

Pentathlon

Next Product to Buy Modeling

Group 14 - Aditya Kumawat

Bindu Priyanka Achalla

Keerthana Raviprasad

Yves Kalin

Introduction to the Problem:

- Background: Pentathlon, a leading European sporting goods retailer, faced challenges with its digital marketing strategy, particularly in how promotional emails were managed across various product departments. The decentralized nature of email decisions led to concerns over customer email fatigue.
- Anna Quintero, the Director of Digital Marketing at Pentathlon initiated a comprehensive approach
- Promotional Email Frequency Test: Conducted to evaluate the impact of email frequency on customer engagement, sales, and email unsubscribes. This test was crucial for understanding the optimal number of promotional emails that should be sent to customers to maximize engagement without overwhelming them.
- Survey Analysis: A survey was sent to customers to gauge their perceptions of the frequency of promotional emails. The findings indicated a significant portion of customers felt they received too many promotional emails from Pentathlon.
- Challenge:
 - A large percentage of customers indicated they were receiving too many promotional emails, leading to potential disengagement.
 - The need for better coordination between departments to manage the overall number of emails sent to each customer, ensuring a more strategic and unified approach.
 - Determining the right balance between engaging customers and avoiding email overload, which could lead to unsubscribes and decreased customer lifetime value.

Problem Statement & Approach:

Optimizing Customer Targeting:

- Data-Driven Email Strategy: Targeted promotional emails based on customer data analysis, ensuring messages were sent to those most likely to engage.
- Effective Customer Segmentation: Utilized purchase history and demographics for precise messaging, increasing campaign influence.

Maximizing Incremental Profits:

- Optimal Email Frequency: Identified two emails per week as ideal, balancing engagement with minimizing unsubscribes, thus efficiently using marketing resources.
- Profit-Focused Messaging: Customized communications based on expected profits from each customer segment, maximizing marketing effectiveness.

Refining Marketing Strategies:

- Advanced Analytics for Insight: Applied analytics (e.g., logistic regression, neural networks) to refine email marketing tactics and understand customer behaviors.
- Strategic Email Content Balance: Adjusted email frequency and content relevance based on customer feedback, optimizing engagement and campaign awareness.

Training Models:

- Logistic regression: Trained and tuned by adding interactions from a three noded neural network.
- XGBoost: Trained and tuned model by adding two best hyper parameters like max_depth and min_child_weight
- Random Forest: Trained and tuned model by adding hyper parameters like max_features and n_estimators
- Neural Network: Trained and tuned model by adding hyper parameters like hidden layers and alpha(learning rate)

Approach to calculate Probability and Expected Profit

Probability: We replaced 'message' column with each type of message(including no message)and predicted the probability of buying for each customer if the corresponding message was sent

Profit: Used Linear Regression model prediction to predict the order size for type of message. Determined the order size for each user by multiplying the probability of purchasing, associated with each message type, by the predicted total order size specific to that message for each user. Subsequently, this value was multiplied by 0.4

Expected Profit Results for each model

Logistic Regression

e_endurance_lrc	0.626940
e_backcountry_lrc	0.594256
e_water_lrc	0.603166
e_racquet_lrc	0.522794
e_team_lrc	0.539639
e_control_lrc	0.430488
e_strength_lrc	0.602624

Random Forest

e_endurance_rf	0.665028
e_backcountry_rf	0.671833
e_water_rf	0.674029
e_racquet_rf	0.600145
e_team_rf	0.608897
e_control_rf	0.509986
e_strength_rf	0.650768

Neural Network

e_endurance_nn	0.617559
e_backcountry_nn	0.578705
e_water_nn	0.616969
e_racquet_nn	0.532587
e_team_nn	0.535989
e_control_nn	0.431539
e_strength_nn	0.596456

XGBoost

e_endurance_xgb	0.621005
e_backcountry_xgb	0.585934
e_water_xgb	0.607161
e_racquet_xgb	0.529251
e_team_xgb	0.537088
e_control_xgb	0.437089
e_strength_xgb	0.589700

Average expected Profit:

average_highest_profit_rf = 337014.70449521573
average_highest_profit_xgb = 310502.69973430404
average_highest_profit_nn = 308779.3290650305

The analysis proceeds to calculate the total expected profit under three scenarios for a campaign targeting 5,000,000 customers:

- a. Random Message Assignment:** Multiplying the average profit of a randomly chosen message by 5,000,000.
- b. Highest Average Message:** Selecting the message with the highest average expected profit and multiplying its average profit by 5,000,000.
- c. No Message (Control Condition):** Multiplying the average profit for the control condition by 5,000,000.

Conclusion:

In summary, the analysis suggests that a strategic approach to selecting promotional messages, even without full customization, can significantly improve profitability compared to random messaging or not sending any promotional messages. However, the implication is that further profit gains could be realized by customizing messages to individual customer preferences, highlighting the potential value of leveraging customer data analytics for marketing strategies.