

Project 2 Proposal

Max Edwards (A02276129)

1. Goal: To use an ANN from a dataset of 13070 songs with values of duration, danceability, energy, and loudness to predict how popular the song will be. The dataset can be found here: <https://www.kaggle.com/datasets/yasserh/song-popularity-dataset> and I have it downloaded.
2. To identify these values I will use Tensorflow to create the network.
3. The deliverables will be:
 1. An ANN that, taking in inputs of duration danceability, energy, and loudness, is able to predict how popular the song will be (as based on the spotify stream number).
 2. I will submit the trained network, the data set used, the source code, and a performance review of different network configurations.
4. To ensure code portability, I will set this up in a Python virtual environment to keep packages and dependencies in sync with one another. I will also submit a video of the project running.
5. My schedule is busy as I have 2 other classes with large projects due over the next few weeks, but a reasonable schedule for me is below:

Through November 18	Through November 25	Through December 2	Through December 9
Play around with dataset and have all input/output configurations setup.	Implement basic network and train. Add performance to the report.	Add on to the dataset with new layers or regression functions.	Finalize network, retrain, analyze and submit project.

Note, I would love to submit this by the 9th so that my finals week is more open for my other classes. However, in the case of something not working or another class having issues, this schedule will leave me finals week as a bit of margin for completing this project.

6. The biggest risks to this schedule is the constraint of my other classes and their term projects. Because I will have 3 projects due at roughly the same time, I will need to be able to organize my focus and time. My best strategy is to have *something* that runs before leaving for Thanksgiving break. Then, after the break, iterate on it with longer training sets and configurations, while working on other projects. Designing the network and experimenting early will allow me to focus my time in December on running the network and working on my other classes.

Notes: In the future, I would love to see which single variable is the best predictor of popularity, but for this project I will use 4 listed above as inputs. If I have extra time, I would be interested in training networks around each variable and seeing which performs the best.