Report on Predicting Total Engagements Using Random Forest Regressor

Introduction

In this report, I present an analysis of a dataset related to marketing campaigns, specifically focusing on predicting the total engagements using a Random Forest Regressor model. The dataset was explored and preprocessed to prepare it for modeling. The predictive model was trained, evaluated, and further tuned to enhance its performance.

1. Exploratory Data Analysis (EDA) and Preprocessing

1.1 Data Loading and Basic Evaluation

The dataset, consisting of information related to marketing campaigns, was loaded into a Pandas DataFrame. Initial exploration included -

- displaying the first few rows,
- examining the data's shape, and checking for missing values.
- Duplicate rows were identified and removed, and 'NA' values were appropriately handled.

1.2 Label Encoding and Standardization

Categorical variables were encoded into numerical format using Label Encoding. Numerical columns were standardized using the StandardScaler to ensure uniformity and ease of interpretation for the machine learning model.

1.3 Outlier Handling and Visualization

Outliers in numeric columns were addressed using the IQR method. The impact of these procedures on the data distribution was visualized through histograms, bar plots, scatter plots, and pair plots. These visualizations helped in understanding the distribution of performance metrics and creative attributes.

1.4 Correlation Analysis

A correlation heatmap was generated to explore the relationships between different performance metrics and creative attributes. This analysis aided in identifying potential predictors for the target variable, "total engagements."

2. Building and Evaluating the Random Forest Regressor Model

2.1 Data Preparation for Modeling

A subset of relevant columns, including both performance metrics and creative attributes, was selected for model building. Categorical variables were encoded, and the data was split into training and testing sets.

2.2 Model Initialization and Training

A Random Forest Regressor model was initialized and trained using the training set. The model aimed to predict the "total engagements" based on the selected features.

2.3 Model Evaluation

The performance of the initial model was evaluated using mean squared error (MSE) and R-squared metrics on the test set. The initial evaluation provided insights into the model's ability to predict the target variable.

Mean Squared Error: 0.19334409672676084
R-squared: 0.07822152626711032

3. Hyperparameter Tuning

3.1 Grid Search for Hyperparameters

Hyperparameter tuning was performed using GridSearchCV, searching for the optimal combination of hyperparameters for the Random Forest Regressor. The parameters considered included the number of estimators, maximum depth, minimum samples split, and minimum samples leaf.

Best Hyperparameters: {'max_depth': 10, 'min_samples_leaf': 1, 'min_samples_split': 10, 'n_estimators': 200}
Tuned Model - Mean Squared Error: 0.17608361774445758
Tuned Model - R-squared: 0.16051179652393144

4. Conclusion

By focusing on creative attributes with a proven impact on campaign performance and incorporating insights from correlation analysis, digital marketers can enhance the effectiveness of their campaigns. Continuous testing, data-driven decision-making, and the use of predictive models contribute to a strategic and adaptive approach, ultimately leading to more successful digital marketing initiatives

1. Focus on Creative Attributes with Significant Impact

1.1 Dominant Color:

The dominant color of visual content appears to have an impact on campaign performance. Consider leveraging colors that resonate with the target audience and align with the campaign's messaging. A/B testing different color schemes can help identify the most effective combinations.

We can see from the below plot that Dominant colour has a huge impact on Spend.
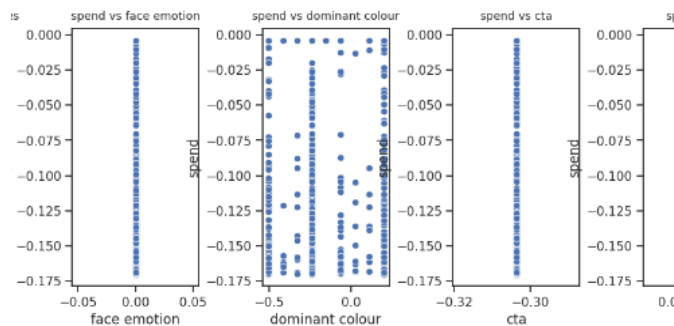
1.2 Call-to-Action (CTA):

The CTA used in campaigns plays a crucial role. Analyze the performance metrics associated with different CTAs and prioritize those that lead to higher engagement. Experiment with variations in language and placement to optimize the effectiveness of the CTA.

1.3 Style and Tone:

The style and tone of the content contribute to campaign success. Explore the relationship between style/tone attributes and engagement metrics. Tailor the content style to match the preferences of the audience, whether it's formal, casual, humorous, or informative.