ENGR 421 DASC 521
Homework 06: Modeling Cash Withdrawals from ATMs
Deadline: December 16, 2019, 11:59 PM

In this homework, you will develop a machine learning solution in R, Matlab, or Python for a real-life regression problem from finance industry. Your machine learning algorithm needs to predict the number of cash withdrawals from 47 different ATMs of a bank using the information given about each ATM and the withdrawal date. Here are the steps you need to follow:

1. You are given two input data files, namely, training_data.csv and test_data.csv. The training set contains 42,958 labeled data instances (47 ATMs x 457 days x 2 transaction types), where each training data instance has 7 columns. IDENTITY column gives you the unique identifier assigned to each ATM. REGION column shows the geographical region of each ATM. DAY, MONTH, and YEAR columns give the transaction date. TRX_TYPE column shows the transaction type (1: card present, 2: card not present). TRX_COUNT is the number of cash withdrawals performed on the specified date. You are also given a very simple solution strategy using a decision tree classifier in the file named quick_and_dirty_solution.R.

2. Develop your own machine learning solution for this problem. You are free to use any publicly available packages in R, Matlab, or Python. The predictive quality of your solution will be evaluated in terms of its MAE (mean absolute error) and RMSE (root mean squared error) values on the test set.

3. Use the trained algorithm from the previous step to perform predictions for the test data set, which contains 940 data instances (47 ATMs x 10 days x 2 transaction types). You are not given the numbers of cash withdrawals for test instances. You need to predict the numbers of cash withdrawals and to write these estimates into a file. For example, the decision tree strategy implemented in quick_and_dirty_solution.R file generates the estimates for the test set and writes these values into a file named test_predictions.csv.

What to submit: You need to submit your source code in a single file (.R file if you are using R, .m file if you are using Matlab, or .py file if you are using Python), the estimated numbers of cash withdrawals that you calculated for the test set (test_predictions.csv), and a detailed report explaining your approach (.doc, .docx, or .pdf file). You will put these three files in a single zip file named as ***STUDENTID.zip***, where ***STUDENTID*** should be replaced with your 7-digit student number.

How to submit: Submit the zip file you created to Blackboard. Please follow the exact style mentioned and do not send a zip file named as ***STUDENTID.zip***. Submissions that do not follow these guidelines will not be graded.

Late submission policy: Late submissions will not be graded.

Cheating policy: Very similar submissions will not be graded.