HitPredict: Predicting Billboard Hits Using Spotify Data

By, Anita Tomy

CONTENTS

ABSTRACT

- The Billboard Hot 100 is the music industry standard record chart in the United States for songs, published weekly by Billboard magazine.
- Chart rankings are based on sales, radio play, and online streaming in the United States.
- In this work, we aim to predict which songs will become chart-topping hits.

FEATURES AND DATA

- We used the Spotify API to create a dataset with approximately 4000 songs.
- Spotify has an interesting API endpoint called get audio features.
- This allows us to get song features like Instrumentalness, energy, liveness,Speechiness, Danceability, Acousticness, Valence, Tempo, Loudness,
- An additional feature, Artist score is collected using Billboard API library .
- The Artist Score metric is assigned a score of 1 to a song if the artist had a Billboard hit before and 0 otherwise.

METHODS

- Data about approximately 4000 songs were collected from Billboard.com and Million Song Dataset.
- The dataset contains information about songs for Billboard Hot 100 between 1990 and 2018.
- Audio features are extracted from Spotify API.

ALGORITHMS USED

- 5 Machine learning algorthims used to predict a song's success:

1) Logistic Regression
2) GDA
3) SVM
4) Decision Tree
5) Neural Networks

CLASSIFICATION

- The dataset is split into 75% training and 25% test set for validation. **Logistic Regression** and **GDA** is then used to find the results.
- A **support vector machine** (SVM) is employed to analyze data for classification.
- **Decision tree** method is a popular tool used for classification and prediction.
- A **neural network** helps to recognize underlying relationships in the dataset through a process that mimics the way the human brain operates.

RESULTS

- Different machine learning algorithms are used to predict the accuracy of the model.
- The different models are then compared with each other.
- The model with highest prediction accuracy is thus determined.

DISCUSSION

- The analysis can be used to find out the time period which was most important for music.
- The analysis can show the song attributes that tend to be higher or lower for the most popular songs.
- The analysis can also be used to find the trends across different music genres.

CONCLUSION

- In summary, we are predicting Billboard hit songs using information about previous hits and non-hits.
- Music is evolving through time and the popularity of songs can be influenced by other factors also.So, to determine a hit song, as of now , is still a moving target.

Thank You