This report explains how I analyzed a dataset about marketing campaigns to predict total engagements using a Random Forest Regressor model. I explored, cleaned, and processed the data to prepare it for Modeling , then trained and evaluated the model, and improved its performance through tuning.

1. Exploratory Data Analysis (EDA) and Pre-processing

1.1 Data Loading and Basic Evaluation

I loaded the dataset into a Pandas Data Frame and looked at the first few rows to understand its structure. I also checked for any missing values and removed duplicate rows.

1.2 Label Encoding and Standardization

I converted categorical variables (like text data) into numerical format using Label Encoding. Then, I standardized the numerical columns to make sure they were on a similar scale.

1.3 Outlier Handling and Visualization

I used the Interquartile Range (IQR) method to handle outliers in numerical columns. I visualized the data distribution with histograms, bar plots, scatter plots, and pair plots to better understand the performance metrics and creative attributes.

1.4 Correlation Analysis

I created a correlation heatmap to see the relationships between different metrics and attributes. This helped in identifying which features might predict the "total engagements" well.

2. Building and Evaluating the Random Forest Regressor Model

2.1 Data Preparation for Modelling

I selected relevant columns for modelling, encoded categorical variables, and split the data into training and testing sets.

2.2 Model Initialization and Training

I initialized and trained a Random Forest Regressor model on the training set to predict "total engagements."

2.3 Model Evaluation

I evaluated the model using mean squared error (MSE) and R-squared metrics on the test set. The initial model had:

Mean Squared Error: 0.193

R-squared: 0.078

3. Hyperparameter Tuning

3.1 Grid Search for Hyperparameters

I used GridSearchCV to find the best hyperparameters for the Random Forest Regressor. The best parameters were:

max_depth: 10

min_samples_leaf: 1

min_samples_split: 10

n_estimators: 200

The tuned model improved performance with:

Mean Squared Error: 0.176

R-squared: 0.161

4. Conclusion

By focusing on creative attributes that significantly impact campaign performance, digital marketers can improve their campaign effectiveness. Continuous testing, data-driven decisions, and predictive models help in developing a strategic approach, leading to more successful digital marketing efforts.