Energy of Song Prediction (from MusicCaps dataset)

Cuetessa Externship, 2025

Aryslanbek Vakilov

Introduction

Goal: Predict the perceived energy level of a song from a text description by musicians.

Why this important:

- Music apps thrive on emotional and contextual understanding.
- "Energy" helps drive mood-based recommendations.

Predicting energy can support:

- Personalized playlists
- Workout/chill mode detection
- Improved song tagging for discovery

Dataset Overview

Dataset: Music Caps with Energy Ratings

- 5,000+ songs with:
- captions (detailed description of song by musicians)
- No labels yet
- Desired target: energy_level
- Features: captions → converted into numeric features Term Frequency - Inverse Document Frequency (TF-IDF)

Why this Dataset?

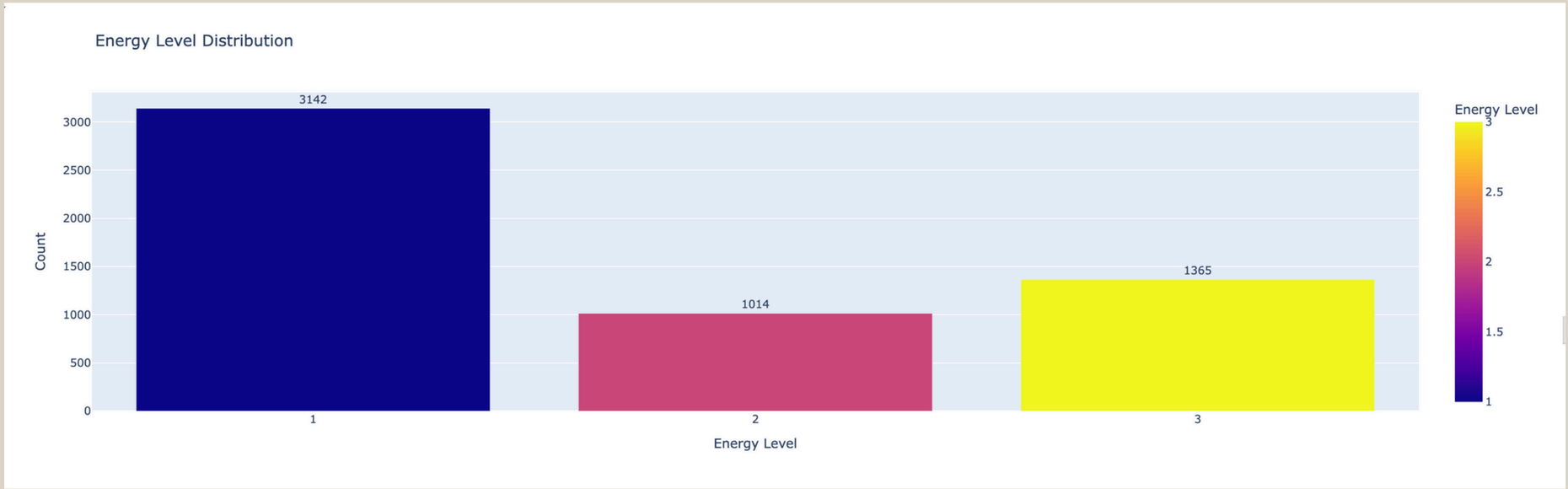
- Real-world, human-annotated.
- Bridges music and NLP.
- The text is solely focused on describing how the music sounds.

Getting the Energy labels

- "We fed the detailed description of songs into ChatGPT for it to predict "energy levels" of the dataset.
- We will leverage LLMs' powerful understanding of language to learn a simpler machine learning model to predict Energy levels
- Simpler machine learning model is interpretable and computationally efficient. Can be easily used by CUETESSA

Exploratory Data Analysis (EDA)

- Bar plot of energy_level distribution (classes 1, 2, 3)
- Class imbalance (class 1 is majority)



from the first civilizations to the most current ones,

Energy Level 1 – Low Energy

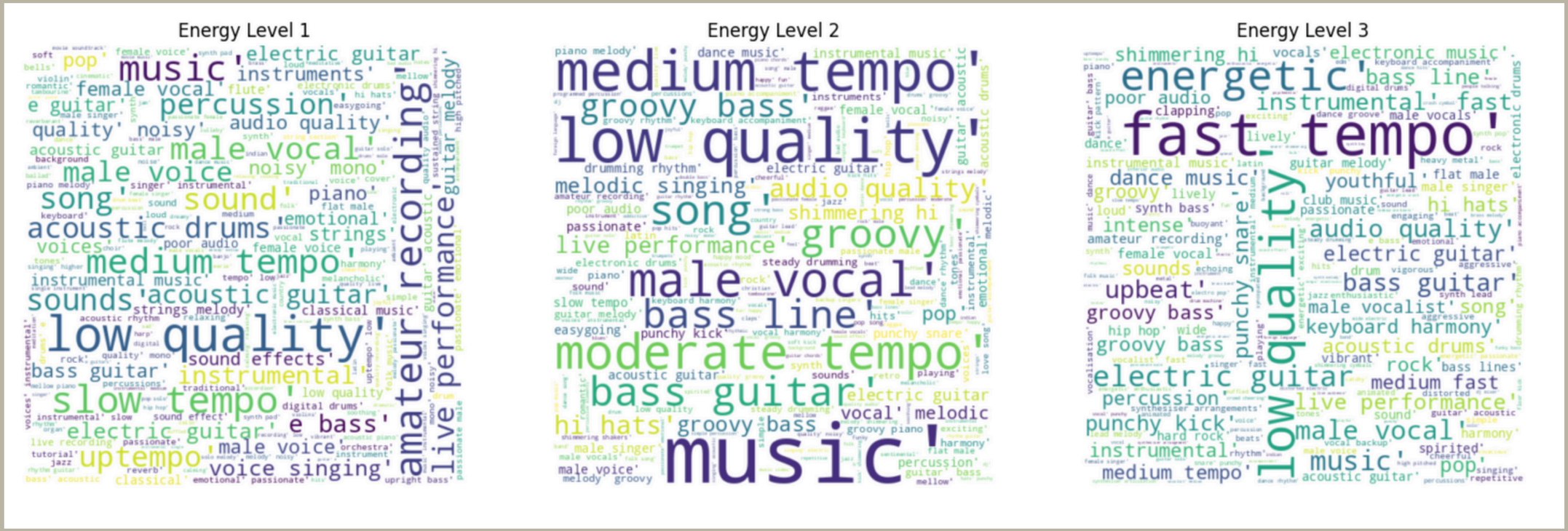
- Common terms: "low quality", "acoustic", "slow tempo", "amateur recording", "drums", "male voice"
- Indicates soft, mellow, and minimalistic compositions.
- Acoustic elements and emotional tones are prominent.
- May reflect relaxing or introspective music (lo-fi, ambient, acoustic sets).

Energy Level 2 – Medium Energy

- Common terms: "moderate tempo", "groovy", "bass guitar", "male vocal", "live performance", "shimmering hi hats"
- Suggests a balanced rhythmic profile—neither too mellow nor too intense.
- Words like “groovy” and “melodic” point to danceable but smooth rhythms.
- Associated with pop, R&B, or chill electronic genres.

Energy Level 3 – High Energy

- Common terms: "fast tempo", "energetic", "electric guitar", "bass line", "punchy snare", "upbeat", "youthful"
- Packed with high-tempo, dynamic descriptors.
- Emphasizes loud, rhythmic, aggressive elements typical of EDM, rock, hip-hop.
- Designed to energize and excite—ideal for workouts, parties, or hype scenarios.



Methodology – ML Pipeline

- Features - captions (text) vectorized with TF-IDF (Top 1000 terms)
- Target - Energy label (from ChatGPT)
- Models tried:
    1. Logistic Regression
    2. Random Forest
    3. SVM

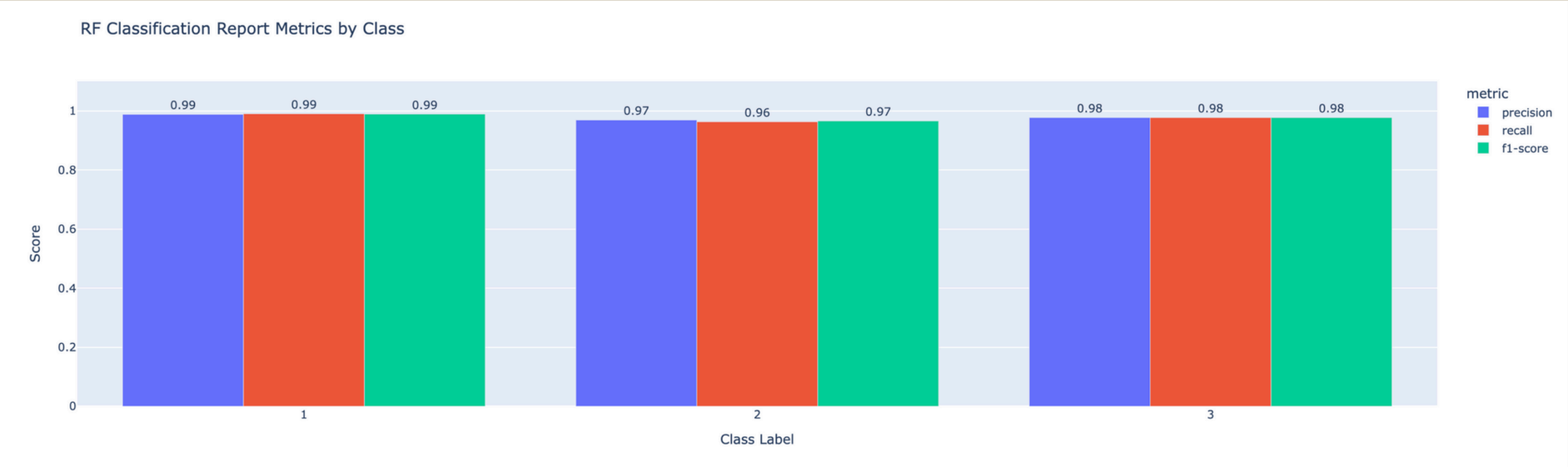
Why TF-IDF + classical ML?

- Simple, interpretable baseline
- Strong for text classification when data is not huge

Random Forest for Energy level Prediction

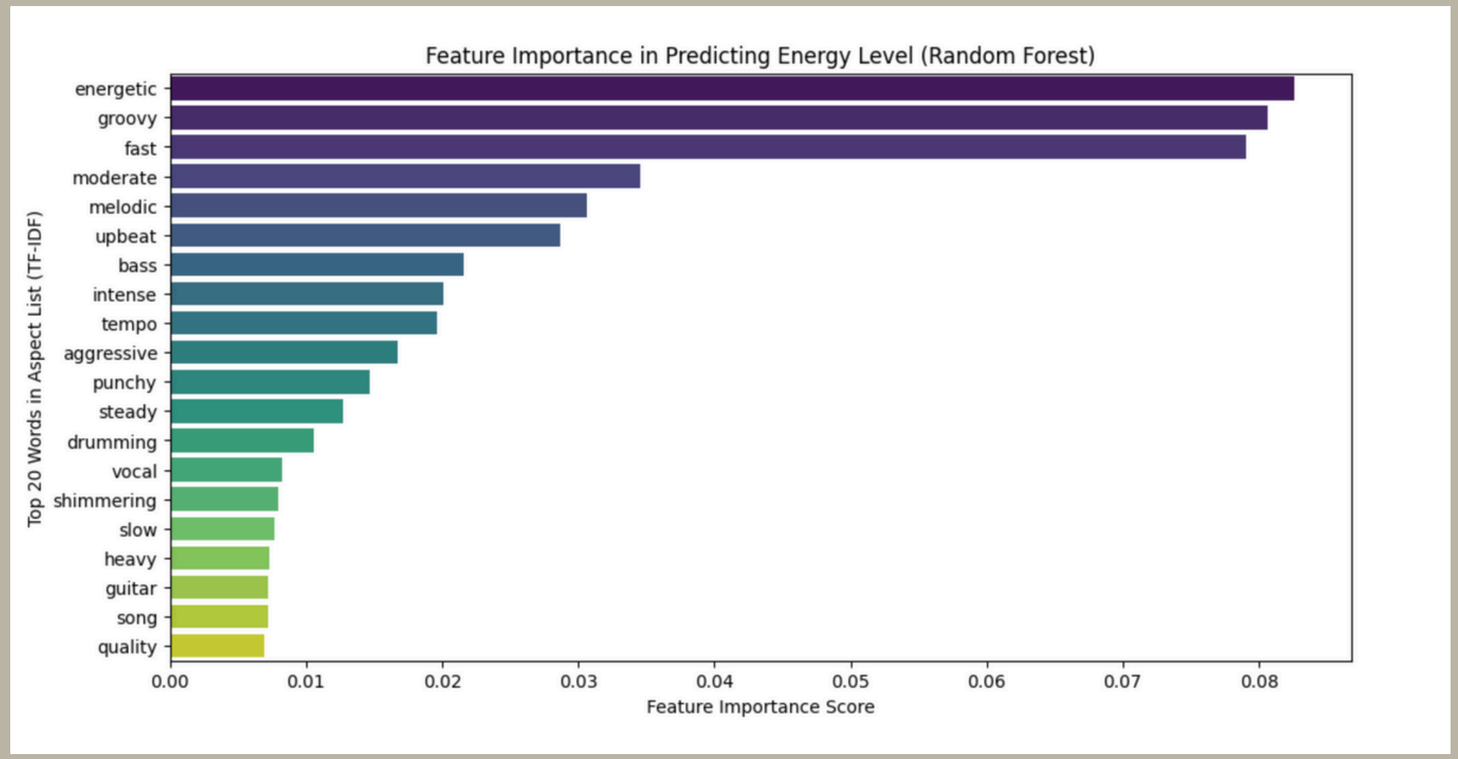
Random Forest Classification Report:

	precision	recall	f1-score	support
1	0.99	0.99	0.99	503
2	0.97	0.96	0.97	162
3	0.98	0.98	0.98	219
accuracy			0.98	884
macro avg	0.98	0.98	0.98	884
weighted avg	0.98	0.98	0.98	884



Feature Importance Interpretation with Random Forest

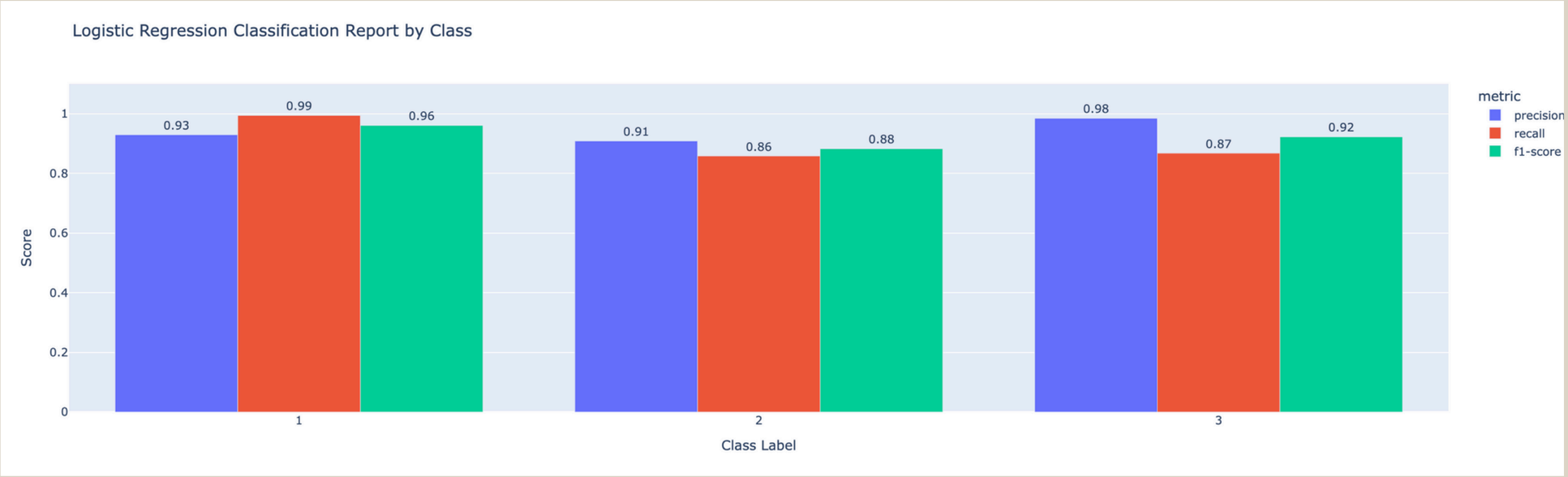
- High-importance words like “energetic”, “groovy”, and “fast” have the strongest influence on classifying songs with higher energy levels.
- Words like “moderate”, “melodic”, and “upbeat” signal medium energy, indicating controlled rhythm and musical richness.
- Words further down the list, such as “slow”, “heavy”, and “quality”, are still influential but less dominant. Some (like “slow”) may relate to lower energy tracks.



Logistic Regression for Energy level Prediction

Logistic Regression Classification Report:

	precision	recall	f1-score	support
1	0.93	0.99	0.96	503
2	0.91	0.86	0.88	162
3	0.98	0.87	0.92	219
accuracy			0.94	884
macro avg	0.94	0.91	0.92	884
weighted avg	0.94	0.94	0.94	884

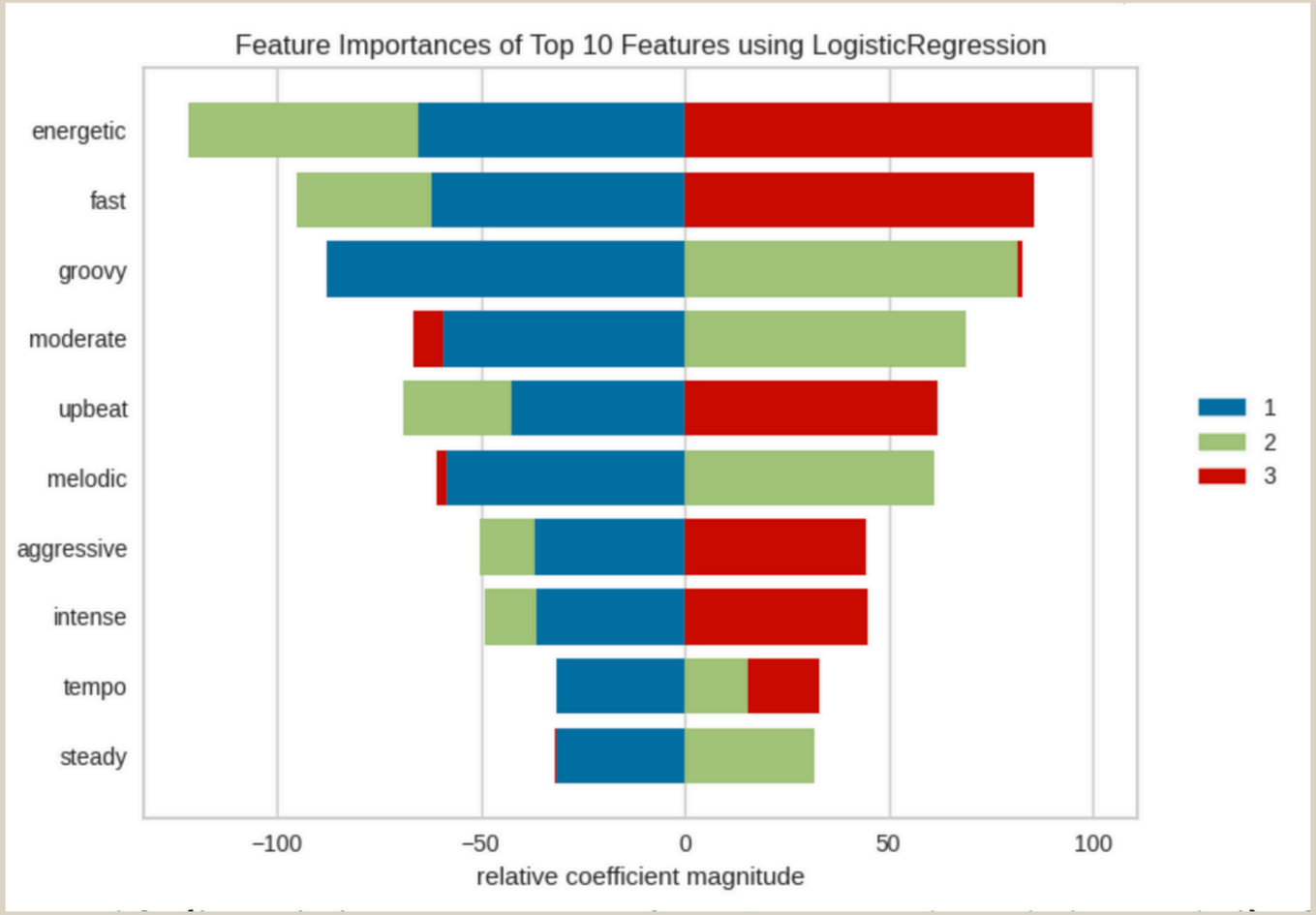


Top 10 Feature Importances (Logistic Regression)

- "Energetic", "fast", "groovy" are strongly associated with **Class 3** (High Energy).Their positive red bars indicate they increase the likelihood of predicting high-energy.

- "Moderate", "melodic", "upbeat" favor **Class 2**, showing a balanced tone or controlled rhythm.Words like "steady", "tempo", and "aggressive" show mixed contributions across classes, suggesting their influence is context-dependent.

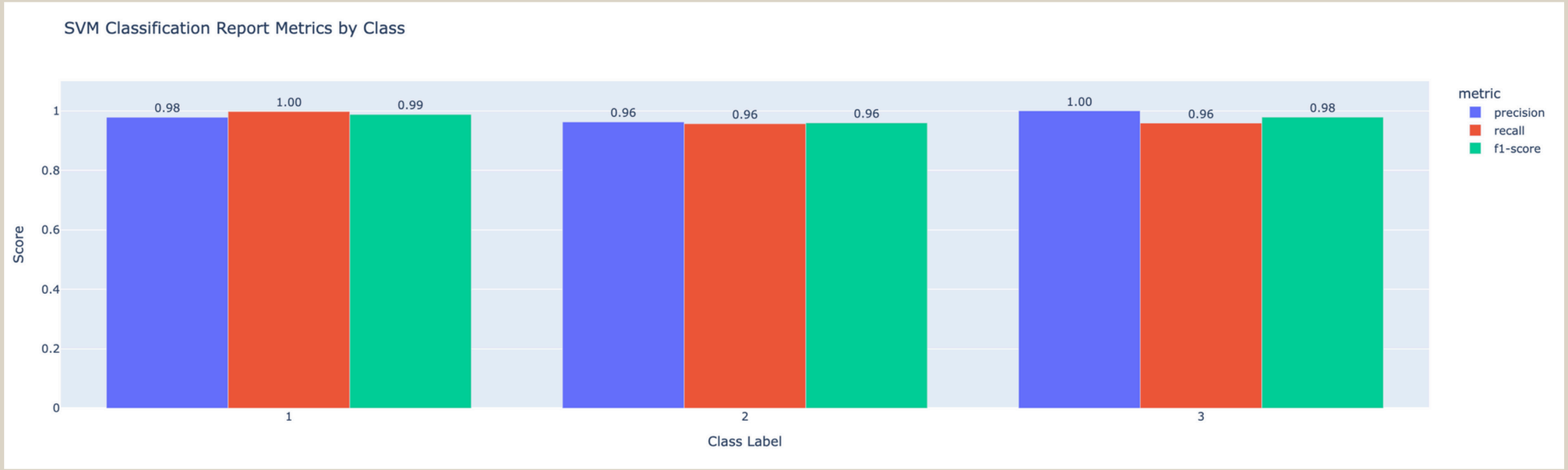
- Negative weights for **Class 1** highlight words that make a song less likely to be low energy.



SVM for Energy level Prediction

SVM Classification Report:

	precision	recall	f1-score	support
1	0.98	1.00	0.99	503
2	0.96	0.96	0.96	162
3	1.00	0.96	0.98	219
accuracy			0.98	884
macro avg	0.98	0.97	0.98	884
weighted avg	0.98	0.98	0.98	884



Conclusion

- We developed Classical Machine Learning models for energy prediction leveraging ChatGPTs ability for language understanding.
- Our models are computationally efficient and interpretable
- We can extend this model to also look at "Depth" in the future


Thank you