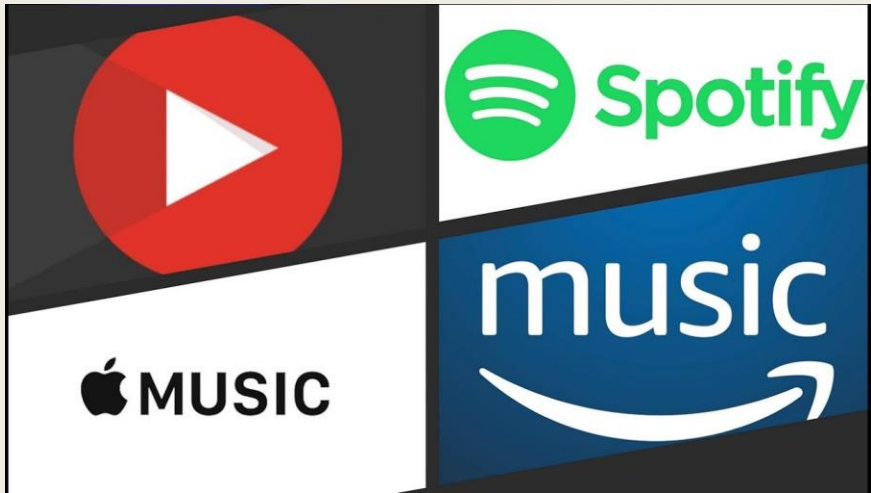
# MoodSwing

Aravind Narayanan, Akshith Rajkumar

# Introduction

- Vitality of efficient music categorization in the digital age spurred by Spotify and Apple Music dominance.
- Shift towards mood-based classification over traditional genres to meet user needs in navigating a vast musical landscape.
- Emphasis on the connection between human emotions and music through NLP and psychological elements to delve into a person's mental state.
- Exploration of emotional analysis by harnessing lyrics and NLP, diverging from traditional acoustic feature-centric approaches.
- Conducting a multi-modal analysis merging lyrical and audio features for a deeper understanding of emotional context in music classification through deep learning and NLP.

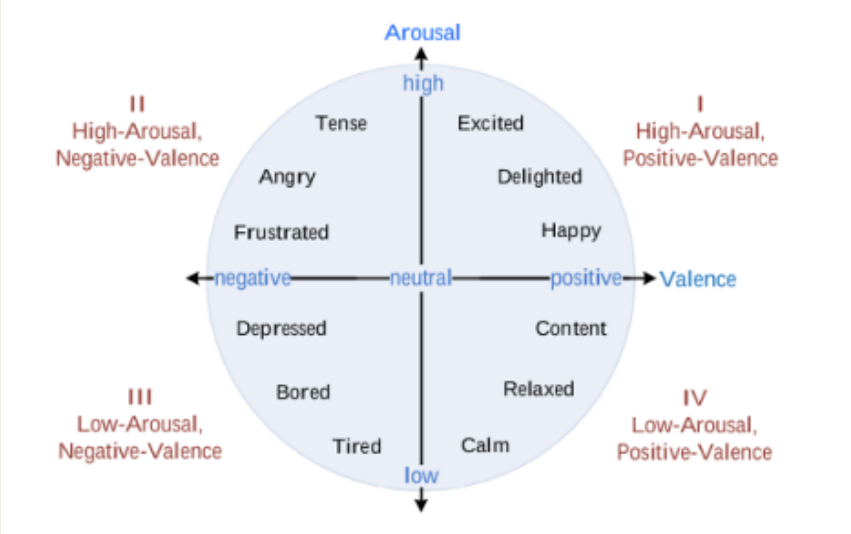


# Rubric for Classification

- Audio features have been traditionally used to classify the songs into different emotions and moods.

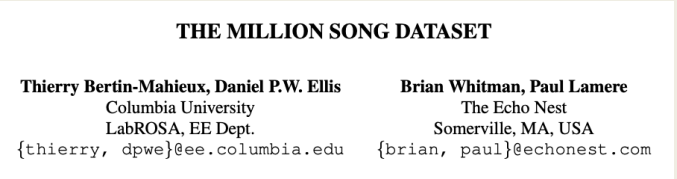
| Mood | Intensity | Timbre | Pitch | Rhythm |
|---|---|---|---|---|
| **Happy** | Medium | Medium | Very High | Very High |
| **Exuberant** | High | Medium | High | High |
| **Energetic** | Very High | Medium | Medium | High |
| **Frantic** | High | Very High | Low | Very High |
| **Anxious/Sad** | Medium | Very Low | Very Low | Low |
| **Depression** | Low | Low | Low | Low |
| **Calm** | Very Low | Very Low | Medium | Very Low |
| **Contentment** | Low | Low | High | Low |

- Using the Valence-Arousal model, we analyze lyrics to classify mood based on emotional dimensions, enhancing song classification by interpreting the underlying sentiments and tones.



# Source of Data and Data Processing

- Using 1% subset of the extensive 280GB Million Song Dataset, tailored for our project's audio feature needs.
- Songs' tags, encompassing various moods and genres, stored separately will require linkage queries.
- We intend to categorize diverse emotional tags from Last.fm into 4 quadrants based on existing literature.
- Musixmatch dataset presents lyrics in a Bag of Words (BoW) format.
- We plan to scrape lyrics from the internet for sequential arrangement, crucial for BERT or GPT2 embeddings.


THE MILLION SONG DATASET

Thierry Bertin-Mahieux, Daniel P.W. Ellis
Columbia University
LabROSA, EE Dept.
{thierry, dpwe}@ee.columbia.edu

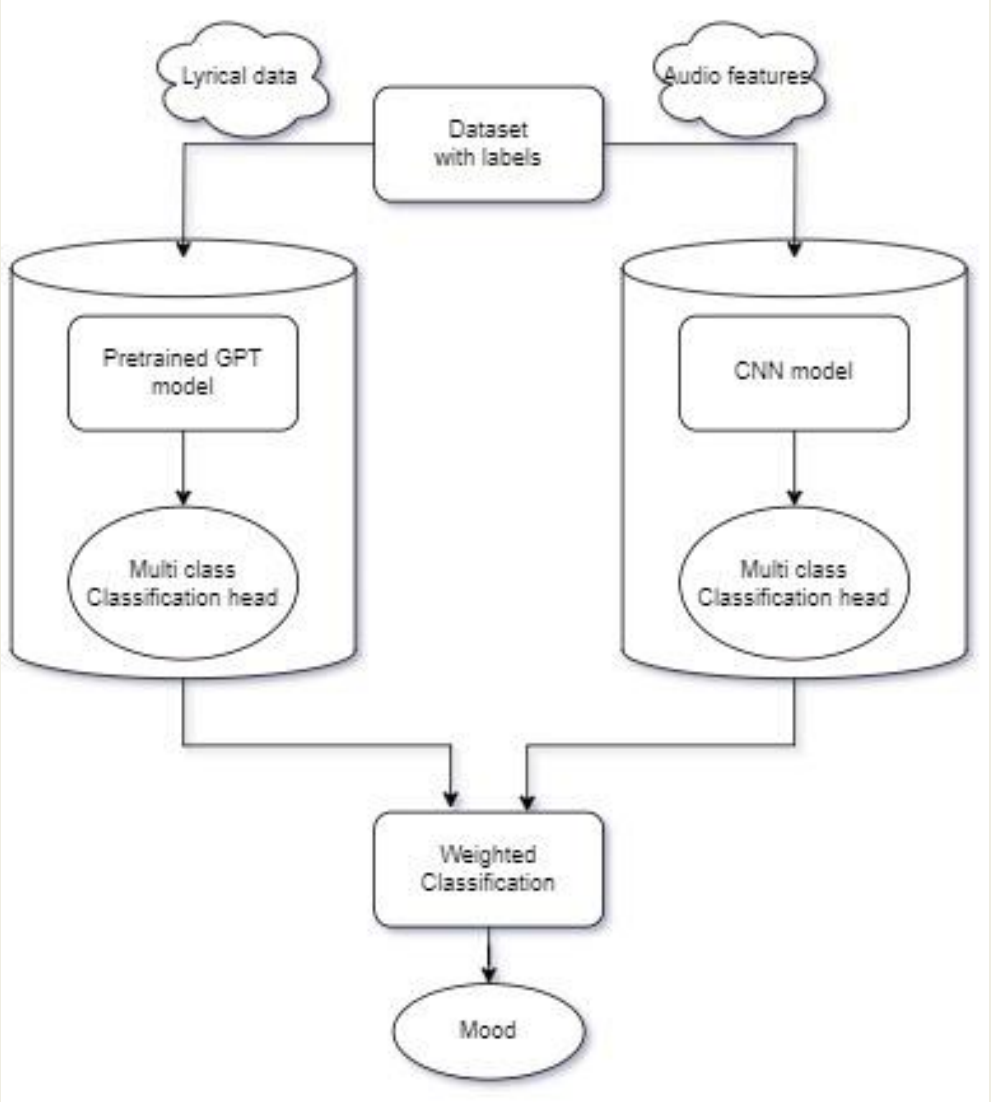
Brian Whitman, Paul Lamere
The Echo Nest
Somerville, MA, USA
{brian, paul}@echonest.com

last.fm

musixmatch®
not only words

# Model Architecture

- Our proposed model architecture is a multi-modal system.
- It comprises of CNN for analyzing acoustic features as well as a pre-trained GPT2 model for analyzing the lyrical data.
- The outputs of the models are combined to finally present a weighted classification of moods.



# Baseline Model and Comparison

- The baseline models used for comparison are from paper titled '*Transformer-based Automatic Music Mood Classification using Multi-modal framework*'.
- The paper presents performance metrics (F1 score, Recall and Precision) and accuracy for various models.
- The plan is to compare results with these published results.

| Model | **Scheme** | **Overall Accuracy (in%)** |
|---|---|---|
| M0 | Acoustic+Textual features (SVM) | 59.55 |
| M1 | Acoustic features (Bi-GRU) | 56.00 |
| M2 | Acoustic features (CNN) | 61.76 |
| M3 | Textual features (Word2Vec+Bi-GRU) | 44.15 |
| M4 | Textual features (GloVe+Bi-GRU) | 50.73 |
| M5 | Multi-modal fusion (Word2Vec+Bi-GRU) | 63.97 |
| M6 | Multi-modal fusion (GloVe+Bi-GRU) | 65.44 |
| M7 | Multi-modal fusion with single attention (Word2Vec+Bi-GRU) | 73.52 |
| M8 | Multi-modal fusion with single attention (GloVe+Bi-GRU) | 76.47 |
| M9 | BERT | 58.08 |
| M10 | XLNet [27] | 57.25 |
| M11 | CNN+BERT [28] | 71.32 |
| M12 | Bi-GRU+BERT | 73.50 |
| M13 | **Proposed multi-modal transformer-based** | **77.94** |

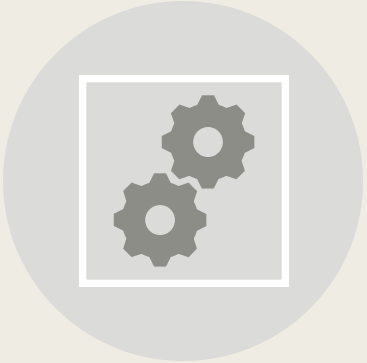
# Risks

- **Limited Lyrics Access:**
  - Copyright constraints might limit complete lyric access, resulting in a smaller dataset or mismatched audio-lyric pairs, affecting performance unpredictably.
- **Audio Feature Reliability:**
  - Existing audio features may underperform, potentially necessitating reliance solely on lyrical data in case of poor performance.
- **Dataset Size and Complexity:**
  - Large individual lyric sizes could pose computational challenges during model training. Mitigation includes exploring optimization methods or potential dataset size reduction.

# Plan

DATASET COMPILATION

**1 week**



MODEL IMPLEMENTATION

**2 weeks**

WEIGHTED LYRICS MOOD CLASSIFICATION

**1 week**

**Expectations for November 21st:**

- A complied dataset that is ready to be trained on.
- One trainable model, either the CNN model or the Transformer model with inferable results.