Integrating Supervised Learning Paradigms for Enhanced Marketing Campaign Outcome Prediction

Abstract - **Businesses are relying heavily on marketing to grow and hence need to evaluate their marketing strategies to get better responses from customers. Machine learning techniques are widely being adopted to perform data-driven predictions for the success rate of these marketing campaigns. This study uses the Extreme Gradient Boosting (XGBoost) model to predict the likelihood of the marketing campaign's success based on various complex customer and product factors. To improve the prediction accuracy, the suggested framework combines feature engineering and ensemble learning methods. Important contributing aspects are methodically examined, including customer annual income, age, credit score, product type, product price range, advertisement intensity, and discount levels. And new metrics to further assess the hidden factors behind customers responses, like the affordability ratio amongst a few others are feature engineered. The model is trained and verified on a synthesized and balanced dataset to simulate real campaign conditions. The model has achieved an accuracy score of 0.8300. The proposed model helps businesses make reliable prediction across different and complex factors that affect customer response and boost their marketing performance and hence overall business.**

Keywords - Marketing campaign, XGBoost, customer response, machine learning, success prediction, consumer behaviour.

I. Introduction

Businesses rely heavily on marketing to acquire customers. They have implemented many surveys and collected a large amount of data in-order to facilitate analysis to improve marketing. But traditional analysis is done on linear relationship between the data which is not sufficient and often misleading. Machine learning can identify hidden patterns and complex relations in the data to predict results like success of the campaign, client responses, etc. Predictive models look at various factors like what kind of product, whether it is a necessity or a luxury, the price of the product, the income of the buyer, the market demand, the discounts offered, the credit score of the buyer, the amount of ads the client came across via different mediums like calls, SMS, social media, TV, etc., and derive strong patterns and relations to facilitate accurate predictions. Linear models like Logistic Regression fails to analyse these complex details simultaneously to provide results.

The proposed method particularly uses a non-linear model that combine statistics and ensemble learning to find and analyse these required complex relations. It even includes certain interpretable features like calculating affordability ratio out of product price and annual income of client and compare the new results with success rates to provide visible linear relations.

The model's goal is to provide businesses with a reliable prediction model that they can use to test and improve their marketing strategies in a simulated database before they actually invest to implement it in real. Saving on time, money and other resources and improving overall business. This model has been trained on a synthesized dataset that is made to reflect realistic customer behaviour and provide balanced and normalized data to train the model for better performance in accuracy, recall, precision and other metrics compared to other linear models and make it practically viable. The business field is highly competitive and having a powerful tool, like the proposed model, to predict the success of marketing campaigns and customer responses can help businesses to iterate and test ideas quickly, and gain an upper-hand against their competitors.

The model can be used to identify and analyse responses thoroughly in individual segments like based on customer demographics and income range, product price and type etc., which will provide more insights into designing personalised campaigns for each of these segments and may boost overall campaign success. More attributes like past behaviour, personality, marital status, profession, etc., could later be added to analyse and train the model to further improve its predictions and account for realistic conditions that affect the campaign and customer responses.

II. Literature Review

Y. Singh et al. [1] used machine learning models like Gradient Boosting (GB), Ada Boost (AB), Cat Boost (CB), Voting Classifier (VC), Stacking Classifier (SC), K Nearest Neighbours (KNN), Random Forest (RF) and Logistic Regression (LR) to analyse consumer personality to predict the performance of marketing campaigns using a normalized dataset. Their research revealed Stacking Classifier to be the best performing model with accuracy of 0.8988. While the model achieved a high accuracy, the minority response classes were not properly identified by this model. Hence, this was not reliable for real-world applications.

B. Jamalpur et al. [2] used different methods like decision trees (DT), random forest (RF), logistic regression (LR), CNN, LSTM for predicting customer lifetime value (CLV) and customer segmentation. The provided comparative results for the models across metrics like Recall, Precision and F1 score for three classes - 0, 1, 2. The method was not flexible across many marketing sectors because it was computationally costly and relied heavily on organized e-commerce datasets.

S. E. Saeed et al. [3] has forecast client subscription to banking products using an ensemble classifier based on hybrid machine learning models like Logistic Regression, Naïve Bayes, K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Decision trees and Random Forest and used 10-folds cross-validation. The Random Forest classifier outperforms others with accuracy of 94.02%. The ensemble was not appropriate for real-time analytics and quick campaign modifications due to its high computational demand and significant preprocessing.

M. Semwal et al. [4] presented a study using Support Vector Machine (SVM) and Business Intelligence (BI) in order to maximize dynamic pricing strategy success in correspondence to the changing e-commerce market. The proposed method has proven to be better across various metrics like accuracy, sensitivity, speciality, F-measure and implementation costs compared to other older methods like BI, ML-BI, etc., The model relied heavily on constant real-time data availability.

C. Senjaya et al. [5] used XGBoost to analyse two years' worth of historical data to help micro, small and medium enterprises (MSME) retain their customers and increase their revenue. It determined dynamic pricing revenue strategies. The proposed XGBoost achieved highest accuracy with R2 = 0.9995 compared to other models like Linear Regression, Random Forest. The study's limited business emphasis and limited dataset diversity limited its applicability to broader or cross-industry contexts.

A. M. Abdullahi et al. [6] used the Random Forest algorithm in RapidMiner to categorize clients according to their behavioural and demographic characteristics. Automated segmentation was highlighted in the study as a means of enhancing targeted marketing campaign's results. However, due to overfitting and poor dataset quality, the model only attained 50.12% accuracy, indicating the need for improved feature selection and hyperparameter optimization.

N. Kodikara et al. [7] in order to find persuasive consumers who are most likely to react favourably to marketing initiatives, coupled dual uplift modelling with XGBoost. Uplift and effectiveness were visualized using the Qini curve, which demonstrated increased targeting accuracy. But the model needed a lot of processing power and big datasets, it wasn't as useful for smaller businesses with less computer infrastructure.

J. A. Frempong et al. [8] used four classifiers – Multilayer Percepton Neural Network (MLPNN), Decision Tree, Logistic Regression and Random Forest in order to predict term deposit subscriptions in bank through telemarketing. The research revealed that Random Forest is the best performing classifier with accuracy 92.7%. The study did not properly handle unbalanced data and feature redundancies.

III. Existing Methods

Current marketing campaign predictions only use customer behaviour and demographics using linear machine learning models like Decision Trees, Logistic Regression or Random Forest, which can't deal with complicated, hidden relations in the raw data. Also, the existing methods relied more on real data which were often unbalanced and inconsistent. Wherein, ensemble methods like XGBoost, Gradient Boosting handle unbalanced data, and deliver better accuracy. But even they require a lot of computing resources.

IV. Methdology

This study collects some essential input factors and uses statistical preprocessing, feature development, model training through an organized, data-driven workflow to predict campaign performance. Essential marketing and demographic data, such as advertisement rates across different mediums like TV, SMS, calls, social media, product kind, whether luxury or necessity, customer annual income range, product price range, discount % offered, credit score of customer and age are gathered at the beginning of the process. These inputs are essential for calculating customer behaviour and predicting the probability of response to the campaign and hence its success. A synthesized dataset was used for this to provide balanced data. First, data is prepared and encoding techniques are used to transform raw input into usable numerical values for analysis and model training.

After preprocessing stage, feature engineering was used to define new relations that improve the prediction model like the affordability ratio (price-to-income ratio), discount intensity, normalized advertisement intensity, credit-afford interaction, income-to-credit and discount-to-affordability. The new engineered features help the model predict the probability of a customer responding to the campaign more accurately by establishing links between customer affordability, product cost, and advertisement intensity.

We used Extreme Gradient Boosting (XGBoost) model to train our model. This enabled the model to handle large dataset and identify non-linear relations between attributes effectively. We also used 5-folds cross-validation and hyperparameter tuning to enhance overall predictions. The model is evaluated using metrics like Accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). The decision threshold is set to 50%. That means, if the predicted probability is greater than or equal to 50%, the campaign is considered successful i.e., the client will respond positively.

After successful evaluation, we used Python Tkinter to deploy the Graphical User Interface (GUI) for the model. Businesses can input values like advertisement intensity - low, medium and high, product category - luxury or necessity, customer income range, product price range, discount offered, credit score of customer and age of customer using the interface's dropdown menus and sliders. It displays the binary results as "YES" for "likely to respond/campaign successful" and "NO" for "not likely to respond/campaign unsuccessful" and additionally even displays the predicted success probability for better interpretability of the results.

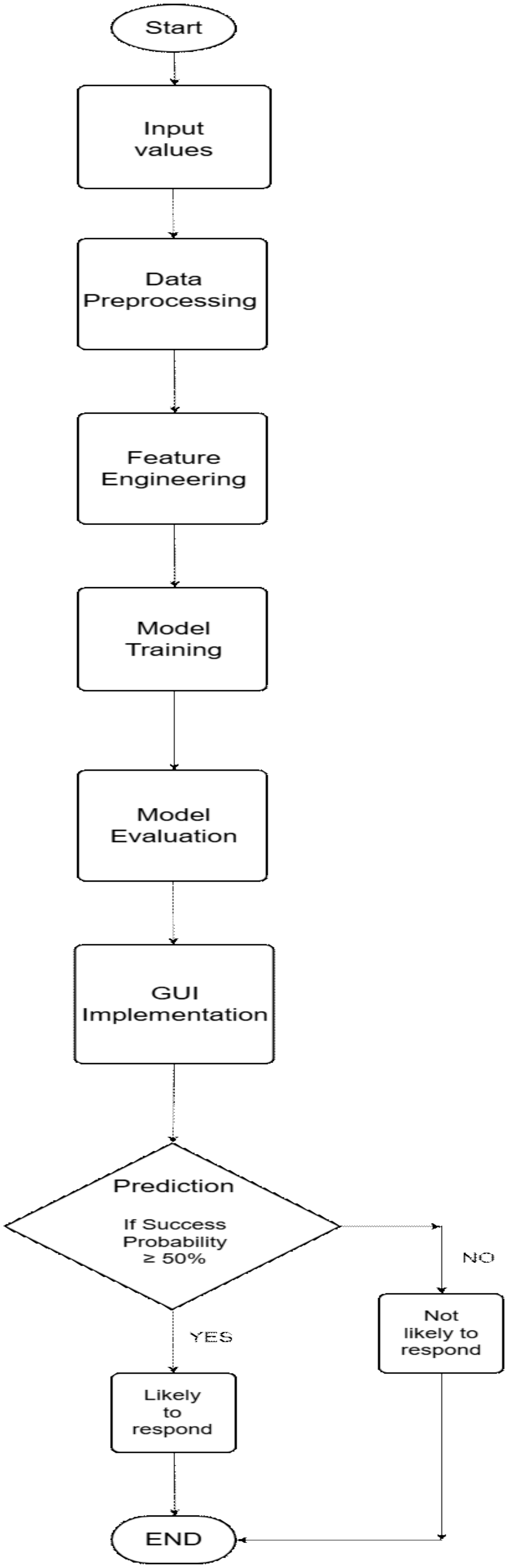

Figure 1: Methodology Flowchart

V. RESULTS AND DISCUSSION

The model achieved an accuracy of 0.8300, precision of 0.8658 and recall of 0.9405 with F1-score of 0.9016. ROC-AUC of 0.7959 and PR-AUC of 0.9489. Feature importance analysis has revealed Log Income, Annual Income, and Affordability Ratio to be important predicting factors for the campaigns' success. The confusion matrix shows an accurate prediction of 1090 "YES" results with few false negatives.

Result of Model training:

1) XG Boost model evaluation:

accuracy	: 0.8300
precision	: 0.8658
recall	: 0.9405
f1_score	: 0.9016
roc_auc	: 0.7959
log_loss	: 0.3758
pr_auc (avg precision)	: 0.9489
balanced_accuracy	: 0.6196
matthews_corrcoef	: 0.3001

Figure 2: Model evaluation metrics

2) Classification report

Classification Report:

	precision	recall	f1-score	support
0	0.5106	0.2988	0.3770	241
1	0.8658	0.9405	0.9016	1159
accuracy			0.8300	1400
macro avg	0.6882	0.6196	0.6393	1400
weighted avg	0.8046	0.8300	0.8113	1400

Figure 3: Classification report

3) Feature Importances

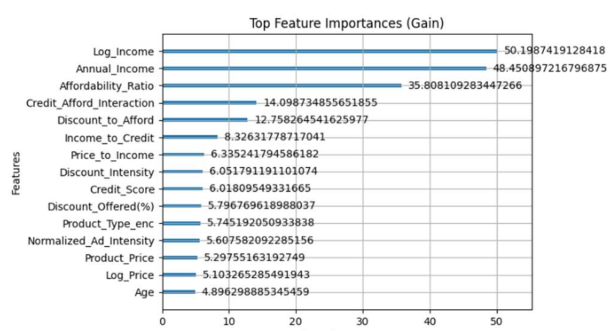

Figure 4: Feature Importance report

GUI:

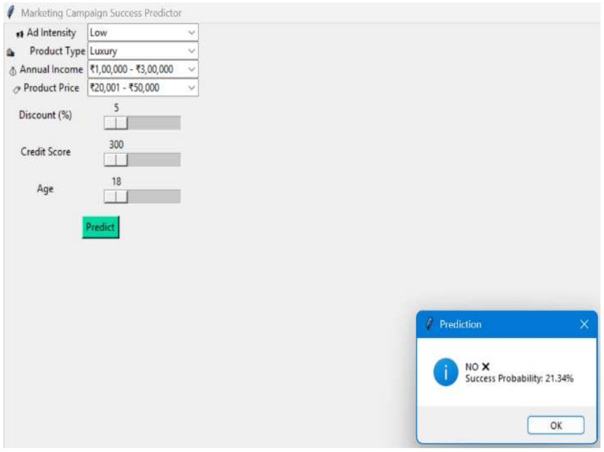

Figure 5: GUI with NO result

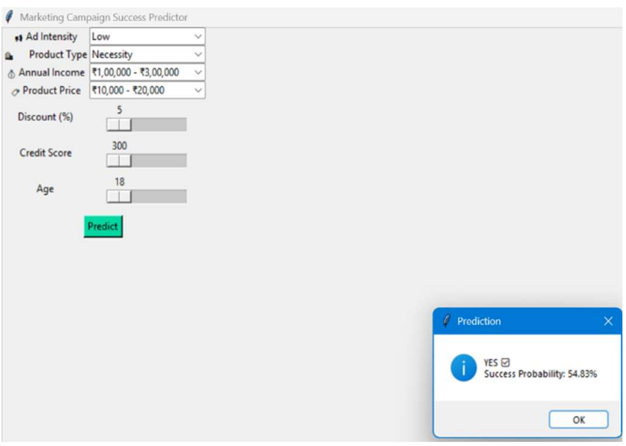

Figure 6: GUI with YES result

Graphs:

1) Confusion Matrix:

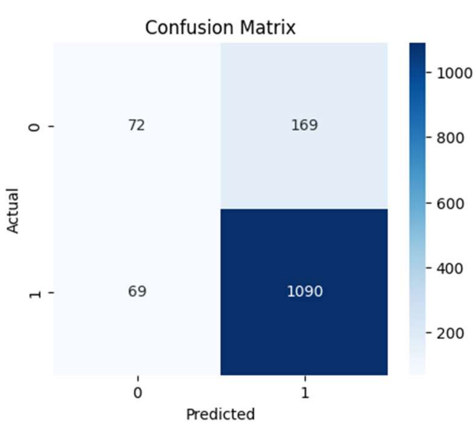

Figure 7: Confusion Matrix

2) ROC Curve:

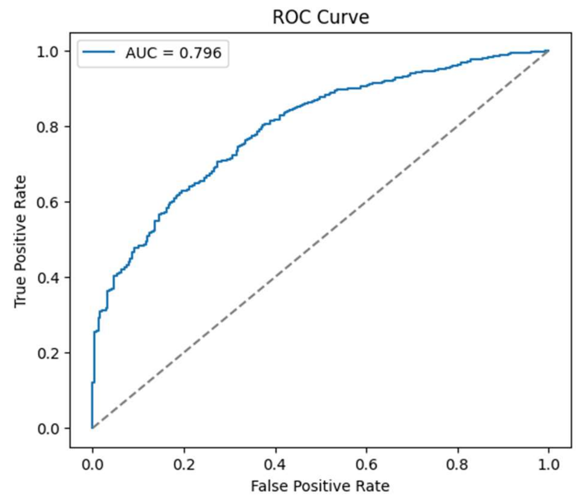

Figure 8: ROC Curve

3) Precision-Recall Curve:

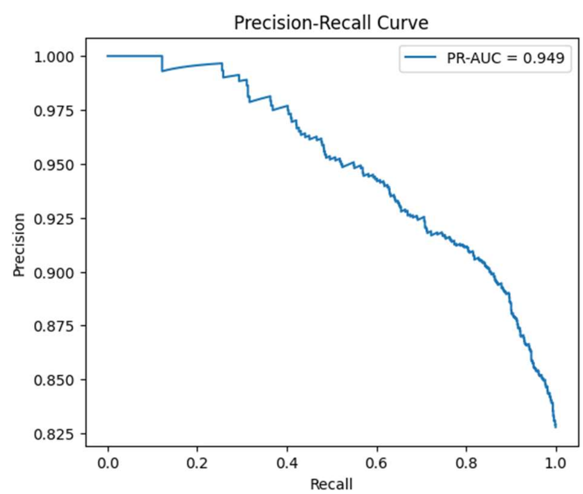

Figure 9: Precision-Recall Curve

VI. CONCLUSION

The XGBoost algorithm was used to predict marketing campaign success based on various customer and product data. To effectively predict the success probability of the campaign and response of the customers the model focused on key factors like the customer's income, product's price, discount offered, advertisement intensity, age, credit score, etc. It delivered an accuracy of 83.00%, and the PR-AUC reached 0.9489. Feature engineering fields like affordability ratio and credit afford interaction, normalized advertisement intensity, etc., improved the model's performance.

Through this study, we can conclude that while there are a lot of factors that affect campaign success the most important ones are linked directly to finances like customer annual income, credit score and the product price. The GUI made it possible to use the model easily and determine the results quickly with binary “YES” and “NO” results along with the predicted success probabilities for better interpretability. These predictions would significantly improve business marketing performance and give them various insights into the factors affecting the campaign’s success and customer responses that can be used to grow their business.

VII. REFERENCE

[1] Y. Singh, J. J. Jena, M. K. Gourisaria, J. P. Singh, V. Kumar, and D. K. Behera, "Predicting Marketing Campaign Effectiveness through Consumer Personality Analysis using Machine Learning," 2024 International Conference on Intelligent Computing and Sustainable Innovations in Technology (IC-SIT), Bhubaneswar, India, 2024, pp. 1–8, doi: 10.1109/IC-SIT63503.2024.10862130

[2] B. Jamalpur, D. Singh, B. S. Kumar, A. Nagpal, S. Pawar, and D. Banerjee, "Applications of Deep Learning in Marketing Analytics: Predictive Modeling and Segmenting Customers," 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE), Warangals, India, 2024, pp. 1473–1477, doi: 10.1109/IC3SE62002.2024.10593465

[3] S. E. Saeed, M. Hammad, and A. Alqaddoumi, "Predicting Customer’s Subscription Response to Bank Telemarketing Campaign Based on Machine Learning Algorithms," 2022 International Conference on Decision Aid Sciences and Applications (DASA), Bahrain, 2022, pp. 1474–1479, doi: 10.1109/DASA54658.2022.9765152

[4] M. Semwal, K. Akila, M. Manasa, P. Sai Raj, Y. Motukuru, and P. Karthik, "Machine Learning-Enabled Business Intelligence For Dynamic Pricing Strategies In E-Commerce," 2024 2nd International Conference on Disruptive Technologies (ICDT), Vaddeswaram, India, 2024, pp. 116–120, doi: 10.1109/ICDT61202.2024.10489724

[5] C. Senjaya and I. D. Sudirman, "Price Adjustment for Micro, Small, and Medium-Sized Enterprise Business Using Machine Learning," 2025 International Conference on Data Science and Its Applications (ICoDSA), Bandung, Indonesia, 2025, pp. 1274–1278, doi: 10.1109/ICoDSA67155.2025.11157087

[6] A. M. Abdullahi, M. S. Hossain, S. A. Htet, S. Ismail, N. N. N. Naing, and M. A. M. Zaaba, "A Prediction of Customer Segmentation using Random Forest in RapidMiner," 2023 IEEE 21st Student Conference on Research and Development (SCOReD), Kuala Lumpur, Malaysia, 2023, pp. 33–38, doi: 10.1109/SCOReD60679.2023.10563451

[7] N. Kodikara and G. Shahtahmassebi, "Predicting Potential Customers in Direct Marketing Using Uplift Modelling and Advanced Machine Learning," 2023 International Conference on Computer and Applications (ICCA), Nottingham, UK, 2023, pp. 1–7, doi: 10.1109/ICCA59364.2023.10401510

[8] J. Asare-Frempong and M. Jayabalan, "Predicting Customer Response to Bank Direct Telemarketing Campaign," 2017 International Conference on Engineering Technology and Technopreneurship (ICE2T), Kuala Lumpur, Malaysia, 2017, pp. 1–6, doi: 10.1109/ICE2T.2017.8216038

[9] Y. Kulkarni, A. Mahamuni, S. Sane, P. Kalshetti, K. Patil, and R. S. Jarad, "Analysing How Marketing Management and Artificial Intelligence are used to Change Customer Engagement," 2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies (TQCEBT), Pune, India, 2024, pp. 1–5, doi: 10.1109/TQCEBT59414.2024.10545064

[10] M. Marciel, J. Gonzalez, Y. M. Kassa, R. Gonzalez, and M. Ahmed, “The Value of Online Users: Empirical Evaluation of the Price of Personalized Ads,” 2016 11th International Conference on Availability, Reliability and Security (ARES), Toulouse, France, 2016, pp. 694-701, doi: 10.1109/ARES.2016.89

[11] A. A. P. Siswadi, A. Suhendra and A. Darmayantie, “Images processing of facial expression to predict the customer opinion towards a product,” 2015 International Conference on Information & Communication Technology and Systems (ICTS), 2015, pp. 109-112, doi: 10.1109/ICTS.2015.7379881

[12] G. Rompolas, “Exploiting time-series analysis to predict customers’ behavioural dynamics in social networks,” 2022 13th International Conference on Information, Intelligence, Systems & Applications (IISA), Corfu, Greece, 2022, pp. 1–6, doi: 10.1109/IISA56318.2022.9904411

[13] P. Samanta, M. Amir, S. Bagga, B. Dubey, M. Khulbe, and R. Gupta, “Influencer Marketing in the Age of AI: Improving Brand Loyalty Through Advanced Analytics,” 2025 International Conference on Intelligent Control, Computing and Communications (IC3), Mathura, India, Feb. 13–14, 2025, pp. 1–6, doi: 10.1109/IC363308.2025.10956901

[14] S. Aburass, O. Dorgham, J. Al Shaqsi, M. Abu Rumman, and A. Al Zarouni, “Optimizing Customer Response Prediction in Auto Insurance: A Comparative Study of Machine Learning Models,” 2024 2nd International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates, Feb. 26–28, 2024, pp. 1–6, doi: 10.1109/ICCR61006.2024.10532980.

[15] F. Ishtiaque, F. R. Mashrur, M. T. I. Miya, K. M. Rahman, R. Vaidyanathan, and S. F. Anwar, “AI-based Consumers’ Preference Prediction Using a Research-grade BCI and a Commercial-grade BCI for Neuromarketing: A Systematic Comparison,” 2023 International Conference on Electrical, Computer and Communication Engineering (ECCE), Chittagong, Bangladesh, Feb. 23–25, 2023, pp. 1–6, doi: 10.1109/ECCE57851.2023.1010