

Final Project
CAC 350: Machine Learning
Due May 6, 2021 by 11:59PM

Complete the following steps for a project of your choice. At each step, you should perform the work and have a brief (only as long as necessary) explanation of what you did and how you did it. Submit your code/write-up (could be a paper corresponding to the code, or you could submit a notebook with markup). The submission should be professional – I have no format requirements, but please submit something that is self explanatory.

Step 1: Choose problem domain. What do you want to create? Do you want a predictor such as housing prices in 2025 or a particular stock price? Do you want to build a classifier, a recommendation engine? Identify the ML algorithm(s) that will best suit your needs.

Step 2: Find a dataset. There are a variety of datasets available at the following locations:

- The UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml/index.php>)
- Kaggle (<https://www.kaggle.com/>) – spotify datasets can be found here
- Google data sets (<https://research.google/tools/>)
- Microsoft Research open data (<https://msropendata.com/>)
- Miscellaneous data made public by various institutions (such as the US Government) and companies.

Step 3: Explore the data. What are the features? What are the datatypes? Which features are relevant to what you want to build? How do you know? Is there any missing data? How should it be handled – delete the feature or fill in the missing data? Clean the data as necessary.

Step 4: Build your model. What hyperparameters did you choose? Why? Did you try additional values?

Step 5: Evaluate your model. How accurate is it (discuss recall vs precision)? Did you try other models to determine if something else was more accurate?

Step 6: Analyze the results of the model you think is best. Explain what the information means.