CDA571 Project Description
Summer 2022

1. **Project Area**

There is one key goal associated with this course. Build a data driven project.

- Go to the drawing board.
- Conceive an idea that solves a problem.
- Mine, simulate, create, scrape, (or use existing) data that can help you solve this problem.
- Solve it- leverage your skills from all your learnings (ML, DL, Statistics, Cybersecurity, Data Mining, Programming and DBMS, DMQL and so on) to infer results from data and elaborate on your findings.

There are no project topics/approaches associated with this course. You are free to choose any area you want to work on- as long as you stay within the constraints set by the "what's not allowed" and the "expectations" sections described at the end of this document.

2. **Expectations**

Form teams of 3-8 members. Each team will be assigned one instructor whom they will coordinate with through the course.

- Join the slack workspace created for the course- this will be the primary source of communication for the course.
- Nominate a team leader. This person will be primarily responsible for communicating with the instructor about the project (any other communication can be sent to instructors by any student).

Brainstorm. Pitch what you want to work on to your instructor. Get their approval.

- Ensure a strong motivation for your project, and clearly enunciate the goal/the problem the project is trying to solve.
- Ensure novelty. Ensure that the project/solution approach brings something new to the table and convince your instructor of this in the pitch.
- Keep the instructor apprised of all developments- good or bad.

Present WIP talks.

- Please refer to the video posted in the course website on what is expected in a WIP talk.
- Detail your work.

Manage expectations and deliverables.

- The 1:1 meetings with your instructors are a key component of this course. While the grading scheme mentions WIPs to be key parts of the course to be graded, these 1:1 meetings set up the WIPs.
- The 1:1 meetings will, at diverse stages of your project, help your instructor align you with the course objectives. In this meeting, your instructor will:
    - Discuss any necessary feedback/concerns.
    - Help with identifying future directions of the project.
    - Track individual contribution.
    - Understand and approve your current progress.
- The meetings (and approval from instructor) before WIP 1 talk will be included in the grading scheme of WIP 1 (same for WIP 2).
- Each project is different, and thereby, each team's grading rubric will have to be customized. The 1:1 meetings are where these customized expectations will be discussed- it is, thus, essential to be well prepared for these as well.

Deploy/Publish. Do consider some of these suggestions.

- Build and deploy a web app that anyone can access.
- Build a dashboard that can be publicly accessed.
- A deployed notebook like interface that can replicate your analysis.

This is not an exhaustive list- for many projects, this may not be possible. In your 1:1 meetings with your instructor, identify some of the best ways to showcase your project.

3. What's not allowed

The list of things allowed in this course is broad. For consistency, there are a few things that aren't allowed (for grading purposes). You are still allowed to do these, but they cannot be the core of your analysis for grading. If you believe that something on this list is essential to your project for any reason- pitch the idea to your instructor and get it approved.

- **Picking arbitrary/common dataset/project ideas and applying a few analytics components to it.** For instance, digit classification using MNIST, survivorship classification using the Titanic Dataset, and so on.

- **Arbitrarily applying a plethora of ML methods and presenting comparative results.** Justify why you pick a particular approach.

- **Using tools and not code.** Several off the shelf tools allow ML models to be fit to data without writing a single line of code. These won't be considered for grading. Eg: Zero Shot Learning from huggingface is a cool mechanism that can perform text classification without any training data- taking this approach however cannot form the core of one's analysis.

- **Plagiarism.** Academic dishonesty will be dealt with severely in accordance with the Department and University Policy. This may include a grade of 0 for an assignment and/or failure in a course.

_____