**Prediction of Region-Wise Popularity for Newly Released Songs**

**Members**:

Rupali Talele (USC ID - 5048687352)

Atharva Rishi (USC ID - 6058219710)

Gayatri Atale (USC ID - 8895871778)

**The Problem**:

When a song is to be released, the artist can prioritize its release in the country where it is most likely to be prevalent, to maximize its popularity. The project will enable the artist to make an informed decision about where to initially release their song, based on the predicted popularity.

**Importance of the problem**:

Most of the time, upcoming talented artists have really good musical content to be released but they don’t have much idea about marketing strategies or how to publicize their music. The initial recognition of a new artist is a major factor in the making or breaking of the artist’s career. The project is intended to give information to an artist, who is releasing a new track, about the region their track is most likely to be popular in. As Spotify has region locked content, if the artist releases his/her song in regions where it is less likely to be popular, the song won't reach its potential. Using data provided by this project the artist can release the tracks where they are most likely to be well received and have maximum streams which will make the song popular.

**Approach**:

For obtaining the songs data, we are planning to create a dataset using Spotify API which will have a song’s general information like name, artist, album, audio features, etc. along with its country wise popularity. As we want to predict the popularity of a new track release, we will train a model on the dataset generated and use it to predict the popularity of the new release, making this a classification problem.

For classification we will be using the Random Forest Classifier by training on track audio features like energy, tempo, danceability and so on as attributes for the classification model. We will then split the dataset country wise and train the classifier on each sub-dataset. For a new song, classification will be done on each of these trained models and a popularity label for each country will be returned. The classifier will sample data from datasets and build several decision trees to return the highest voted label. We will be using the popularity measure as a label for the decision tree.

The output of the project will be a list of countries where the new song is most likely to be popular.