DS 500: Data Science: Tools and Techniques
Project Description: Fall 2019
Deadline: Dec 13th, 4:00 pm (In Class) **Total Marks: 10**

As discussed during lectures, typical data science project consists of the following 6 steps. In this project, you are required to document and apply these steps on a challenging data science problem. The project is quite open and its your imagination how to efficiently able to identify problem, gather raw information about it followed by data preparation / cleaning, propose a solution with effective implementation.

- Setting the research goal
- Retrieving Data
- Data Preparation
- Data Exploration
- Data Modeling
- Presentation & Automation

Group Size is Strictly TWO.

1 Marking Scheme

1.1 Four Pages Report about Problem

[7 Points] All Students are required to write a four page report about the project in your own words describing each steps in data science process. Double column and 11 Points (Times New Roman). You report must be written in LaTeX. Your report will be evaluated based on the following. Screen shots can be added as appendix if necessary.

- Setting the research goal. The marks will be awarded based on challenging research problems (1 Point)
- Retrieving Data (1 Point)
- Data Preparation (1 Point)
- Data Exploration (1 Point)
- Data Modeling (2 Points)
- Presentation & Automation (1 Point)

1.2 Presentation

[3 Points]

1.3 Marking Criteria

Since the elements above are wide ranging, general criteria are given that are applied as a percentage to each component of the portfolio. In the following, "writing" is understood to apply both to coding and English.

(0 - 49%)

A very poor contribution showing little awareness of subject area. Lack of clarity. Communication of knowledge is either inarticulate and or irrelevant. Code fragments from the Internet may have replaced student written content to the extent that it is not possible to determine what the student has understood. Only partial functionality has been achieved.

(50 - 64%)

Knowledge is limited or superficial. Some awareness of concepts and critical appreciation are apparent, but there are major omissions or misunderstandings. Writing is not clear and there is no argument. Incorrect solutions or non-functioning software solutions have been given.

(64 - 79%)

Knowledge base is up-to-date and relevant to an appropriate breadth and depth as final year students. The student has demonstrated the ability to apply theory and concepts, across domains and identify their interrelationship. A critical appreciation is demonstrated, which is supported by appropriate references. Writing is clear if a little uneven. Source code is functional, structured and commented. Code is valid and mostly secure.

(80 - 100%)

As above but there is clear evidence of independent thought and reasoned conclusions. Literature is fully supported by citation using appropriate references and there is development of a critical appreciation of opposing arguments. Presentation of work is fluent, focused and accurate. Source code is clear and properly commented. Clearly exceeds taught material.