

CENG 4512 Data Science and Analytics Final Homework
2020/2021 Fall

Classification Pipeline

Tasks:

1. Create a **pipeline** in R to classify using a certain dataset. Your pipeline should be a wrapper function with certain sections
 - a. Section 1: Data import
 - b. Section 2: Data exploration
 - c. Section 3: Data Preprocessing
 - d. Section 4: Training
 - e. Section 5: Testing and comparisons
 - f. Section 6: Prediction

Important! Please make sure your pipeline works with any dataset with binary response.

Important! Please provide source code in .R format

Important! Please stick with **caret** library for the above tasks.

2. **Demonstrate** your pipeline with Online Shoppers Intention dataset

Download Online Shoppers Intention dataset from

<https://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset>

This dataset contains feature vectors belonging to 12,330 online sessions. The dataset was formed so that each session would belong to a different user in a 1-year period to avoid any tendency to a specific campaign, special day, user profile, or period. The dataset consists of 10 numerical and 8 categorical attributes.

The **'Revenue'** attribute should be used as the final class label.

Further details of the dataset can be found in <https://link.springer.com/article/10.1007/s00521-018-3523-0>

Details about submission:

1. You are supposed to submit a source code file in .R format and a report via **RMarkdown**.
2. Please post your.html files and source codes. The name of the files should be your name_surname_final (like eralp_dogu_hwmidterm.html) through DYS. Make sure your html file includes both code and results.