**Background: (PPT 3)**

As an ancient art form that profoundly influences human emotions, music plays a central role in various cultures and societies and is essential in our emotional expression and experience.

**Significance:(PPT 4)**

My project aims to disentangle the emotions conveyed by the lyrics and audio features of music. By developing a model that captures these elements, I hope to deepen our understanding of the emotional impact of music. This can enhance emotional well-being and may help address mental health issues.

**Research: (PPT 5)**

Besd on my motivation and my aiming My research focuses on music emotion recognition (MER) is a field in music information retrieval. MER takes inspiration from the Russell emotion model and provides a more specific classification of emotions. My goal is to create a more advanced MER model that considers both lyrics and audio. I build on an important paper in the field that uses a Bi-directional Long Short-Term Memory (Bi-LSTM) model with GloVe word representations and achieves 91.08% accuracy in classifying sentiment from lyrics. My goal is to go beyond this benchmark and improve the accuracy and wider applicability of the model.

**Method:**

**Datasets: (PPT 6-7)**

My project uses two music sentiment datasets, "MoodyLyrics" and "MoodyLyrics4Q", both of which are divided into 4 categories based on the Russell sentiment model. The Dataset1 uses only the emotional dimension of the lyric text as labels, while the Dataset2 uses the overall music labels of Last.fm.

Due to copyright issue, I collected the lyrics myself, initially using the Genius API, but later designed a custom web crawl in order to be more accurate. This method ensures the correct match of song and artist and reduces the risk of wrong lyrics.

**(PPT 7)**

I also used the Spotify API to extract audio features from the songs. After collecting the data, I cleaned and standardized the lyrics, using custom regular expressions to eliminate unnecessary elements and filter out non-English lyrics.

**(PPT 8-9)**

Because the data distribution, I decide to balance the first dataset, I randomly removed 90 songs from the "happy" category, keeping a specific random state to ensure repeatability. Before training the model, I shuffled the entire dataset to avoid any bias in the original order of the data.

**Reproducing the paper (PPT 10)**

**1.**The next key stage of my project is to reproduce the research papers of Bi-LSTM and GloVe from the literature review, which is a fundamental aspect of scientific validation. This step is able to verify the reliability and validity of the original research results. It provides deeper insights into the methodology and logic of the original study, as well as identifying potential problems and directions for improvement.

**2**.In the reproduction phase, I strictly followed the hyperparameters and structure detailed in the paper. The study utilizes pre-trained GloVe 100-dimensional vectors for word embeddings.

**3**.However, for the Naive Bayes (NB) model, which cannot handle negative values, I chose the TF-IDF method for word embeddings, while the other models continued to use GloVe.

**4**.Following the guidelines of the paper closely, I was able to reproduce several models including Naive Bayes, K-nearest Neighbors, Support Vector machines, Convolutional neural networks, Long Short-Term Memory networks, and bidirectional long short-term memory networks, achieving similar accuracy as reported in the original paper.

**Experimental design: (PPT 11)**

Based on my analysis and result of the paper, I set out to improve the performance and accuracy of the model through different experimental methods. This involves three key areas: embedding techniques, preprocessing techniques, and incorporation of audio features.

**Word embedding: (PPT 12)**

**1.**First of all, I reduced the maximum sequence length (max\_len) from 1000 to 250 based on the actual text sequence length in the dataset to minimize the effect of zero-padding. Then I realized that for repetitive and rhythmic texts such as lyrics, it may be effective to experiment with words and smaller contexts.

**2.** So next I tried various word embedding techniques for the algorithms in the paper, including Bag of Words (BoW), Term Frequency-Inverse Document Frequency (TFIDF), and Word2Vec300d, setting up a unified preprocessing step, lemma + lowercase + noise removal, and stop word removal. This ensures that the data is clean and consistent.

**3.**In addition, the parameters of the machine learning model and the deep learning model are fine-tuned to optimize their performance using grid search and learning curve and loss curve.

**4.**Experimental results show that BoW, TFIDF, and Word2Vec outperform GloVe, exceeding the baseline accuracy of 91% of the original paper.

**Preprocessing:** **(PPT 13)**

**1.**I experimented with the four top performing models using different preprocessing strategies from the embedding stage.

**2.**This includes exploring the impact of techniques such as stemming, lemmatization, noise remove, and stop word remove.

**3.**For deep learning models such as Text-CNN and BiLSTM, I adopt a cyclic testing approach to account for random weight initialization during training to ensure stable and credible results.

**4.**With these approaches, I have achieved an important milestone: models like SVM, Naive Bayes, and Text-CNN have surpassed the baseline accuracy of the original paper in single lyrics analysis. **(PPT 14)** Notably, the SVM model achieved an impressive accuracy and F1 score of 94%. This phase is crucial and demonstrates the potential of improved embedding and preprocessing techniques to improve the accuracy of lyric emotion recognition.

**Audio feature exploration: (PPT 15)**

**1.**After studying word embeddings, tuning, and preprocessing, I focused on studying the impact of audio features on sentiment. Initially, I was prepared to understand the data first. Due to the particularities of the data, I standardized the audio features to maintain the consistency of the data, which is crucial for model training and analysis.

**2.**My analysis started by visualizing the distribution and clustering test of audio features, using tools such as heatmaps, PCA(principal component Analysis), and t-SNE (t-distributed Random Neighborhood embedding), to assess the clustering trends by the Hopkins statistic. Although the Hopkins test contains some clusters, dimensionality reduction techniques such as PCA and t-SNE show some overlap and randomness in the data points. **3**.Heatmap analysis and distribution showed limited correlation between audio features and emotion categories, for example, energy and loundess had the strongest positive correlation with the angry class but the other features did not have strong positive or negative correlation.

**4.** Subsequent investigations using heatmaps and random forests models on specific audio features such as keys, modes, and time signatures showed little contribution to sentiment classification and decided to drop in order to avoid the curse of dimensionality.

**5.**Given these findings, I excluded these features from future model training. Models trained only with audio features did not produce satisfactory results.

**(PPT 16) DATASET1数据可视化展示**

**(PPT 17)**

**6.**I then incorporated audio features into the previously successful model. This includes pre-fusing features for the SVM model and adding an audio feature input layer to the Text-CNN and BiLSTM models. I also experimented with stacking ensemble learning methods, using random forest for audio features, TFIDF and SVM for text, and XGBoost as a meta-classifier.

**7.**Subsequent experiments show that audio features do not significantly improve the performance of the model. It is possible that this is because Dataset 1 is labeled according to the lyric sentiment dimension and the audio features do not have a very strong influence.

**8.** To test whether the audio features enhanced the generalization ability and performance of the model, subsequent performance tests using Dataset2 confirmed this.

The overall accuracy on the test set is not particularly high this is probably due to the fact that there are 7775 Unseen Tokens in the test set which is related to the vocabulary size of dataset 1. However, for the results it is found that the integrated model is generally better than the single-modal model, especially on the CNN model, the integrated model achieves an F1 score of 38%. The f1 score is significantly improved for the angry class. This may be due to the positive correlation between the energy and loudness of dataset 1 and the angry categories learned by the model. This proves that considering both audio and lyrics can improve the generalization ability of the model.

**9.**To further validate my findings, I trained the composite model on dataset 2 and compared it to a benchmark study using XL-NET and lemma. My model outperformed the F1 of benchmark 59, achieving an F1 score of 67%. Also the single model for lyrics and audio features did not outperform the combined model.

**10.** According to this result, Then I found a strong correlation between audio features and emotion categories in dataset 2; The PCA and heatmap results show the clustering and correlation between audio features and sentiment, which verifies the effectiveness of the fusion of lyrics and audio features for music sentiment classification

**(PPT 18) DATASET2数据可视化展示**

**Use case: (PPT19）**

**1.** In order to verify the potential and generalization ability of the model. I used my model to analyze the sentiment of the top 100 Spotify songs from the last decade.

**2**.I used a cnn to perform the experiments because it cross-validated the best on both datasets. **(PPT 19)综合CNN Architecture展示**

**(PPT19）**

**3.** The results of the model predictions reveal some interesting trends, such as sad songs reaching 10-year highs in 2020 and 2022. Happy songs gradually decreased from 2020 and reached a near 5-year low in 2022, which may reflect global unrest and negative emotional events such as the COVID-19 and the Russia-Ukraine conflict in 2022.

In summary, the model's predictions reveal significant trends in music preferences in recent years, in particular an increase in the popularity of sad songs and a decrease in the popularity of happy songs, which may reflect global socio-political events. These findings highlight the complex relationship between social emotions and musical taste.

**Evaluation and Discussion: (PPT20）**

**1**

In order to improve the accuracy and effectiveness of music sentiment analysis, I adopt a comprehensive approach. The process started with a two-stage lyric collection method, initially using the Genius API and later moving to custom web scraping to improve accuracy. To ensure a high quality dataset, downsampling techniques are applied for balancing and consistent random shuffling is used to prevent bias in training.

2

The project involved reproducing key studies to confirm the reliability of prior research and to deepen the understanding of the methods involved. This reproducing process provided valuable insights, particularly in the use of BiLSTM in combination with GloVe embeddings, identifying areas for improvement.

3-5

An important focus of the experimental design is the integration of word embedding techniques, preprocessing, and audio features. Studies have shown that techniques such as Bag of Words (BoW), TFIDF, and Word2Vec are more effective than GloVe when analyzing text with rhythm and repetition, such as lyrics. This highlights the need to choose a task-specific embedding method. Furthermore, careful model tuning and evaluation, including global search and analysis of learning and loss curves, gives me deeper understanding of model behavior, resulting in improved accuracy and efficiency.

One of the most critical findings is the important role of audio features in improving model generalization and performance. Extensive testing on a variety of datasets highlights the importance of including audio elements in music sentiment analysis.

6

Practical applications of my model include analyzing the top 100 Spotify songs from over a decade, demonstrating the potential and generality of my model in real-world Settings. The application validates the relevance and applicability of the developed model in the analysis of contemporary music trends.

7

In summary, the design of the project at each stage resulted in key insights and ongoing optimization opportunities. The conducted experiments highlight the untapped potential of combining lyrics with audio features to advance the field of music sentiment analysis.

**Next Steps: (PPT21）**

Conduct a detailed annual emotional analysis of Spotify's Top 100 songs.

Complete a comprehensive report on the findings, contributing to the field of Music Emotion Recognition.