AnaLyrics Engine
Predicting Music Genres with NLP

Meghana Pakala

—

February 2024
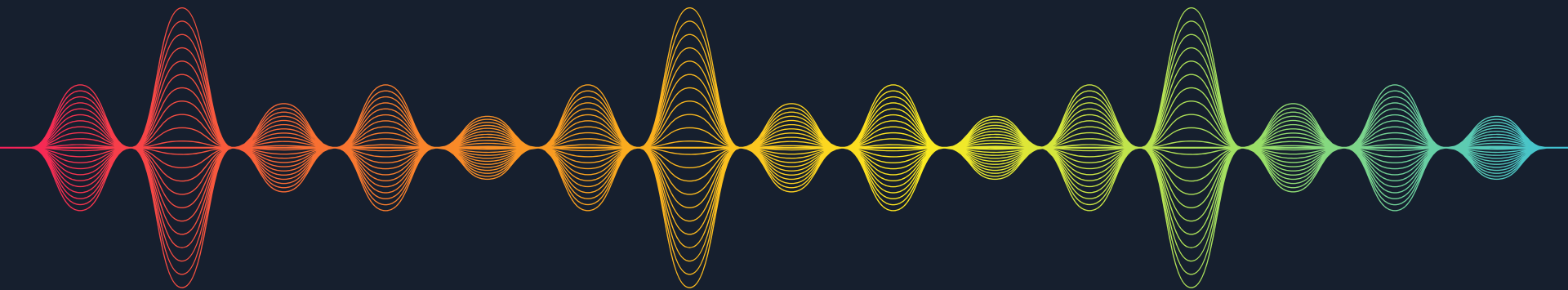
Prelude

Data Collection
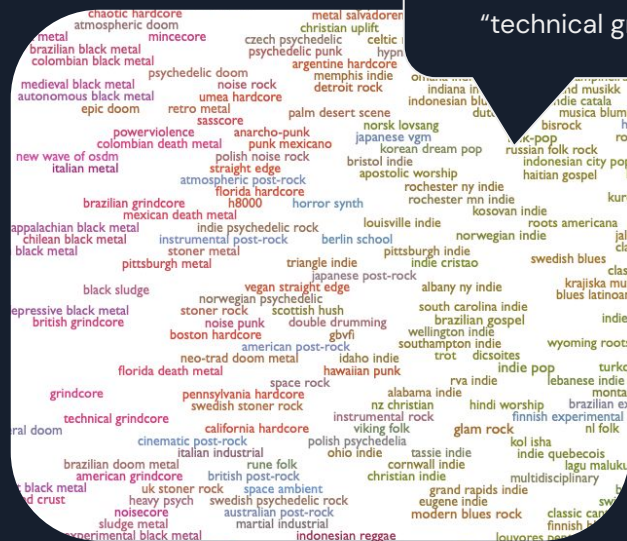
The Sound of...	Song Count	Distribution
Pop	515	25%
Rock	630	31%
Hip Hop	385	19%
Country	494	24%

"vegan straight edge"

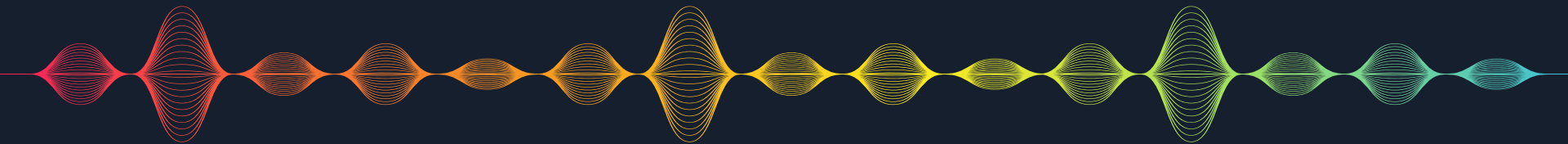"medieval black metal"

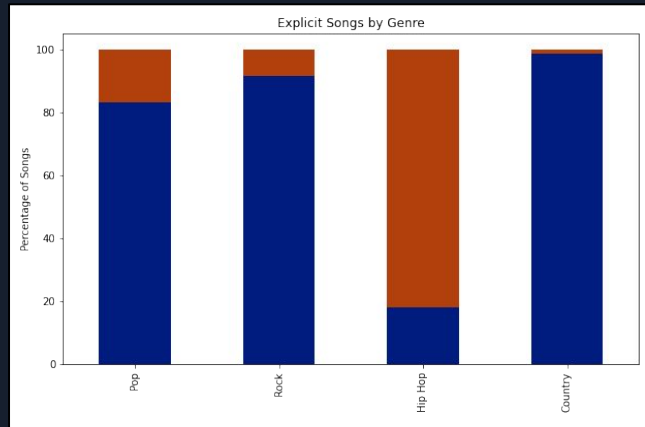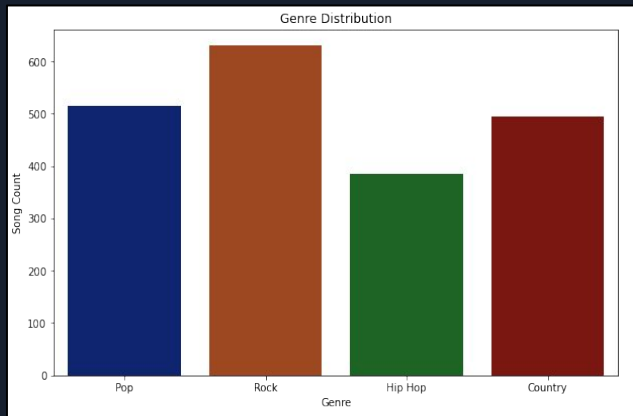"technical grindcore"
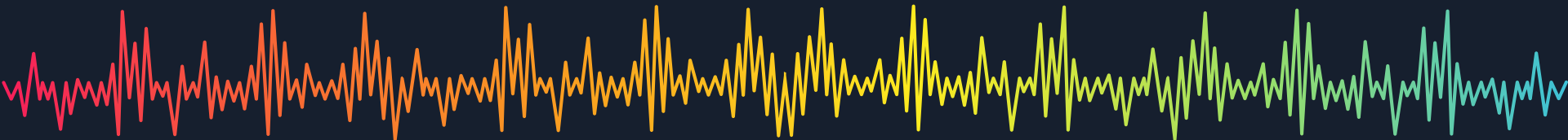
Modeling

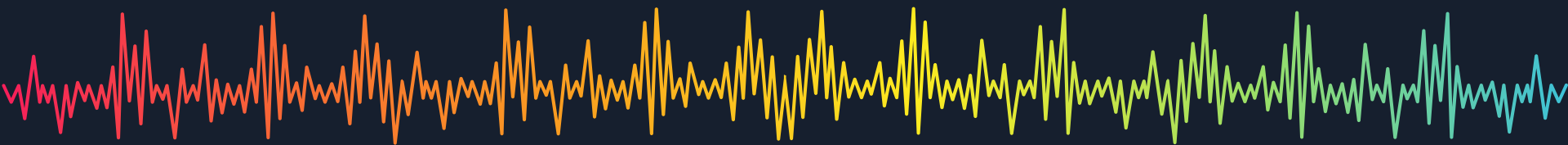**Vectorization**
- Count
- TF-IDF

Sampling
- SMOTE
- Random

Classification
- Random Forest
- Multinomial Naive Bayes

Genre Distribution

Explicit Songs by Genre

Outro

1. Collect a larger variety of lyrics to address sample size issues
2. Limit time frame to consider the evolution of language
3. Combine audio features to focus on song structure
4. Analyze performance on songs from top charts

Encore!

Q&A

Github
https://github.com/meghana-pakala/capstone

LinkedIn
https://www.linkedin.com/in/meghanapakala/