Email Campaign Effectiveness Analysis for Sephora

Hunter Berberich

Introduction

Customized email campaigns are a crucial tool for businesses like Sephora to engage customers and drive sales. By analyzing user interactions with email advertisements, we can assess the campaign's effectiveness and optimize future strategies. This report focuses on evaluating the current email campaign's performance and providing actionable recommendations to enhance customer engagement and sales. Various machine learning models were employed to predict user clicks on email advertisements, using historical click data, product features, and user characteristics. The study aims to provide insights into the campaign's success and identify areas for improvement.

Method

Data Collection
The dataset comprises three components: click history, product features, and user characteristics. The click history dataset includes user_id, product_id, and a binary clicked variable indicating whether the product was clicked. The product features dataset contains product_id, category, on_sale status, number_of_reviews, and avg_review_score. The user characteristics dataset includes user_id, number_of_clicks_before, ordered_before status, and personal_interests.

Data Preprocessing
Data preprocessing involved merging the datasets on user_id and product_id, handling missing values, and removing outliers. Negative values in avg_review_score were filtered out. The number_of_clicks_before variable was converted to numeric after removing '+' signs. One-hot encoding was applied to the category variable. Features were normalized using MinMaxScaler.

Model Building
Several machine learning models were employed to predict the clicked variable. A Decision Tree was configured with a maximum depth of 10 and a minimum sample split of 5. Logistic Regression was used with a maximum iteration of 100. A Random Forest model was configured with 100 estimators and a random state of 0. Naive Bayes utilized Gaussian Naive Bayes. Finally, AdaBoost was configured with 50 estimators and a learning rate of 0.5.

Model Evaluation
The models were evaluated using accuracy scores on both training and testing datasets. Additionally, confusion matrices and classification reports were generated to provide insights into the models' precision, recall, and F1-score.

Results

The Decision Tree model achieved an accuracy score of 0.75 on the training set and 0.78 on the testing set. The Logistic Regression model achieved an accuracy score of 0.62 on both the training and testing sets. The Random Forest model demonstrated strong performance with an accuracy score of 0.72 on the training set and 0.85 on the testing set. The Naive Bayes model achieved an accuracy score of 0.64 on the training set and 0.65 on the testing set. The AdaBoost

model showed promising results with an accuracy score of 0.77 on the training set and 0.76 on the testing set.

Table 1: Model Performance Comparison

Model	Training Accuracy	Testing Accuracy
Decision Tree	0.75	0.78
Logistic Regression	0.62	0.62
Random Forest	0.72	0.85
Naive Bayes	0.64	0.65
AdaBoost	0.77	0.76

Discussion

The Random Forest and AdaBoost models outperformed the other models in predicting user clicks, with the Random Forest model achieving the highest accuracy on the testing set. Logistic Regression, while straightforward and interpretable, showed the lowest accuracy, indicating it may not capture the complex relationships in the data effectively. The performance of the models highlights the importance of model selection and tuning in predictive analytics.

Conclusion

This analysis demonstrates the application of various machine learning models in predicting user clicks on email advertisements. The Random Forest and AdaBoost models were identified as the most effective, providing significant insights for enhancing user engagement and sales. Future work could explore additional features and more advanced models to improve prediction accuracy further.

Recommendations

Based on the analysis of the email campaign's effectiveness, it is recommended to optimize email content by personalizing messages based on user interests and behaviors. Segment users for targeted campaigns and conduct A/B testing on elements like subject lines and sending times to identify effective strategies. Continuously monitor campaign performance and establish a feedback loop for ongoing refinement. Enhance user experience on landing pages to increase conversions and focus on high-impact segments for better ROI. Implementing these strategies will improve customer engagement, increase website visits, and boost sales for Sephora.