

Email Engagement (Campaign) Analysis Report

~ by Keivalya Pandya

Objective

Perform data analysis, apply machine learning, and communicate findings.

Exploratory Data Analysis (EDA)

The dataset comprises email interaction metrics, including whether an email was opened, whether the meeting link was clicked, and whether there was a response.

Correlation Analysis

A correlation matrix was generated to examine the relationships between the different variables. The matrix revealed:

- A moderate positive correlation between 'opened' and 'meeting link clicked'.
- A weaker positive correlation between 'opened' and 'responded'.
- A very weak negative correlation between 'meeting link clicked' and 'responded'.

Distribution Analysis:

Histograms and pairplots were used to visualize the distributions of individual variables and their pairwise relationships. The analysis showed:

- Bimodal distribution for the 'opened' variable.
- A concentration of lower values for 'meeting link clicked' and 'responded' variables, indicating fewer link clicks and responses.

Cumulative Metrics Over Time

A cumulative open rate plot over time suggested that the open rate stabilizes just below 80%, indicating a high likelihood of emails being opened over time.

Model Development and Evaluation

Several machine learning models were developed to predict the likelihood of email engagement based on the features available in the dataset.

Class Imbalance

It was noted that the dataset is imbalanced, with a greater number of opened emails compared to unopened ones.

Performance Metrics

Due to the class imbalance, precision, and recall were identified as critical metrics for evaluating model performance. The models consistently showed a high precision rate for predicting opened emails but failed to identify unopened emails.

Accuracy

Despite the model's high overall accuracy, it was biased toward predicting the majority class (opened emails).

Future Work

Plans to incorporate Large Language Model fine-tuning to predict whether the email is 'engaging' enough to be opened or not and whether the e-mail body is capable of attracting the readers to click on the meeting link. In both cases, a generative model trained on this data can help in crafting better content for the audience.