Project-1 Details

September 25, 2020

1 Class Project - 1

1.1 Overview

In this project you will be putting be applying the skills you learned in the class so far. More specifically, you will be asked to train a classification or regression model on a dataset of your own choice.

1.2 Due Date

You will be submitting your project and present your work on 10/13/2020 at 4:30 PM.

If you have any questions about homework please check with me as soon as possible.

1.3 Dataset

- You will be using a dataset of your own choice.
- You can use API's,webscraping and available open data repositories to get your data. Some suggestions are:
    - UCI ML Repo
    - Kaggle
    - Wikipedia - Kaggle ML
    - Spotify API
    - Yelp API
- When you choose your dataset please be aware of the following pitfalls:
    - Before spending too much time for cleaning the dataset make sure that it's good for the problem you would like to solve. So first make plan and then invest time and effort.
    - Stay aware from time series data. We didn't studied the techniques for handling time series data.
    - Make sure that you find a dataset that is not studied extensively online. For example: No Titanic, House Price, Iris, MNIST etc. datasets.
    - If you don't have a compelling reason don't work with a small dataset (at least 500 rows and 5 features.)

1.4 Deliverables

- For this project there will be three deliverables:

1. A Github repo that contains your project.

2. A Jupyter Notebook (I will refer it as `technical notebook` or `report`).

3. A presentation (No more than 5 minutes and 10 slides)

1.4.1 Github Guidelines

- Put an attention graping title to the repo.

- Github should include a licence.

- ReadMe should be included.

- ReadMe should include these sections:

  - Summary
  - Data description and links to source.
  - Summary of the folders/files in the repo.
  - Check this link for a clean repo

1.4.2 Report Guidelines

This will a jupyter notebook and it will be uploaded to your github repo.

Should include the following sections:

```
- Abstract
- Introduction
- Motivation
- Related work
- Proposed method
- Experiments
- Results and discussion
- Conclusion and summary
- Limitations and later work.
- References and contributions
```

1.4.3 Presentation Guidelines

This should be very brief presentation and targeted to your fellow colleagues in the class.

I will be considering the following aspects of your presentation as I'm grading it:

- Is the motivation for the project explained in a compelling way?

- Figures are all legible, relevant and explained clearly.

- Are the ML techniques adequately discussed?

- Are the results presented adequately?

- Overall quality of slides and visualizations

- Whether slides contains too much text or not.

- Time use and pace