ECON498 Machine Learning II(Spring 2021) Homework 1 Name: _____

Due: 2th March 2021 (23:59) UID: _____

---

1. (100 points) Train a model which predict the type (Primary Type) of the reported crime in Chicago.

   - Download the crime dataset from the Chicago data portal https://data.cityofchicago.org/api/views/ijzp-q8t2/rows.csv?accessType=DOWNLOAD. It is a CSV file (approximately 1.8 GB)
   - This dataset contains data of reported crime occurred in the City of Chicago since 2001, which includes the time, location, type .....etc.
   - You do not need to use ALL the data. You should use only a subset of it. For example, you can take only the data in year 2013-2015.
   - Some of the observations are obviously problematic, you may need to drop them before you do your analysis.
   - Train a model to predict the type (Primary Type) of the reported crime. You can use everything in the dataset except for the column "Description".
   - In your repository, there should be:
     - One (or more) python file, which contains your machine learning code
     - A README.md file, which contains instructions of how to run the python files.
     - A document which contains a three pages (maximum) report of your model. You should write down the results, how you choose the parameters, how you subset your dataset, any limitations of your model.
   - The date and time give you a lot more information than merely a bunch of numbers.
   - You should hand in via Github. Please set your Github repository to a private repository and invite me to be your collaborator. Your repository should include your code and also your write up. However, DO NOT commit your dataset into the repository. I will download the dataset from the link to test run your code.
   - Again, do not print out your answers and hand in a hardcopy to me, you will fail this class if you do that.
   - There is no perfect answer for this task. You need to make simplifying assumptions to make sense of the data. Feel free to simplify the problem and write down your reasons in the write-up document.
   - The deadline is 23:59pm 2th March 2021. If you choose to hand in late (between the 3rd March 2021 and the 9th March 2021), there will be a 20% reduction in homework score.