ECE 498 DS Spring 2019

Graduate Project Requirements Document

Graduate projects are an important part of the course through which we expect the graduate students in the course to delve deeper into the process of Data Analytics for solving challenging real-world problems. The important aspects in solving any data science problem are as follows:

1. Defining the problem and problem statement
2. Getting access to data
3. Data pre-processing and cleaning
4. Summarizing the data; feature-engineering
5. Building a model to solve the problem
6. Evaluating the model
    - Includes comparing with existing methods and solutions
7. Conclusion
    - Includes identifying limitations and directions for future work

For the project, students are expected to perform each of the above steps while working in a group. They are encouraged to work closely and brainstorm with Prof. Iyer and the teaching staff to ensure timely progress in the right direction. The project amounts to 1 credit hour, so we expect commensurate effort. For projects that reach a certain level of quality, we can work with you to get them published.

Below are details regarding selecting a problem statement and project deliverables.

Problem statement:

There are three avenues where you can find inspiration for project topics.

A. Mini projects
You can propose a meaningful and reasonable extension of Mini Project 1 or Mini Project 2. Having already completed the projects, you are familiar with the data and the questions that are answered using it. What are other interesting questions that can be answered with that data, or by augmenting it with data from other sources?

B. Online competitions (Kaggle)
As the website mentions, Kaggle is “the place to do data science projects”. Organizations can host data science challenges/competitions on it by sharing a problem statement of interest and data to solve the problem. Groups of people can participate in the challenge to solve the problem within a given time period and compete with other participants.
Although there are several interesting challenges mentioned on the website, we have shortlisted a few for you to pick from based on the time-restrictions of this course.
1. Challenge 1 – Microsoft Malware Detection
2. Challenge 2 – Accelerometer Biometric Competition
3. Challenge 3 – Eye Movement Verification
4. Challenge 4 – Africa Soil Property Prediction
5. Challenge 5 – How much did it rain?

You are expected to choose the data from the challenge of your choice. The challenge solutions are available, so study them and pose your own problem i.e., create your own problem statement that is different from the problem statement in the challenge, which can potentially be answered with the data. Note that the above competitions are closed. You are not allowed to pull code directly from Kaggle and must build your own model.

C. Self-motivated project
Do you have a problem in mind that you are interested in solving? For example, you could choose something that is close to your research. Or something that you want to try out just for the fun of it. Feel free to propose such an idea for the project. However, keep in mind the following:
1. The project must be new (and **not** something that you are already working on)
2. You must get access to the data yourself

In case you are looking for datasets, UCI Machine Learning Repository is a popular source.

Recommendation: Considering the limited time available for the project, we would strongly recommend choosing a problem for which you don't have to spend a lot of time pre-processing and understanding the data. This would give you enough time to work on creating a good model and interpreting and evaluating it (these are very important).

Timeline and Deliverables:

Timeline:

Mar 29 – Propose two ideas for the project (one paragraph for each project)

Apr 1-3 – Meet with course staff to discuss about project ideas and finalize one project

Apr 5 – Submit project proposal (not the same as proposing two ideas; see below for details)

Apr 19 – Submit mid-project progress report

May 6-8 – Final project presentation

May 10 – Final project submission (includes report, presentation and code)

Project proposal:

2-page report that includes background and motivation for the problem statement, description of the data, related work, and solution plan.

Mid-project progress report:

Up to 2-page report describing the methods and results (if any), and plan for the upcoming weeks.

Final report:

Up to 8-page IEEE or ACM conference style report that includes introduction, related work, methods, results and conclusions. Results should include a thorough evaluation of the proposed model and interpretation of results. Also include details about individual contributions in the group and a section on challenges faced in the course of the project.

Final presentation:

15 min presentation (10 minutes slides + 5 minutes QnA).

We will give a template for the slides. Please submit your slides 2 days before the presentation for early feedback.

Grading:

Project proposal – 15 points
Mid-project progress report – 25 points
Final presentation – 30 points
Final submission (report, presentation, code) – 30 points

Miscellaneous:

1. All submissions are due by 11:59PM on the given date. Late submission policy applies to all deadlines.
2. Unless specified otherwise, all reports should be single-spaced with font size 11 text.
3. Submissions to be made on Compass2G.