**Trend Analysis Of Emotional and Psychological Themes in Music over time and across genres**

**Harsh Ramani, Dhruv Charan, and Aryan Patil**

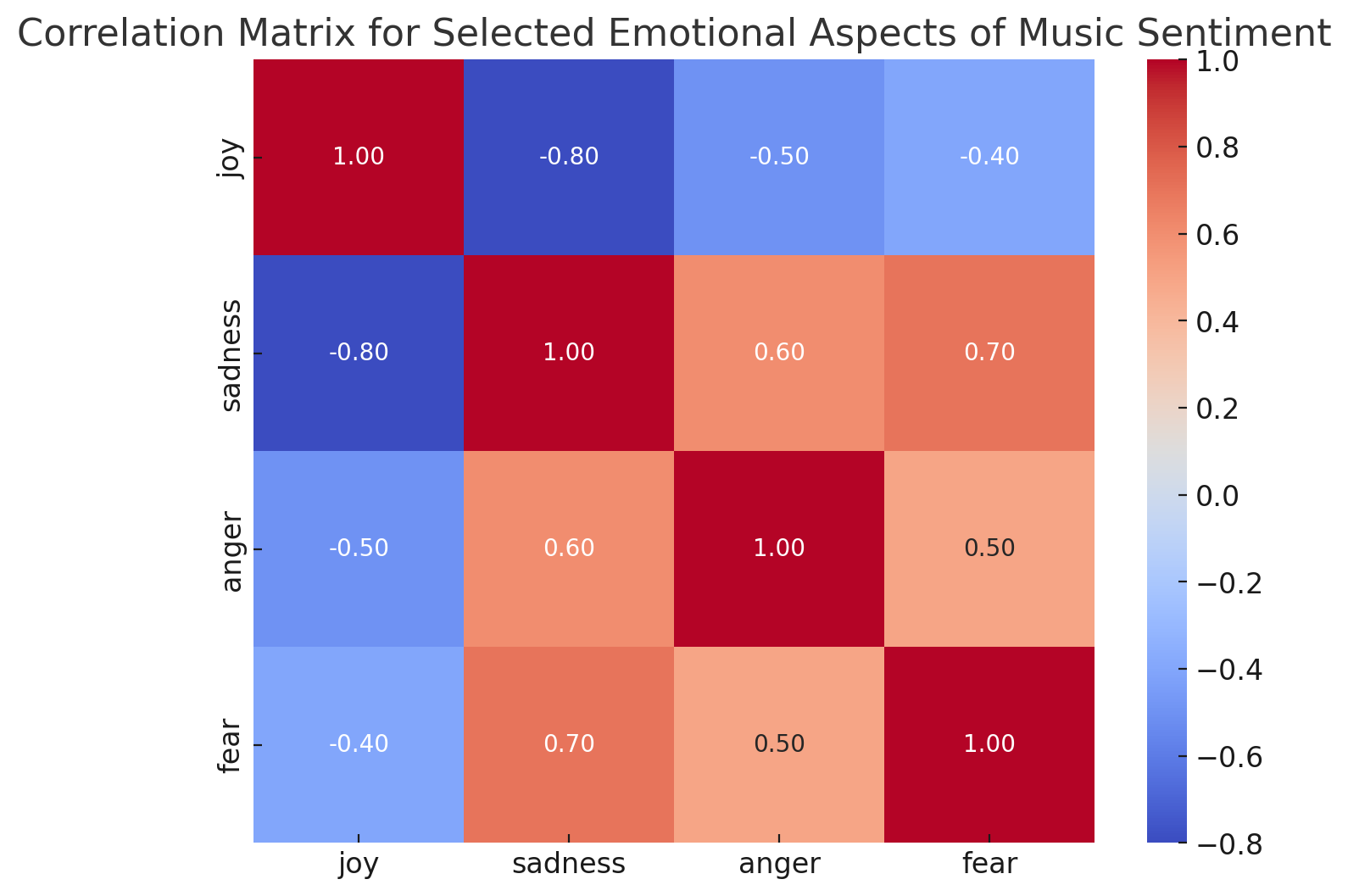
**Introduction:** We will analyze the sentimental and emotional content of music lyrics across genres, eras and artists using BERT and XLNet[[1]](#ulug1koxwqi3), to develop a comprehensive understanding of how human emotion is portrayed in one of the major cultural forces of the last century- music. This allows us to see the evolution of how individual and societal feelings are encapsulated within such an important medium of self expression. This can give us an insight into how emotional intelligence and music are intertwined with each other and can give us an insight into the psychological responses the lyrics aim to evoke.

**Background:** Inspired by Glenn McDonald’s ‘Every Noise at Once’ project at Spotify[[2]](#nub0ssh7te2i), which created a musical-genre space for user exploration, our aim is to delve into lyrical content for enhanced music discovery. By analyzing lyrics, we aim to uncover insights that could lead to a platform facilitating content discovery and stratification.

**Dataset**: We'll be utilizing the chloeliu/lyrics[[3]](#3efizwvpghh0) dataset from the Hugging Face datasets library for our project, which includes stemmed song lyrics along with crucial metadata like artist, genre, and release date. Additionally, it features musical aspects such as acousticness and instrumentalness, scored from 0 to 1. This dataset, with 28,372 total entries, is ideal for capturing thematic, emotional, and musical elements in song lyrics. We'll use the first 25,000 entries for training and the remaining 3,372 for testing across all defined tasks.

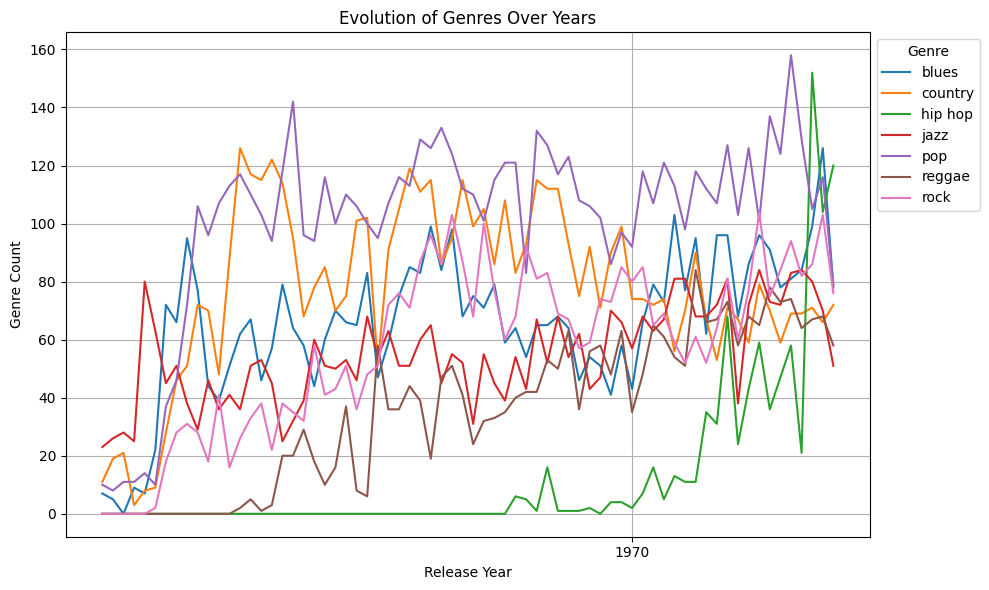
**Method:** We will be performing sentiment analysis and personality detection on the lyrics from our dataset using two models- BERT and XLNet[[4]](#8uleu93ogihy) to classify songs into different sentiments. We will also use the probabilistic LDA model to perform topic modeling on this dataset to be able to reference sentiment information with the topic classifications. Lyric Generation will be a Language Modeling task which will be achieved by training a language model(GPT) on existing lyrics. Lastly, trend Analysis will show how lyrical themes and sentiments have evolved over time which will be the base for many music Recommendation systems . Our choice of the BERT and XLNet models stems from their utility in tasks like classification due to their strong performance in understanding the context of language. Logistic Regression will serve as a baseline model for the classification task, topic modeling, while we use 2D CNN for the remainder.

**Potential Results:** Our results would be derived from the insights gained from the sentiment analysis in terms of trends over time for themes and sentiments, and also like comparative analysis of the performance of BERT and XLNet on our tasks over various metrics.

****

*Fig 1: Comparing a few emotions to identify correlations between the musicality and the sentiment insights.*

**Takeaways:** This can help uncover psycholinguistic patterns in song lyrics, which can help us learn more about the emotional response to music and lyrics and the interplay of genres and sentiments.



*Fig 2: Time series demonstrating the fluctuation across various genres, over years, capturing different patterns*

**References:**

[1]Singh, S., Singh, W. AI-based personality prediction for human well-being from text data: a systematic review. Multimed Tools Appl (2023). <https://doi.org/10.1007/s11042-023-17282-w>

[2]<https://everynoise.com/>

[3]https://huggingface.co/datasets/chloeliu/lyrics

[4]Erdenebileg Batbaatar, Meijing Li, and Keun Ho Ryu. 2019. Semantic-Emotion Neural Network for Emotion Recognition From Text. IEEE Access, 7:111866–111878.