Data Scientist Test Task [v2.0] inPowered

Optimizing CPE across groups

Rafael Henrique Martello

1. Introduction

As a data scientist working in an advertising agency, one of our core objectives is to optimize the performance of our ad campaigns. A critical component of this optimization process involves determining the most effective bid and budget strategies for individual advertising items. These items, which are grouped into distinct campaigns, each possess unique targeting parameters and creative elements (such as headlines and summaries). Our key performance indicator (KPI) in this context is the cost per engagement (CPE), a metric that we aim to minimize to enhance the efficiency and cost-effectiveness of our ad delivery.

To achieve this, we have been provided with a dataset that contains historical performance data for various advertising items. The dataset includes a variety of features such as the bid and budget amounts, engagement metrics, media spend, and detailed targeting information. Additionally, it encompasses creative attributes like the ad's headline and summary. The challenge is to leverage this historical data to develop a machine learning model capable of predicting the CPE for future advertising items based on their targeting, creative, bid, and budget variables.

2. Objective

The primary objective of this project is to build a predictive model that can accurately forecast the CPE of an ad item. By doing so, we can make data-driven decisions on how much to bid and budget for each ad item, ultimately improving our ad delivery and reducing costs. The model will take into account the following key variables:

3. Approach

To address the problem, we will:

1. Exploratory Data Analysis (EDA): Analyze the dataset to understand the distribution of variables, identify any missing values, and uncover relationships between different features and the CPE.

2. Model Selection: Experiment with various machine learning algorithms to identify the best model for predicting CPE. Potential models include linear regression, decision trees, random forests, and gradient boosting machines.

3. Model Evaluation: Evaluate the performance of the models using appropriate metrics (e.g., mean squared error, R-squared) and select the best-performing model.

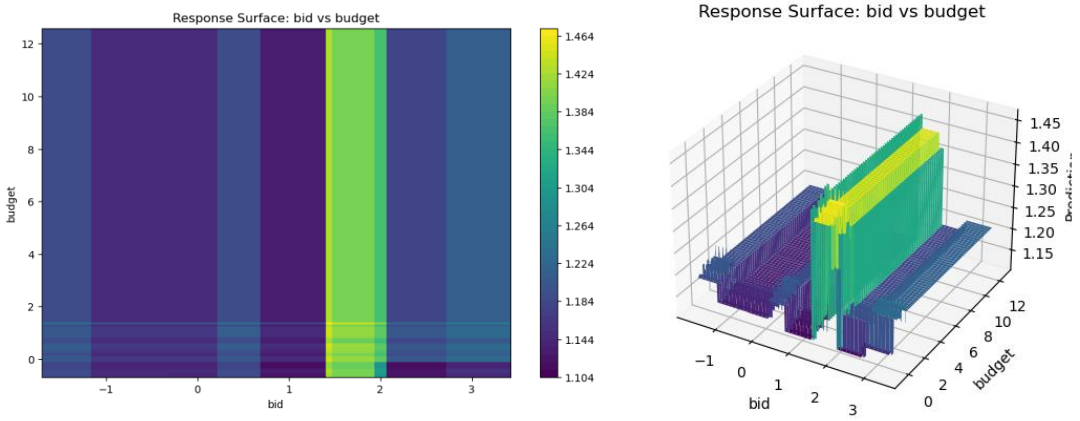
4. Model Tuning and Validation: Fine-tune the selected model and validate its performance on a holdout dataset to ensure it generalizes well to new data.

5. Response Optimization: After selecting and fine-tuning the best-performing model for predicting CPE, the next step is to optimize the advertising strategy based on the model's predictions. This involves identifying the optimal values for the variables that influence CPE to achieve the lowest possible CPE.

Results:

Using a GradientBoosting Regressor

I have generated a search space for the cpe target:



And using a minimizing function I have reached a CPE of **1.11** with a **bid of 0.6** and a **budget of 383.11.**