CTR Prediction for an Email Campaign

Submitted by

- Name:Anushree Tomar
- Email:anushree.tomar@gmail.com

Objective

Our Goal is to build a machine learning-based approach to predict the CTR of an email campaign.

Exploratory Data Analysis (EDA)

1. The Dataset is clean dataset as there is no missing values are present.
2. Checked for unique values in the dataset and found that "is_timer" column has only '0' value.
3. Most of the features are Categorical in nature.
4. Visualized pairwise bivariate distributions in a dataset, and found that the variables are not linearly separable and right skewed
5. Checked for the presence of multi-collinearity for independent variables using correlation Matrix this would help in feature selection also.
6. Vary few variables are correlated to each other.

Feature Engineering

1. Dropped the "campaign_id" and "is_timer" variables from the test and train dataset.
2. Created dummy variable for "times_of_day" column in test and train dataset.
3. For Linear Regression performed feature scaling standardization on the train and test dataset and for Deep Learning Model added Normalization layer.

Feature Selection

1. In One of the approach we have created Model using Rndom Forest on selected features.
2. For feature selection we have used Backward Feature Elimination using SequencialFeatureSelector from the library mlxtend and LinearRegression from the sklearn library.

Machine Learning Models

1. Machine Learning Algorithms used for this dataset:

- Random Forest Regressor using RandomizedSearchCV
- Light GBM using traditional Gradient Boosting Decision Tree (boosting_type : 'gbdt')
- Light GBM using Gradient-based One-Side Sampling (boosting_type : 'goss')

- Linear Regression, Ridge and Lasso Regression
- Deep Learning Model using tensorflow

1. Light GBM using Gradient-based One-Side Sampling (boosting_type : 'goss') gave the best results for the training dataset.

Future Scope

1. In order to increase the R-Square value we can also perform feature transformation techniques on the target variable as log transformation etc.
2. We can try other Regression models also as Decision Tree regressor,XGB regressor,SVM and Gradient Boosting.
3. We can also took the average of all predictions in order to achieve good R-Square value.