

## SUMMARY

### Email Campaign Effectiveness Analysis:

Email marketing has become an essential tool for small and medium-sized businesses to keep their customers informed about their products, services, and updates. However, sometimes emails are ignored due to various reasons such as improper structure, excessive images or links, complex vocabulary, or length. In this context, the objective of this project was to develop machine learning models that can predict if an email will be ignored, read, or acknowledged and identify the important features that prevent an email from being ignored.

The exploratory data analysis (EDA) revealed that different email campaign types have varying levels of email engagement, with Campaign Type 1 having a high likelihood of being read despite fewer emails sent, while Campaign Type 2 was mostly ignored despite more emails being sent. Moreover, an average word count of 600, a subject hotness score of 0.6, and an average of 10 links and 3 images in email content lead to a better engagement rate. Distribution plots showed that Word Count follows a normal distribution, while Total Links and Total Images are positively skewed with the presence of outliers. The correlation matrix supports earlier observations, suggesting positive correlations between Email Campaign Type and Total Past Communications with email engagement, and negative correlations between Word Count and Subject Hotness Score with email engagement.

The machine learning models employed in the project were Logistic Regression, Decision Tree, Random Forest, and XG Boost. The evaluation results showed that the XG boost on the SMOTE data performed the best with an F1 score of 0.80 on the test set. This model was able to correctly classify the majority class (class 0) with high precision and recall while also improving the performance on the minority classes (class 1 and class 2) compared to the other models.

Based on the EDA and model evaluations, some recommendations were made. Email Marketing Campaign Types 1 and 3 performed better than type 2, so concentrating on improving type 2 may suffice. The word count should be appropriate, with content that is concise and to the point, with a few marketing gimmicks thrown in for good measure. Keeping the number of images and links to a minimum is also recommended. It was observed that total previous communications had a positive influence, so having a healthy relationship with customers is crucial.

Overall, the project highlighted the importance of email marketing in building relationships with leads and customers. It also demonstrated the potential of machine learning models in predicting email engagement and identifying the factors that prevent an email from being ignored. The recommendations provided can be useful for small to medium business owners who rely on email marketing strategies to stay in business.

However, it is important to note that the performance of the models can still be improved. The low F1 scores for the minority classes indicate that the models may not be effective in predicting engagement for all customers. It may be worthwhile to investigate other machine learning techniques, such as deep learning, to improve the model's performance. Additionally, collecting more data and including more features related to customer behavior and preferences may also improve the model's performance.

at last, this project successfully developed machine learning models to predict email engagement and identified the important features that prevent an email from being ignored. The recommendations based on the EDA and model evaluations can help small and medium-sized businesses to improve their email marketing strategies and increase customer engagement.