Project Report Title: AI-Powered Music Generation Using Emotion and Artist Style (Done by: Sujal thakkar)

Abstract:

This project aims to develop an AI-powered music generation system that combines emotional content and artist style to create unique music tracks. By leveraging deep learning techniques, we intend to enable the generation of emotionally expressive music that can reflect the style and preferences of specific artists. This project addresses the growing demand for personalized music experiences in a digital age where people seek more tailored content.

Overview:

The primary motivation behind this project stems from my deep passion for working in the field of Artificial Intelligence. I identified a significant gap in the market where users face difficulties in finding personalized songs tailored to their preferences. To address this, I decided to take initiative and create a solution myself.

Initially, I selected well-known Hindi/Bollywood singers to build a comprehensive dataset. I chose 20 artists, each with a catalog of over 20 songs, resulting in a total of 400 songs from the Hindi music industry in India. To add diversity to the dataset, I extended the process to include English songs as well. I picked around 22 artists and gathered about 20 songs per artist, totaling nearly 440 songs. This diverse dataset forms the foundation of my project.

Given the large size of the dataset, I adopted a "divide and conquer" strategy, breaking it into manageable parts and tackling each one individually. As a result, you'll find several IPython notebook (IPYNB) files on my GitHub page. These files include:

1. **Data Collection and Conversion**: This file collects songs from YouTube and converts them into MP3 format, as I only needed the audio for playback.
2. **Emotion Extraction**: Using Wikipedia as a source, I extracted emotions associated with each song. For some songs without available emotion data online, I manually assigned emotions using random values, ensuring high accuracy.
3. **Feature Extraction**: This IPYNB file converts MP3 files into numerical formats, managing any missing information to create a complete representation of each song.
4. **Model Evaluation and Music Generation**: The final file contains the models used to evaluate and generate new music, leveraging the features and emotions extracted from the dataset.

What is the problem?

The primary problem is to design an AI model capable of generating high-quality music that mirrors both the emotional content and stylistic traits of specific artists. The challenge lies in understanding and extracting these traits from various data sources, such as audio tracks, lyrics, and artist information, and translating them into meaningful musical outputs. This problem is particularly interesting because it opens up new avenues for music personalization and creative expression using AI.

What is the approach you propose to tackle the problem?

The approach to tackle this problem involves using deep learning techniques, specifically generative models like Variational Autoencoders (VAEs) and Recurrent Neural Networks (RNNs), to generate music based on the emotional content and artist style. These models are well-suited for capturing sequential data and can learn the stylistic elements of different artists from training data. The VAEs can encode the emotional content of music, while RNNs can help model the sequence of musical notes and chords that define an artist's style.

What is the rationale behind the proposed approach?

The proposed approach is grounded in the idea that music can be represented as a sequence of notes and chords, which can be learned by RNNs, and as a set of emotional cues, which can be encoded by VAEs. Previous works have explored using deep learning for music generation, but they often focus solely on style or emotion without combining both aspects. By integrating VAEs and RNNs, our approach aims to provide a more holistic model that considers both the emotional content and stylistic traits. This differentiation from previous works is key as it addresses a more complex problem space: generating music that resonates emotionally and stylistically.

What are the key components of the approach and results?

The key components of the approach include:

1. **Data Preprocessing**: Extracting features such as emotional tags and stylistic elements (e.g., rhythm, harmony) from audio tracks.
2. **Model Architecture**: A combination of VAEs for emotional encoding and RNNs for style modeling. The VAEs will encode the emotional information and generate latent vectors that represent the emotional profile. The RNNs will then synthesize the musical sequences that correspond to the stylistic traits of the artists.
3. **Experiment Setup**:
    - **Dataset**: We use a curated dataset containing music tracks, artist profiles, and emotional annotations.
    - **Computing Environment**: Experiments are executed on a GPU-equipped server with TensorFlow/Keras for deep learning model implementation.

- **Model Parameters**: We use a 2-layer LSTM for RNNs and a 2-layer encoder-decoder for VAEs with latent size 128.

Experiment Results:

Main Results:

The main result of the experiments shows that the proposed model can generate music tracks that retain both the emotional content and stylistic traits of the artists. Evaluation metrics, such as user feedback and similarity measures to the original music, suggest that the model successfully captures the stylistic preferences and emotional context of the music.

Supplementary Results:

Parameter choices like the latent size in VAEs and the number of LSTM layers in RNNs were chosen to balance complexity and performance. The selection was guided by preliminary tests to ensure the models can capture meaningful patterns without overfitting.

Discussion:

The results are promising, showing that combining VAEs and RNNs can effectively generate music that combines emotional and stylistic aspects. Compared to existing methods, which may focus solely on style or emotion, our approach offers a more nuanced generation process. Future work could include enhancing model robustness, expanding the dataset to include more diverse genres and emotional expressions, and improving user interaction mechanisms.

If the results are less satisfactory, it may be due to the limited scope of the dataset or the choice of model parameters. Further exploration into hybrid models and the integration of other forms of data, such as visual cues or user feedback, could improve performance.

Conclusion:

This project demonstrates the feasibility of generating personalized music by combining emotional and stylistic elements through AI. The approach offers a promising path forward for music personalization and opens up new possibilities for the creative application of AI in music. The ultimate goal is to create a system that can autonomously generate music that is not only technically sound but also emotionally resonant and stylistically relevant.

References:

- Link of the whole data
- Link of X% of Dataset
- Papers Read:

    - https://arxiv.org/abs/1706.03762
    - https://arxiv.org/abs/2305.10435
    - https://arxiv.org/pdf/1906.02691
    - https://arxiv.org/abs/1809.04281
- My GitHub repo