Springboard

Capstone Proposal

What is the business problem?

Every artist wants their songs to reach the top of the charts. With data readily available on Spotify, there is ample opportunity to use machine learning to predict which songs will hit or remain on their charts. The focus of this project is to identify the key factors that impact the performance of a song and create an algorithm that will predict the performance based on those factors.

Who are the intended stakeholders, and why is this problem relevant to them?

Intended stakeholders are musicians, A&R, and other associated music professionals. It will allow them to help pick songs on in an artists portfolio that they believe will have the best chance for reaching the top of the charts.

Where are the datasets available from?

Numerous data sets exist in various streaming platforms: Spotify, LastFM, Apple Music/iTunes, etc. In this project, we will use Spotify data. The weekly charts can be downloaded into a csv which can reference the Spotify API to pull additional details associated to the song via the track ID.

What data science approaches do you anticipate you will use to model the business problem as a data science problem?

This analysis will be a time series analysis with supervised regression analysis. Besides chart history, we can also include a number of features from the APIsuch as danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, duration, and time_signature. Based upon exploratory data analysis, we will determine which features show statistical correlation to the track performance on the charts.

How do you anticipate that you will evaluate the performance of each of the data science approaches that you currently envision?

To identify the best model, we will look at specific statistics to indicate the fit of the model. We will want an accuracy, precision, recall, AUC and F-1 statistic to be as close to 1 as possible to indicate how often we get a correct prediction from the model.

Certain features may be more important to the model than others. We may want to include some of these statistics when looking at their impact to the result when all other variables remain equal.

How do you anticipate that the intended clients will use the results of your CP3 to address the original business problem?

The algorithm can be used to make sure the songs are well matched to the current market preferences to preach the top of the charts. The better these songs perform, the more exposure and profit the clients can expect to receive.

Deliverables

- Jupyter notebooks
- Final report
- Presentation slide deck