

Reza Sayyari

Student ID:0079105

Report for the Third Homework of Advance Data Analysis with Python

In this homework we are going to create a model for predicting whether a respondent is likely to vote in the last presidential election or not to review the fundamentals of machine learning concept.

A variety of models and methods, such as Random Forest, Linear Discriminant Analysis and Decision Tree, have been tested in the code.

“sklearn.model_selection.SelectKBest” is used in this code for reducing overfitting, improving accuracy , and reducing training time. Also, “sklearn.preprocessing.QuantileTransformer” is used for removing unwanted variables from the data set, such as “MISSING”, “REFUSED.”

After the mentioned processes, the model trains again and the new results are obtained. Finally, five best classifiers and regressors are looped for finding the best hyperparameters.