

Problem Selection:

Predicting how much a person likes to listen to classical, country, pop, or rock music based on hobbies and interests.

Data Collection:

We have one dataset which is a survey of exploring young people's interest. The survey consisted of asking musical preferences, movie preferences, hobbies & interests, phobias, health habits, personality traits, views on life & opinions, spending habits, and demographics. Out of this dataset, we'll be using the hobbies and interests data specifically to predict taste levels in 4 different types of music.

Data Preparation:

The dataset currently is categorical, with each column having a preference rating of 1-5 (very low to very high). In the interest of simplicity, we'll be compacting the rating values down to -1 (dislike), 0 (indifferent), and 1 (like), with ratings 4-5 -> 1 (like), 3 -> 0 (indifferent), and 1-2 -> -1 (dislike).

- Compact all inputs

Exploratory Analysis:

- Calculate and plot summary statistics
- Pick three hobbies/interests, break them down by each of the 4 labels (look back to p1 if you need examples)

Data Modeling:

- Pick a classification technique and build a model (look at p2 for examples)
 - Test out different parameters and hyperparameters
- Pick a clustering technique and build a model (look at p2 for examples)
 - Test out different parameters and hyperparameters

Communication:

5-10 page Report + 5 minute presentation.