JOBATHON

Our task is to build a machine learning-based approach to predict the CTR of an email campaign. It was a regression problem. The target value is a continuous value. Several models are tried and based on r2_score value Lasso regressor is used to predict the test values. In exploratory data analysis the shape of data, basic information and statistical description are found.

In pre-processing missing values are checked. Both train and test data set didn't contain any null values. Then outliers are detected using quartiles. Boxplot of features are plotted to graphically represent the outliers. Outlier values which are greater than upper limit was handled by replacing with upper limit and values which are less than lower limit was handled by replacing with lower limit. One of the columns was categorical.

Label encoding is used to transform the categorical values. Two columns are dropped in the Feature engineering section based on correlation. Correlation was represented using heatmap.

Three scaling was used: Min-Max, Standard and Robust scaling. Six models were used namely Decision Tree Regressor, Linear Regression, Ridge Regressor, Lasso Regressor, Random Forest Regressor, K Nearest Neighbors Regressor. A total of 15 models were created by using 5 regressor on 3 types of scaled data. Based on the r2_score values Lasso Regressor model is chosen.