**Customer Campaign Response Prediction Report**

Prepared by: Esparance Tuyishime  
Date: June 12, 2025

# 1. Executive Summary

This report details a predictive modeling project aimed at enhancing customer targeting for marketing campaigns. Using historical campaign data, we developed a machine learning classifier to predict which customers are most likely to respond positively. The insights derived from the model support smarter outreach strategies and optimize resource allocation for future campaigns.

# 2. Objective

The goal of this project is to predict customer responses (yes or no) to marketing campaigns using historical data. This predictive capability enables targeted marketing and increases campaign efficiency.

# 3. Dataset Overview

* The dataset, `hist.csv`, contains approximately 260,000 rows and includes the following features:  
  week\_id  
  customer\_id  
  attribute1  
  state\_id  
  Sex  
  campaign\_id  
  response (target variable)
* No missing values were found in the dataset.

# 4. Data Preprocessing

* Categorical features (`attribute1` and `Sex`) were label-encoded.
* All features were numerically represented for modeling.
* Additional features such as `campaign\_frequency`, `campaign\_mod\_3`, `campaign\_group`, and `prior\_responses` were engineered to improve predictive power.

# 5. Exploratory Data Analysis

# The response variable is imbalanced: more customers did not respond than those who did.

# A heatmap revealed weak correlations between features and the target, with `Sex` showing a slight negative correlation.

# The countplot and correlation matrix are presented below.

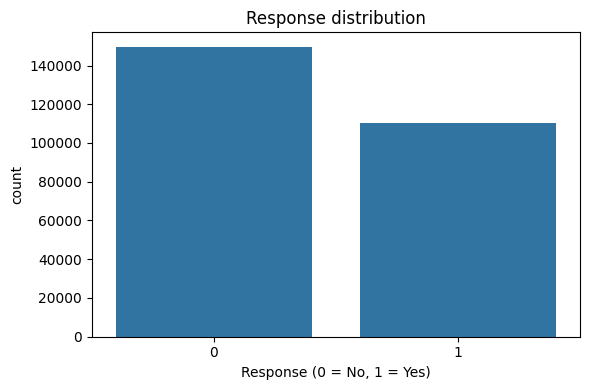


Figure 1: Distribution of Response Variable

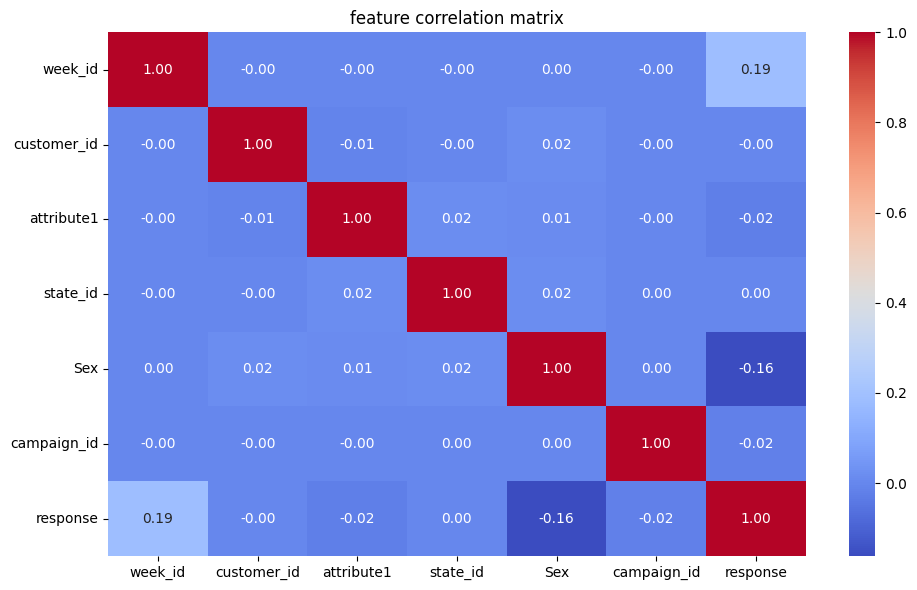


Figure 2: Correlation Matrix of Input Features

# 6. Model Building

* A Random Forest Classifier was selected for its robustness and performance.
* The dataset was split into training and testing sets using a standard 80/20 split.
* Model performance was evaluated using accuracy and classification metrics (precision, recall, F1-score).

# 7. Feature Importance

* The Random Forest model revealed the most influential features:  
  - campaign\_frequency  
  - state\_id  
  - prior\_responses  
  - attribute1  
  - campaigns\_seen

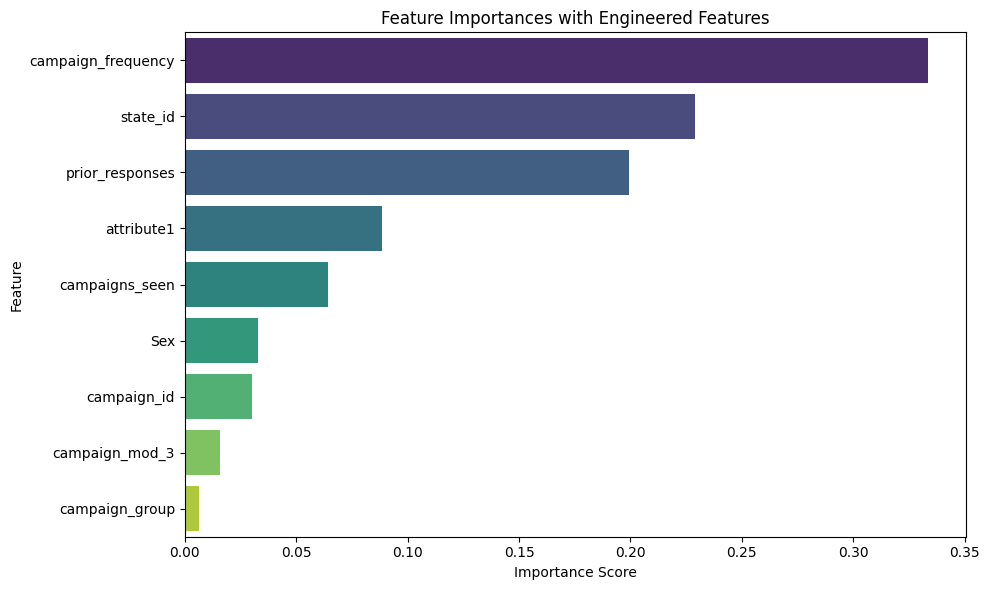


Figure 3: Feature Importances with Engineered Features

# 8. Top Customer Recommendations

Based on predicted probabilities, the following customer IDs are in the top 25% most likely to respond positively to a campaign. These customers should be prioritized for email outreach:

5819, 3649, 7769, 5869, 5496, 9126, 12596, 21, 8632, 2749

# 9. Forecast for Week 27

For Week 27, the expected response rate among the top 25% predicted responders is:

Expected Response Rate: 74.19%

# 10. Conclusion and Recommendations

The Random Forest model demonstrated strong predictive capability and identified key features influencing customer behavior.  
The model supports the design of personalized, data-driven campaigns.  
It is recommended to:  
 • Focus outreach on top-scoring customers.  
 • Monitor the actual response rate in Week 27.  
 • Continue refining models with updated campaign data.  
Further experimentation with boosting methods (e.g., XGBoost) may yield even higher performance.