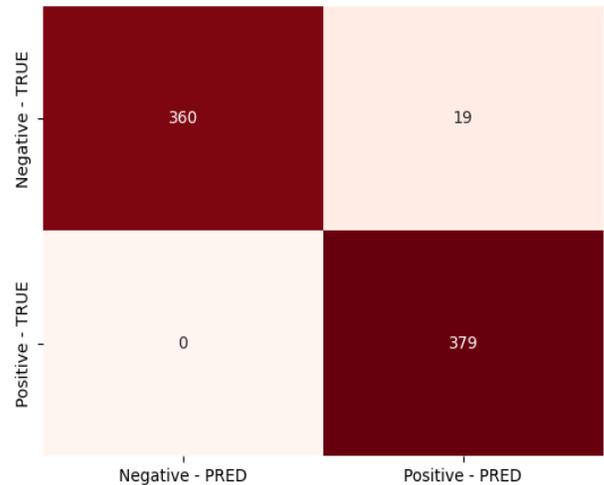


Figure 4: Random Forest Confusion Matrix

The perfect recall score of 1.000 signifies that it successfully described each true positive case, demonstrating its ability to comprehensively detect positive outcomes. Similarly, Random Forest achieved an impressive accuracy score of 0.975, reflecting its overall correctness in predicting both positive and negative instances. Hence, this denotes that Random Forest is a robust model, capable of accurately identifying positive cases while maintaining high accuracy levels.

4.2. Decision Tree

Decision Tree exhibited a strong performance, particularly in recall and precision. With a precision score of 0.883, it effectively identified true positive instances with a relatively low rate of false positives. Additionally, the recall score of 0.992 denotes high sensitivity to positive cases, obtaining the majority of true positives. Fig. 5 shows the confusion matrix of the Random Forest ML technique. Although slightly lower than Random Forest pertaining to accuracy, Decision Tree achieved a creditable accuracy score of 0.930, revealing its capability to make correct predictions throughout the dataset. Basically, Decision Tree's performance strengthens its effectiveness in accurately identifying positive outcomes while maintaining competitive accuracy levels.

4.3. Support Vector Classifier

Support Vector Classifier demonstrated a solid performance across the precision, recall and accuracy metrics. With a precision score of 0.891, it exhibited a high level of accuracy in identifying positive instances while minimising false positives. Fig. 6 illustrates the Support Vector Classifier Confusion Matrix. Its recall score of 0.966 signifies that SVC effectively acquired the

majority of true positive cases, indicating its robustness in detecting positive outcomes.

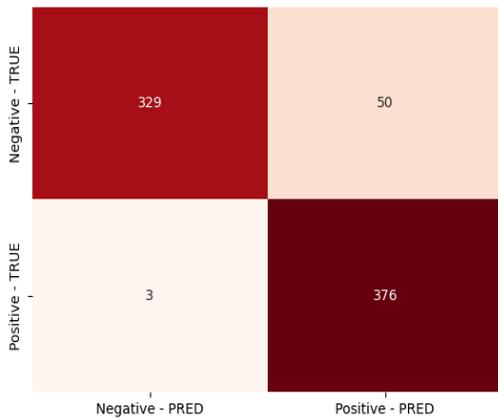


Figure 5: Decision Tree Confusion Matrix

Despite a slightly lower accuracy score compared to Random Forest and Decision Tree, SVC achieved a commendable accuracy score of 0.923, reflecting its overall correctness in predicting both positive and negative instances. This emphasises SVC as a reliable model to accurately predict positive outcomes with competitive accuracy levels.

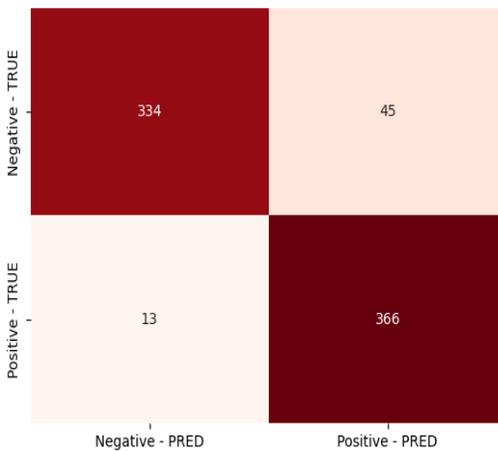


Figure 6: Support Vector Classifier Confusion Matrix

4.4. K-Nearest Neighbour

K-Nearest Neighbour displayed a notable performance, particularly regarding recall, indicating its effectiveness in obtaining positive instances. Fig. 7 illustrates the Confusion Matrix of the k-Nearest Neighbour ML technique that reports a precision score of 0.843. It accurately identified positive cases while maintaining a reasonable false positive rate. Its high recall score of 0.992 denotes that KNN effectively recorded the majority of true positive instances, displaying its sensitivity to positive outcomes. Notwithstanding a slightly lower accuracy score compared to other models, KNN attained a respectable accuracy score of 0.904, reflecting its ability to make correct predictions throughout the dataset.

Principally, KNN's performance confirms its efficacy in accurately identifying positive outcomes while maintaining competitive accuracy levels.

4.5. Logistic Regression:

While Logistic Regression established a slightly lower performance compared to other models, it demonstrated reasonable accuracy in predicting positive outcomes.

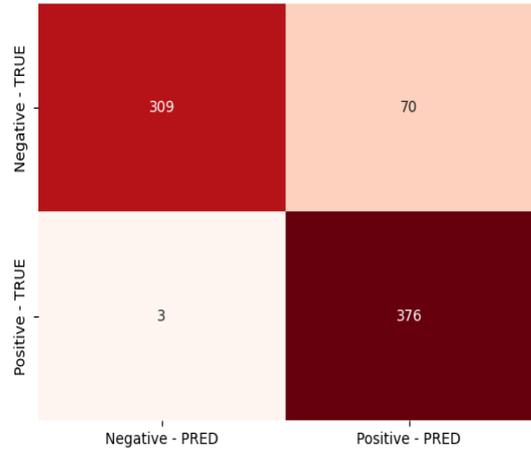


Figure 7: K-Nearest Neighbour Confusion Matrix

Fig. 8 illustrates the Confusion Matrix of the Logistic Regression ML technique that reports a precision score of 0.793, it accurately identified positive instances with a moderate rate of false positives. The recall score of 0.836 indicates that Logistic Regression effectively represented a good portion of true positive cases, demonstrating its sensitivity to positive outcomes. Despite a lower accuracy score in comparison to other models, Logistic Regression achieved a reasonable accuracy score of 0.809, revealing its overall correctness in predicting both positive and negative instances. Essentially, Logistic Regression provides a reliable approach that can be applied to predict positive outcomes, albeit with a slightly lower performance compared to other models.

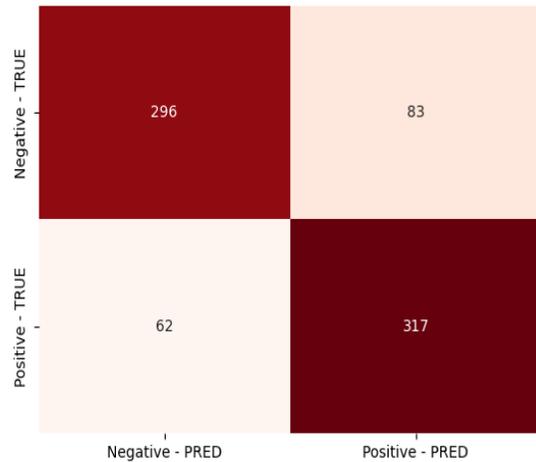


Figure 8: Logistic Regression Confusion Matrix

Table 1: Comparative Analysis of Precision, Recall, Accuracy

Model	Precision	Recall	Accuracy
Random Forest	0.952	1.0	0.975
Decision Tree	0.883	0.992	0.93
SVC	0.891	0.966	0.923
KNN	0.843	0.992	0.904
Logistic Regression	0.793	0.836	0.809

Classification reports deliver a detailed breakdown of each model's performance, permitting a comprehensive evaluation. Table 1 and Fig.9 provides a performance comparison between all the ML techniques addressed in the study.

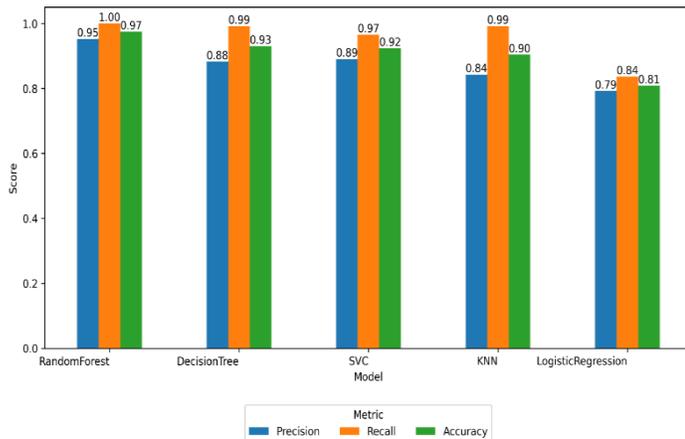


Figure 9: Comparative Performance of Different Models

5. Conclusion and Future Work

The evaluation of multiple machine learning models to predict customer response has yielded significant insights into the effectiveness of these techniques within marketing contexts. Notably, the Random Forest model demonstrated superior predictive capabilities, achieving a high accuracy of 97.5%, a precision of 95.2%, and a recall of 100%. This strong performance underscores Random Forest's ability to accurately identify both positive and negative customer responses, highlighting its potential as a valuable tool for marketing professionals. Decision Tree and Support Vector Classifier (SVC) also exhibited robust predictive power, further validating the utility of machine learning in this domain, while K-Nearest Neighbors (KNN) provided reasonably accurate predictions, albeit with slightly lower performance metrics. In contrast, Logistic Regression showed the least effective performance, with an accuracy of 80.9%, a precision of 79.3%, and a recall of 83.6%.

These findings collectively emphasize the capacity of machine learning techniques to effectively analyze customer response data in marketing scenarios. The Random Forest model's exceptional recall, in particular, suggests its strength in capturing a high proportion of positive customer responses, which is critical for targeted marketing interventions. The comparative performance analysis of the models provides valuable guidance for selecting appropriate algorithms for specific marketing objectives. Ultimately, this study reinforces the importance of leveraging advanced analytics to enhance the precision and efficacy of marketing decision-making processes and customer targeting strategies, leading to improved resource allocation and campaign ROI.

However, to further refine predictive accuracy and expand the applicability of these models, future research should prioritize several key areas. First, exploring ensemble methods, which strategically combine the predictive strengths of multiple individual models, holds promise for achieving even higher predictive performance. Second, a systematic investigation of feature engineering and selection techniques is warranted to optimize model inputs and potentially uncover more salient predictors of customer behavior. This could provide valuable

insights into which customer attributes are most influential in driving campaign response. Third, integrating domain-specific knowledge and external data sources, such as social media activity, sentiment analysis, and broader demographic information, could enrich the models' predictive capabilities and provide a more holistic understanding of customer behavior. Finally, conducting longitudinal studies to assess the temporal stability and generalizability of the developed models is essential to validate their robustness and effectiveness in dynamic real-world marketing environments. Such studies would contribute to a more comprehensive understanding of the models' long-term utility and inform adaptive marketing strategies.

Acknowledgment

This research was supported and funded by the research sector, Arab Open University- Kuwait Branch under decision number 25128.

References

- [1] Y. Kulkarni; A. Mahamuni; S. Sane; P. Kalshetti; K. Patil; R. Jarad, "Analysing How Marketing Management and Artificial Intelligence are used to Change Customer Engagement," in 2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies, 1-6, 2024, doi: 10.1109/TQCEBT59414.2024.10545064.
- [2] N. Rane, M. Paramesha, S. Choudhary, J. Rane, "Artificial Intelligence in Sales and Marketing: Enhancing Customer Satisfaction, Experience and Loyalty," *Advances in Artificial Intelligence*, 17, 2024. doi: 10.18178/JAAI.2024.2.2.245-264.
- [3] K. Dhiwar, M. Bedarkar, "Circular Economy for Fashion Waste in the Indian Fashion Industry: Challenges and Opportunities," *Fashion Practice*, 16(3), 344, 2024. doi.org/10.1080/17569370.2024.2393495.
- [4] R. Gill, P. Kumar, M. Patel, H. Kumar, "Integrating Marketing Data Ecosystems: Merging Diverse Data Sources for Holistic Insights," *Data Engineering for Data-driven Marketing*, Emerald, 2025.
- [5] S. Verma, R. Sharma, S. Deb, and D. Maitra, "Artificial intelligence in marketing: Systematic review and future research direction," *Information Management Data Insights*, 1, 100002, 2021. doi.org/10.1080/17569370.2024.2393495.
- [6] M. Shaffer, I. Okundaye, *Marketing, Sales, and Distribution*, Innovation in Nephrology, Elsevier, 2025.
- [7] H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Computer Science*, 2, 160, 2021. doi: 10.1007/s42979-021-00592-x.
- [8] H. Gacanin, M. Wagner, "Artificial intelligence paradigm for customer experience management in next-generation networks: Challenges and perspectives," *IEEE Network*, 33(2), 188-194, 2019. doi: 10.1109/MNET.2019.1800015.
- [9] Q. Nguyen, A. Sidorova, "Understanding user interactions with a chatbot: a self-determination theory approach," in 2018 AMCIS Proceedings, 3, 2018.
- [10] A. Maxwell, S. Jeffrey, M. Lévesque, "Business angel early-stage decision making," *Business Venturing*, 26(2), 212-225, 2011. doi.org/10.1016/j.jbusvent.2009.09.002.
- [11] S. Chatterjee, S. Ghosh, R. Chaudhuri, B. Nguyen, "Are CRM systems ready for AI integration? A conceptual framework of organizational readiness for effective AI-CRM integration," *The Bottom Line*, 32(2), 144-157, 2019. doi.org/10.1108/BL-02-2019-0069.
- [12] R. Seranmadevi, A. Kumara, "Experiencing the AI emergence in Indian retail – Early adopters approach," *Management Science Letters*, 9, 33-42, 2019. doi: 10.5267/j.msl.2018.11.002.

- [13] W. Wisetsri, C. Aarthy, V. Thakur, D. Pandey, K. Gulati, "Systematic Analysis and Future Research Directions in Artificial Intelligence for Marketing," *Computer and Mathematics Education*, **12**(11), 43-55, 2021. doi: 10.17762/turcomat.v12i11.5825.
- [14] S. Nazim, M. Rajeswari, "Creating a Brand Value and Consumer Satisfaction in E-Commerce Business Using Artificial Intelligence with the Help of Vosag Technology," *Innovative Technology and Exploring Engineering (IJITEE)*, **8**(8), 2019.

Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).