

Music Meets AI: Can You Classify Songs by Lyrics Alone?

Scenario: You recently obtained an internship position at Spotify because the data science department wants to develop new approaches for music genre classification. The traditional method of music genre identification through tempo and pitch and rhythm analysis does not interest your manager because he wants to know if lyrics alone can determine a song's genre.

The data science task you need to accomplish involves determining whether machine learning systems can identify music genres through lyrics analysis without audio data. Your research will assist Spotify in determining the extent to which song lyrics reveal genre differences.

Your Mission: You will use Python together with NLP techniques and Machine Learning models on the given dataset to accomplish the following tasks:

- Clean and preprocess song lyric data.
- The analysis will test both basic ML models that use TF-IDF and state-of-the-art BERT embeddings.
- Evaluate the effectiveness of these methods.
- Create a concise report that presents your research outcomes together with your findings and proposed solutions for Spotify's team.

Are you ready to uncover the defining characteristics of musical genres?

GitHub: <https://github.com/Az0202/CS3/tree/main/Music-Genre-Case-Study>