CSE351 Introduction to Data Science

Spring 2020 - Exercise 3 (Submission deadline: Saturday 11 April)

In this exercise, you will solve a data challenge as a group project using the following approach:

1) Visit the data challenge page.
2) Read carefully the problem to be solved.
3) Download dataset.
4) Figure out suitable approaches for solving the problem.
5) Preprocess data for creating a dataset suitable for experimentation.
6) Use your favorite data analysis tool to solve the problem.
7) Report the approach taken, steps for using your favorite data analysis tool and your results in a PPT format and submit on Blackboard.
8) Create a Github repository, upload your preprocessed data, steps for using your favorite data analysis tool and your results. Submit the Github link on blackboard.

Total points: 15

Rubric:

1) Brief description of dataset and concise description of problem statement – 2 points
2) Description of approaches suitable for solving a problem – 4 points
3) Data preprocessing and creating final dataset – 4 points
4) Using your favorite data analysis tool and documentation of involved steps – 4 points
5) Concise description of results – 3 points
6) Valid submission of blackboard: 1 point
7) Valid submission on Github: 2 points

The focus of this exercise is more on a methodological approach to solve a particular challenge and not to evaluate the accuracy of your solution.

Choose **one of the following** challenges. It is preferred that each group takes a separate challenge.

1) Predict purchase amount of sales transactions. Refer to Black Friday dataset at https://www.analyticsvidhya.com/blog/2018/05/24-ultimate-data-science-projects-to-boost-your-knowledge-and-skills/
2) Identify which tweets are hate tweets and which are not. Refer to Twitter classification dataset at https://www.analyticsvidhya.com/blog/2018/05/24-ultimate-data-science-projects-to-boost-your-knowledge-and-skills/
3) Who survived the shipwreck of the Titanic? https://www.kaggle.com/c/titanic
4) Where is a particular taxi cab is going? https://www.kaggle.com/c/pkdd-15-predict-taxi-service-trajectory-i
5) How long a given taxi trip will take? https://www.kaggle.com/c/pkdd-15-taxi-trip-time-prediction-ii