**Deep Dive Project – GPA Analysis**

**Group 3**

**1 Group Information**

The following are our project group members and contact information.

* Leyang Zhang Email: [leyangz2@illinois.edu](mailto:leyangz2@illinois.edu)
* Lingfeng Zhao Email: [lz20@illinois.edu](mailto:lz20@illinois.edu)
* Tianyi Huang Email: [tianyih6@illinois.edu](mailto:tianyih6@illinois.edu)
* Yue Su Email: [yuesu4@illinois.edu](mailto:yuesu4@illinois.edu)

**2 Project & Dataset Description**

In this project, our goal is to predict GPA of a UIUC course based on previous course GPAs and related information. The information was stored in a dataset from <https://github.com/wadefagen/datasets/tree/master/gpa> (“uiuc-gpa-dataset.csv”).

For the dataset, we have in total 66818 rows and 22 columns, each row representing a specific course and each column is either a feature or label (GPA). More precisely, the features are:

* Year and term: numerical, discrete data. The dataset covers courses from 2010 to 2022
* Subject and course number: categorical data. There are in total 172 subjects
* Course title: nominal data
* Schedule type: categorical data. There are in total 18 schedule types.
* Primary instructor: nominal data

and the labels are

* Grade of courses from A to F and W (Withdraw): ordinal data

Not all the cells in the dataset are filled. In fact, the column “Sched Type” only has 56764 data and the column “primary instructor” has only 66640 data, meaning that they have some missing data. One reason for the missing data might be due to the FOIA requirements: “All data contained in this repository is data contained in public documents released in response to FOIA requests. Some data was excluded by The University of Illinois to adhere to privacy laws”, as mentioned by the website.

There are many ways to exploit how GPA is related to the given features. This may depend on different needs. For example, to predict the distribution of GPA for a course, we may work on the number of people getting *each* grade. To get an idea of how a general student would perform in a course, we may look into the average GPA for a course, or the number of people passing it.

**3 License**

In this project, all work follows Creative Commons Licenses.