

# **KPI Prediction Tool: A Look Into The Future**

# 

# **Final Report**

**Team 4**

**Economics 599.88/611**

**Department of Economics**

**University of Calgary**

**April 29th/2022**

# 

# Project Partners

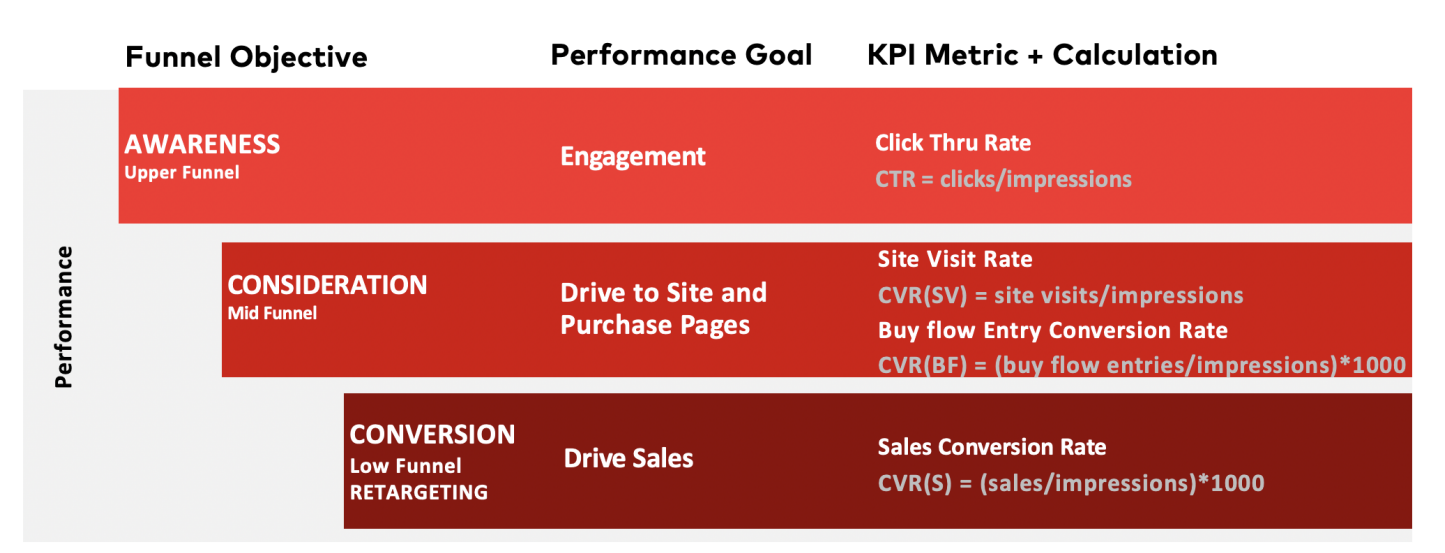
| **NAME** | **ORGANIZATION** |
| --- | --- |
| Ady Gray | Critical Mass |
| Gunpreet Singh | Critical Mass |
| Kimberley McDonough | Critical Mass |
| Dan Lewis | Critical Mass |

# Team Information and Role Assignment

The Critical Mass (CM) team, sometimes also referred to as ‘Team 4’, consisted of five talented and hardworking individuals. Our team, blessed with different academic backgrounds, instantly felt motivated by the CM project and selected it as our first choice. We were intrigued by the challenges this project presented and felt that our combined skill sets were perfect for creating the tool that the partner was seeking. Our team members along, with their roles and responsibilities, are as follows:

| **TEAM MEMBER NAME** | **ROLE** | **RESPONSIBILITIES** |
| --- | --- | --- |
| Feyre Gezahegn | Report Lead | * Responsible for communicating with Critical Mass * Converts information obtained from data into readable language * Ensures the objective, process, and results of the project are clearly communicated through verbal and written methods |
| Wes Warman | Project Lead | * Work together with each Lead to ensure project flows * Ensure deadlines are going to be met and targets are reached |
| Muhammad Usman Khan | Engineering Lead | * Responsible for the creation and design of a GUI * Assist in the creation of the Model and the machine learning algorithm |
| Saul Chirinos | Data Lead | * Analyze, identify, and interpret patterns and trends in the data * Assist Modeling Lead with developing and tuning the forecasting algorithm * Assign data tasks to other team members such as data cleaning and exploratory data analysis. |
| Yutong Liu | Modeling Lead | * Doing the model construction, including the discussion of the model selection, feature engineering method and validation method. * Leading on aspects of comparing the different model performance to generate a more robust outcome. |

# Project Objective

Forecasting makes it possible to determine what the unknown future may be. That is to say, creating a forecasting model for a Key Performance Indicator (KPI) or performance indicator is beneficial. It allows us to see how much effect we have. Generally, forecasting a KPI allows us to identify opportunities to enhance the performance of the KPIs. In this project, we used the data set given to us by CM to predict time series KPI in terms of several collected features. We also worked with CM to build a predictive tool that will forecast KPI for new creatives as they launch. Each performance indicator is linked to a specific funnel which is divided into three parts. First, in the upper funnel, consumers are ‘aware’, that is to say, they engage with the company. Second, in the mid-funnel, consumers ‘consider’ the company by exploring products they find on the company website. Lastly, in the low funnel, the ‘conversion’ phase, consumers purchase products from the company.

TThe project was divided into four main parts. (1) Start by data wrangling and transforming the given dataset to make it usable for our models. (2) Perform exploratory data analysis (EDA) to further understand the data and find answers to the primary and secondary learning outcomes. (3) Analyze the data and determine the importance of each predictor to the target variables, then build the funnel-specific predictive model for evaluation. (4) Due to the nature of the project, it was unlikely that a better method of visualization could be created in four months. As a result, the forecast tool has no built-in visualization methods and instead outputs a CSV file that can be imported into Tableau or any other visualization tool. This is an ideal solution as it ensures CM is capable of using an already established tool that they are familiar with.

# How Project Creates Value for CM

The importance of forecasting KPI for a given creative will allow CM to see how and where resources should be allocated to maximize the efficiency and performance of a given creative. The project first outputted the average expected (funnel specific) KPI to develop a performance trend line for the creative life cycle. To do this, Team 4 required historical data to construct a quantitative model to anticipate our KPI. Patterns from historical data are quantified and turned into a forecasting algorithm. The more historical data on our KPI and measurements of its affecting elements we can collect, the more accurate our forecasting model will be. Even if we can employ AI, we will feed it this data so it can construct a model for us. The project also built CM a user-friendly graphical user interface (GUI) tool for users to easily handle the forecast task and acquire data visualization for KPI within several clicks.

In addition to the first two, Team 4 identified the most important features for KPI prediction, optimized the number of creatives per funnel and publisher, and identified the impact of the number of creatives in the market on performance. Once the project has been completed, CM will have further ability to make effective and impactful business decisions through data science.

# How CM Defines Success

Concerning the KPI prediction, CM considers the model a success in three respects: accuracy, improvement, and learning. (1) Since KPI forecasting is more accurate than price forecasting, the project will be considered a success if it has a 55 percent or more accuracy to the true value. Hence, the project yields a higher value for CM as the relationship between the predictor and target variables improves. (2) If our KPI prediction tool can improve CM’s existing simple model by 10 percent, the model will be considered a success. (3) The KPI prediction can be considered a success by contributing to the existing or new primary and secondary learning outcomes, which further feeds into CM’s motto: “measure, optimize, and adjust”. However, Team 4 was constrained by time and hence was only able to determine the primary learning outcomes with some secondary learning outcomes.

# Prototype Description

## 6.1 Data

The data set contains over 13,000 creatives/ads, each with selected features and KPI results. The features included: channel, funnel, publisher, line of business (LOB), product, theme, creative version, KPI audience, price, price placement, discount, offer placement, offer group, length, asset type, video type, and ad size. As for the creative KPIs, weekly data was provided for when the creative was in the market.

## 6.2 Algorithm Design

Our final algorithmic design for predicting funnel-specific average KPI includes a random forest model to predict CTR and CVR(SV), and a light gradient boosting machine model to predict CVR(BF) and CVR(S). The change to the LGBM model was found helpful since we were averaging a lower error score and higher R-squared score on the holdout set.

**Pseudocode:**

1. Select funnel specific data based on the KPI to predict
2. Split data into training and testing sets
3. Create a column transformer object for preprocessing
   1. Encode categorical features
4. Create model (RF/LGBM) passing in parameters we found to perform well
5. Create a pipeline object that transforms the desired columns and fits the preprocessed data into the model
   1. Cross-validate the pipeline using the training data from step 2 (the cross validation will already split the training data into another set of training and validating sets)
   2. Compute R-squared and mean absolute error
   3. Compare average model score with previous scores
6. Fit the pipeline to the training data from step 2 and predict on the hold-out testing data also from step 2
   1. Compute R-squared and mean absolute error
   2. If we observe lots of overfitting, repeat the process with a different approach

## 6.3 Estimation Equations

Since our models are tree-based models, there is no explicit formula for showcasing the model. Random forest and LGBM work by creating N decision trees and aggregating their prediction for an ensemble prediction. One advantage for using tree based models, however, is that we can quantify the impact that each feature has on average KPI.

**Feature Selection**

When doing prediction, we may need to execute further feature selection. Generally, we want to reduce the number of input variables, which in turn, lowers the computational cost of modeling and in certain situations, may improve the performance of the model. For this project, two tree-based model methods, random forest and Light GBM, were used to determine the feature importance.

The tree-based model uses a collection of decision trees, which consists of Internal nodes and leaves. The selected characteristic utilizes the internal node to decide how to partition the data set into two different sets with comparable replies. Then, the characteristics of the internal node are chosen using a criterion such as Gini impurity or information gain for classification tasks, and variance reduction for regression. At this point, we can see how each characteristic reduces the split's impurity. Moving on, we calculate how each feature reduces impurity on average for each feature. Finally, the feature significance is calculated as the average of all trees in the forest.

The final results of the selected features using two models are as follows:



# Validation

To test our models we generated two performance metrics, R-Squared and Mean Absolute Error. We also used 5-fold cross validation and tested our model on the hold out data, generating R-Squared and Mean Absolute Error scores. This was done for each funnel specific KPI prediction model, and the scores are as follows:

**Performance Metrics for Model with Training/Validation Data**

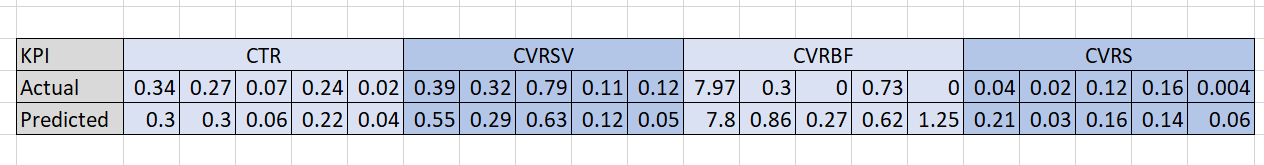
| KPI | CTR | CVRSV | CVRBF | CVRS |
| --- | --- | --- | --- | --- |
| R-Squared | 0.65 | 0.67 | 0.58 | 0.42 |
| MAE | 0.054 | 0.12 | 1.01 | 0.27 |

**Performance Metrics for Model using Hold-out Data (Testing)**

| KPI | CTR | CVRSV | CVRBF | CVRS |
| --- | --- | --- | --- | --- |
| R-Squared | 0.59 | 0.65 | 0.53 | 0.37 |
| MAE | 0.051 | 0.12 | 1.15 | 0.30 |

Avoiding overfitting the model was the main objective in tuning them. If our performance metrics dropped substantially after evaluating the model on the withheld data then we should have been worried. However, our validation metrics only drop by about 10 percent on average across all models. For this reason, we do not think that our models are overfitted.

We also used the first five rows of the validation sets to visually show how our model works, and the predictions are as follows:



From these predictions, we can see that predicting the KPIs looks to be consistently accurate for CTR and CVRSV and less accurate for CVRBF and CVRS. We believe this is due to the higher variance in these mean KPIs.