CS360 Final Project

Step 1) – Choose a dataset and teammate(s) by October 20th (**and have it accepted by me!!!!!)**

**Some starting points:**

<https://datasetsearch.research.google.com/>

<https://archive.ics.uci.edu/ml/datasets.php>

<https://www.kaggle.com/>

You will need to justify how your dataset meets the criteria for this

project (look at the data, what big questions are you trying to answer, etc).

If you can’t seem to find what you are looking for, shoot me an email or swing by office hours, I’ve got other websites I can direct you towards. If you have data from either an internship or somehow have acquired it, I am OK with you using it ***PROVIDED*** you have written permission from the owner of the data (you can prove that you have permission however you wish). When you have locked down your dataset, post your team members and dataset to piazza, this means nobody else can do that dataset!

Step 2) –

by November 15st you and your team will need to schedule an appointment with me. You’ll

present your findings and work so far, any roadblocks, status updates, etc. Think of this as diagnostic, if things are going off the rails this meeting should instill some sense of panic into you to get back on track, because there won’t be any penalties at this point. If progress is going alright, this will simply just be an update meeting.

Step 3) –

December 13th from 5 – 10 pm you will present your project to the class. Depending on how many projects we have, I’ll give you a time limit (lower bound, upper bound) but plan on approximately 15 minutes at this point. **This document will be updated once teams are formed.**

I’d recommend a nice PowerPoint, with not too many words, some graphics and trying to minimize how much code you present. **If you work with an algorithm in this class which we don’t cover (which is possible) a very succinct explanation would be encouraged.** Equations are fine to have in your slides, but you want to walk the line between technical content and having a polished presentation.

Step 4) – **Dec 13th, Submit your presentation, code, etc on Canvas. One submission per team please!**  This is also the date for teammate evaluations which will be submitted to Canvas, nobody but myself will have access to what you write down.

**Project breakdown:**

**Problem presentation [15 points]**

Describe the purpose of the data set your team chose, perhaps what reasons was it collected? For a classification or clustering algorithm would is a good measure of effectiveness?

**Related work [10 points]**

Find existing work either regarding your dataset or your general problem. What approaches have been tried? You might consider trying similar approaches but you will need to diversify your project from prior works. **If you cannot find any existing work I highly encourage you to keep looking, if I find existing work that you did not you will not be happy!!!**

**Data handling [20 points]**

Describe the data, scales, values, etc for each feature in your dataset.

<https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining>

Pictures/Graphs are great includes here. Also include statistics that might be meaningful for your data including **(but not limited too)** range, mode, mean, median, variance, counts, etc. Was your dataset balanced if it has classes, or not equal representation. What metrics are useful to explain your project?

**Teammate evaluation [0 points with possible penalties]**

Was the workload shared? Did everyone deliver what was committed? This is your chance to verify hard work of your teammates. It is anonymous to your teammates so please be honest, take credit for work you have done. Not submitting an evaluation will lower your personal grade, not your team grade.

**Presentation [20 points]**

Are you able to pose your problem and answer it? Can you give the audience some quick background knowledge/domain? What worked? What didn’t work?

**Code [20 points]**

Are you cross validating when applicable? Have you tried several algorithms? What hyperparameters did you tune? Is your code clean, and understandable? What algorithms did you not use (it can be useful to talk about why you didn’t do something!).

**Conclusions [15 points]** What are the key takeaways for your project? What could you have done better? What would you have explored with more time? Did you answer the questions you set out to answer? Did new questions arise in the middle of your project?