CSCI 3360 Team Course Project

Kaggle (https://www.kaggle.com) is a site that holds data science competitions (from which you could win monetary prizes). A large number of datasets is also hosted there.

Introduction

The team course project is to be completed by teams of 3 or 4 members. Your team is to find a dataset on Kaggle to work on. The dataset could be associated with a competition (https://www.kaggle.com/competitions) or not (https://www.kaggle.com/datasets). To look for a specific type of datasets, you are encouraged to search by tags (https://www.kaggle.com/tags). Here are some example tags that you may find useful:

- https://www.kaggle.com/tags/beginner
- https://www.kaggle.com/tags/regression
- https://www.kaggle.com/tags/classification
- https://www.kaggle.com/tags/clustering

Often times other people have worked on these datasets and shared their work under the "Notebooks" tab. You are encouraged to study other people's work and learn from them. Of course, your course project cannot be a replica of other people's work.

Formation of Teams

You are free to work with any fellow students of your choice. Each team must consist of 3 or 4 members. Once you have formed a team, please **sign up through eLC** by clicking "Tools"→"Groups". You are encouraged to look for a dataset on Kaggle first and make use of Piazza to look for other students who are also interested in working on the same dataset.

Please try to sign up for a team on eLC by **Tuesday, November 3 @11:59:00 PM**. Otherwise, you will be assigned a team.

Project Structure

The course project consists of 3 components:

1. Project Proposal

The project proposal must be a **typed pdf document named proposal_team#.pdf**, where # is to be replaced by your team number (from eLC, e.g., proposal_team1.pdf). The proposal describes your plans for the course project and must include the following:

- The team number and names of your team members.
- A brief description of the dataset of your choice, including the Kaggle link to the dataset.
- Explain what type of exploratory data analysis that your team is planning to do, such as what summary statistics to collect and the type of figures to plot in order to visualize the data.
- List the models that your team is planning to use. The minimum number of models your team must use is the same as the size of your team (teams of 3: at least 3 models; teams of 4: at least 4 models). **At least 1 model must be a new model that has not been discussed in class**.
- Explain the evaluation metrics your team is planning to use and the outputs to be generated.
- Describe how the workload is to be shared among your team members.

The proposal for the course project is to be submitted to eLC by **Friday, November 6 @11:59:00 PM**. *Only one submission per team is needed.*

2. Presentations

Each team is to give a 15-minute presentation on their course project. The presentations will take place during class times on the following dates:

- TUE 11/23/2020
- MON 11/24/2020
- MON 11/30/2020
- TUE 12/01/2020
- THR 12/03/2020

Sign ups for a time slot will be announced later via eLC. Each team member must be present for the presentation and be ready to answer questions related to the course project. You will be asked to provide an intuitive explanation of the new model and how it is similar to and different with a model that was discussed in class.

3. Project Report

The project report must be a **typed pdf document named report_team#.pdf**, where # is to be replaced by your team number (e.g., report_team1.pdf). The report should provide details of the entire process of data analysis with supporting figures, results, interpretations, what you learned, etc.

The report for the course project and your Python source code are to be submitted to eLC by **Friday, December 4 @11:59:00 PM**. *Only one submission per team is needed.*