

**Executive Summary**

Designing a Genre prediction Tool for Music on Spotify

Soundmen Ltd.’s Data Scientists designed a machine learning system for grouping music into popular genres. This system will address the issues faced by Spotify and music listeners by providing a standardised method of grouping music based on the experiences they evoke and common compositional elements. With our product, musicians will be able to say with confidence the type of music they produce, and music lovers might be able to give a definite answer about songs in their favourite genre.

Currently there are hundreds of music categories available and most of the time, a song may fall into multiple. Consequently, these some of the constraints we faced when trying to build a system for grouping music:

* Identifying the genres that are widely accepted as the major music categories and mapping every other sub-category to these genres.
* Identifying the features of a song that aid in genre identification.
* Getting a large dataset from Spotify’s API so that we can train a machine learning model to learn how to categorise music.
* Identify which models have the highest success rate at categorising songs.
* Making an app that can run the machine learning model in a user-friendly environment.

The five Data Scientists from Soundmen successfully handled these constraints by:

* Obtaining a pre-made data set from Kaggle that compiled 27621 rows of data, pulled from Spotify API.
* Performed data cleaning with pandas by dropping rows with null values, simplified the dataframe by making each row contain one track with one unique genre(so that if a track had two genres, it would appear twice on the dataframe) and created a target genre list from a dictionary of all the sub-genres.
* Performed exploratory data analysis with Boxplots, heatmaps, and bar-charts that explored the correlations between features and identified outliers in data.
* Split the data into testing data and training data and used them to select a classification model. Also, performed metric tests that graded the accuracy of each model.
* Tested the performance of the final model on a new unseen set of songs from Spotify and obtained positive results.
* Built a final website with Tableau visualisations and a machine learning page using JavaScript, Html, CSS, python’s flask library and bootstrap 4.

Final Product

The final product was a website hosted on Heroku. It performed genre predictions by taking in 12 song features from the user and processing the input with a random forest classifier in the back end of the website. Furthermore, it provided information on the exploratory data analysis process with 3 embedded tableau visualisations and gave information on the project process in a methodology page.

Our product has been completed but there is still room for improvement. The model had a precision of 63%. Based on on-going research by a few members of our team, we are approaching a consensus that a cluster machine learning model could have served us better. Moreover, we could improve the useability of the predictor by creating a system whereby the user can select a song from a dropdown list of latest billboard songs rather than a long form that asks for features that an average person knows nothing off.