

Project plan: Spotify Song Recommendations Project

Group D11

Ron-Aaron Vahtra, Reno Feliks Lindvere, Martin Koltsov

1. Exploring the dataset and understanding of patterns (everyone ~3 hours)

- After gathering the data we should familiarize ourselves with it and identify parameters that correlate with track popularity. Each team member can conduct their own research and share their findings with the rest of the team. This way we will have better intuition of parameters that affect track popularity and understand the patterns we should look for.

2. Model building (everyone ~10-15 hours)

- Decide whether we remove or keep rows with missing values.
- Figure out the way to get more testing data ([Spotify deprecated several API endpoints on November 27th 2024](#)).
- We will need to split the data into two separate parts for training and testing purposes. Try different ratios (20/80, 10/90, etc). We will use *sklearn* for these purposes.
- Try out different algorithms, also try more than one model. Definitely should experiment with Linear regression, random forest, gradient boosting.

3. Testing and tweaking (everyone ~10-15 hours)

- Define popularity more accurately (what the numerical value exactly represents).
- Decide how model should be tested, which output values should be.
- Test the model and analyze the results
- Last step is to save trained model using *joblib*

4. Simple UI (optional ~5 hours)

- If we will have enough time left we can make simple web/ui application for demonstrating purposes.

5. Project Presentation materials (everyone ~5 hours)

- Make poster and notebook which will describe research methods and the results.