Project Proposal

Music Emotion Recognition: An Integrative Study of Lyrics and Audio Features

Student Name: Yuchen Zhu

Student ID: 2335100

Supervisor Name: Jizheng Wan

Project Category/Topic: AI

Project Aim:

- Develop an ensemble model that integrates lyric text and audio features to better predict the sentiment of a song. We anticipate that this ensemble model will outperform models that rely solely on a single feature in predicting song sentiment, revealing the interplay between lyrics and audio, and providing new insights into the way we understand music's psychological impact.

- **Significance**: Music has a powerful and transformative impact on human emotions. The motivation for this project stems from the growing emphasis on emotional well-being and the important role of music in influencing the emotional and psychological states of individuals. The project aims to provide users with a deeper and more nuanced understanding of the emotions conveyed by the music by incorporating the lyric content and audio features of the song in the analysis. By demonstrating that the ensemble model outperforms single models, this project aims to help users better understand and cope with the emotions they experience through music, thereby improving emotional well-being and potentially mitigating mental health issues such as anxiety and depression. In addition, the initiative is committed to driving research and innovation in the field of sentiment analysis.

- **Relevance to AI**: The project plans to use machine learning methods to build separate models based on text and audio features of lyrics, and combine the two models through ensemble learning techniques to more accurately predict the sentiment of songs.

Literature Review:

In recent years, Music Emotion Recognition (MER) [1] has gradually become a research hotspot and has emerged as one of the key research fields in the field of Music Information Retrieval (MIR) [2]. MER is committed to deep insight and accurate identification of the emotions and emotions expressed in music works, to realize personalized recommendations and accurate classification of music. Under the guidance of Russell's emotion model [3], which offers a more nuanced classification of emotions, the field has received sustained and extensive attention and research. The deep exploration of music emotion recognition not only promotes the rapid development of personalized music systems but also shows great social value and commercial application potential.

From a technical point of view, research on music emotion recognition (MER) has focused on using features of lyrics and audio to predict the emotional properties of music. In terms of lyrics analysis, a study [4] proposed to improve the accuracy of sentiment classification by using the style (StyBF), structure (StruBF), and semantic (SemBF) features of lyrics. Another study "MoodyLyrics" [5] proposed a lyric sentiment annotation method based on the emotional specificity of content words. Meanwhile, "LyBERT" [6] uses transfer learning and the BERT model to perform detailed multi-category sentiment classification of lyrics. In addition, a study [7] used naive Bayes to classify the positivity of music and found that the lyric-based classification showed high accuracy in positivity. These studies demonstrate the significant value of lyrics in the research of musical emotions. On the other hand, in the realm of audio features, research [8] used MFCC and residual phase features and adopted a Support Vector Machine (SVM) to classify music emotion. Another study [7] used SVM to classify the arousal of music and found that audio classification also achieved high accuracy in terms of arousal. In addition, there is research [9] that uses vector distance calculation, combines Spotify's valence and energy feature values, and refers to Russell's emotion model to define the emotion classification of music more accurately. Separately, another study [10] evaluated the performance of logistic regression in predicting song emotions using Spotify's audio features, concluding with high accuracy. These studies further confirm the significant value of audio features in emotion research.

Although the current single feature has shown some effectiveness in MER, its limitations gradually emerge with the deepening of research. Therefore, most researchers believe that integrating multiple features is an important way to improve the accuracy of music emotion recognition. For instance, studies in [11] and [7] suggest that combining audio and lyric features can enhance emotion classification, highlighting the value of integrating features from both dimensions. In addition, ensemble learning, as an effective method, has also been proven to significantly improve classification accuracy. As emphasized in [12], ensemble learning plays a significant role in improving the accuracy of sentiment classification, while [13] explores the application value of feature-level and decision-level fusion methods in sentiment analysis. Based on these analyses, this project aims to integrate lyrics features, audio features, and ensemble learning methods effectively to develop a more comprehensive and precise music emotion recognition model.

Project Objectives/Deliverables

Objective 1: To Curate a Comprehensive Dataset

- To meticulously gather and integrate diverse English-language music tracks, creating a dataset enriched with audio features, lyrics text, and sentiment labels.
-

Objective 2: To Enhance Data Preprocessing

- To design and implement efficient preprocessing methodologies, ensuring data quality and consistency for effective training and evaluation.

Objective 3: To Train Models and Develop an Ensemble Approach

- To first identify and train the most optimal individual sentiment analysis models focusing on audio features and lyrics text respectively, and then utilize an ensemble approach that strategically combines the strengths of the previously trained optimal models for improved sentiment prediction.

Objective 4: To Exceed Established Benchmarks in Model Evaluation

- To assess the developed models with the aim to surpass publicly acknowledged benchmarks in music sentiment prediction using the same publicly available test sets.

Deliverable 1. Comprehensive Dataset Collection and Integration

- Develop and curate a comprehensive, accurate, and reliable dataset containing diverse English-language music tracks, enriched with audio features from Spotify, corresponding lyric information, and sentiment labels.

Deliverable 2. Data preprocessing and partitioning:

- Implement efficient preprocessing methodologies to standardize and clean data, ensuring consistency and accuracy for effective training and evaluation.
- Strategically partition the data into training, validation, and test sets while ensuring a balanced distribution of sentiment labels across subsets.

Deliverable 3. Model Identification, Construction, and Training

- Identify and train the most optimal sentiment analysis models focusing individually on audio features and lyrics text.
- Develop an ensemble approach that strategically combines the strengths of the previously trained optimal models, aiming for improved sentiment prediction.

Deliverable 4：Evaluation and Benchmark Surpassing

- Rigorously assess the developed models with the aim to surpass publicly acknowledged benchmarks in music sentiment prediction using consistent and publicly available test sets.
- Compare the performance of the models against baseline models in existing literature using the same evaluation metrics to verify the effectiveness of our approach.

The project aims to create a powerful framework for music sentiment analysis by integrating various aspects of a music repertoire, including audio features and lyrics. Objective 1 focuses on curating a comprehensive dataset enriched with essential elements such as Spotify audio features, lyric text, and sentiment labels. This dataset lays a solid foundation for subsequent phases. Objective 2 highlights the importance of meticulous data preprocessing, ensuring that the data is standardized, cleaned, and strategically partitioned for efficient model training and evaluation. Objective 3 highlights the need to first identify and train optimal sentiment analysis models using audio features and lyrics text separately, and then develop an ensemble method that collaboratively combines these models to enhance sentiment prediction. Finally, Goal 4 aims to use a publicly available test set for music sentiment prediction that not only meets but also exceeds established benchmarks.

Overall, the deliverables ensure a thorough approach to music sentiment analysis, from data collection and preprocessing to model training and benchmark beyond evaluation.

Methodologies:

1. Data collection and integration:

Data Collection:

- A dataset consisting of audio features, lyric text, and sentiment tags was collected from Kaggle and other platforms. Spotify and Genius APIs are utilized to supplement missing audio features or lyrics information.

Data integration:

- Data sets from different sources are uniformly formatted and integrated to ensure the consistency and coherence of data in attributes, ranges and units.

2. Data preprocessing and partitioning:

Data cleaning:

- Improve the overall quality of the dataset by scrutinizing and removing outliers, duplicate entries, and non-English lyrics.

Text preprocessing:

- Word segmentation, stop word removal, and necessary stemming or lemmatization are performed on the lyrics text.

Audio feature Normalization:

- The audio features are normalized to ensure that the features are on the same dimension.

Data partition:

- The curated dataset is divided into training set, validation set and test set, and the distribution of sentiment labels in each subset is ensured to be balanced.

3. Model building, training, and ensemble:

Audio feature models:

- Test models based on time series data, such as one-dimensional Convolutional neural networks (1D CNNS), and select the best model based on performance for audio feature sentiment analysis.

Lyric text models:

- Test a range of algorithms, including basic support vector machines (SVMS), logistic regression, random forests, decision trees, and deep learning models such as LSTM, etc., and select the model with the best performance on lyric text sentiment analysis.

Ensemble methods:

- Ensemble learning strategies such as Stacking, Bagging, or Voting are utilized to integrate models of audio features and lyrics text. This strategy aims to allow the model to learn the complementary relationship between audio and lyrics in predicting sentiment.

Model tuning:

- Cross-validation is used for hyperparameter tuning, and the model structure is optimized by comparing the performance of different models on the validation set.

4. Model evaluation and comparison:

Performance Evaluation:

- In the testing phase, evaluation metrics such as precision, recall, and F1 score are used to comprehensively evaluate the prediction performance of the model.

Model comparison:

- The developed model is compared with the benchmark model in the literature on

the same test set to ensure that the evaluation metrics used are consistent with the publicly recognized benchmark.

Beat the baseline:

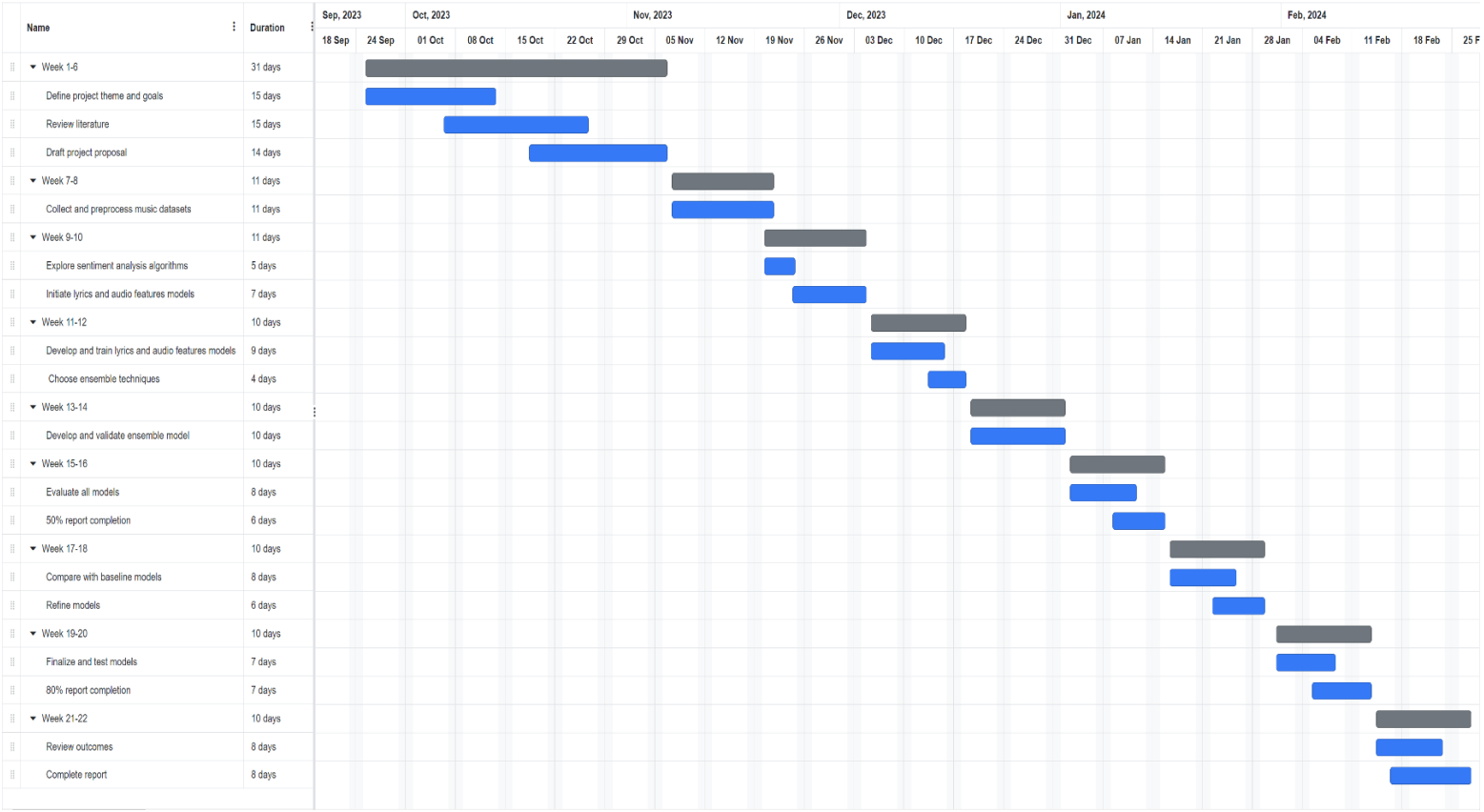
- Analyze and pursue the possibility that the model can outperform the existing baseline model on various evaluation metrics.

Project plan:

Week	**Dates**	**Tasks**
Week 1- 6	Sep 25 - Nov 5	- Define project theme, goals, and requirements. - Review related literature and datasets. - Draft and submit project proposal.
Week 7-8	Nov 6 - Nov 19	- Collect music datasets from sources such as Kaggle, Spotify API, Genius API, and other potential APIs. - Preprocess data for consistency and reliability.
Week 9-10	Nov 20 - Dec 3	- Explore and select algorithms for sentiment analysis. - Begin development of models for lyrics text and audio features. - Initiate report writing.
Week 11-12	Dec 4 - Dec 17	- Complete development of lyrics and audio features models. - Train models on dataset. - Select ensemble techniques for model integration.
Week 13-14	Dec 18 - Dec 31	- Develop ensemble model by combining lyrics and audio features. - Validate model performance against individual models.
Week 15-16	Jan 1 - Jan 14	- Evaluate all models using metrics such as precision, recall, F1 score, cross-validation, ROC curves, and confusion matrices. - 50% completion of the report.

Week 17-18	Jan 15 - Jan 28	- Compare models' results with existing baseline models. - Refine and optimize models based on feedback.
Week 19-20	Jan 29 - Feb 11	- 80% completion of the report. - Analyze results, make final adjustments, and conduct final tests on all models.
Week 21-22	Feb 12 - Feb 25	- Review project outcomes - Completion of the report

Gantt chart:



Risks:

- Finding sufficient, high-quality, and accurately annotated datasets of music and lyrics can be difficult.
- Copyright issues may be involved when processing and analyzing music and lyrics data.
- Even when ensemble learning and feature fusion are used, the performance of the algorithm may not be as good as expected.

Contingency Plan:

- If complete data is not available from the database, I would explore alternative public data sources or consider using public music APIs such as Spotify, Genius, Last.fm, etc., to supplement or replace missing values.
- If ensemble learning does not achieve the desired performance, we re-examine data quality, model diversity, and feature engineering, and consider data provenance or annotation issues as potential performance bottlenecks.

Hardware/Software Resources

1. Hardware Requirement:
   CPU – 11th Gen Intel Core i7-11800H @ 2.30GHz
   GPU – NVIDIA GeForce RTX 3060 Laptop GPU
   RAM – 16GB
   Storage – 100GB

2. Software Requirement:
   OS – Windows
   Python 3.11
   VS Code/PyCharm/Jupyter Notebook

3. Student and Supervisor have access to these resources.

Data

- The primary datasets utilized in this research were sourced from Kaggle and other public datasets mentioned in the relevant academic literature within the domain.
- In the event of missing values or additional data requirements, we might consider leveraging music-related public APIs such as Spotify API, Genius API, and Last.fm API.
- Student and Supervisor have the access to this data.

References

1. Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., ... & Turnbull, D. (2010, August). Music emotion recognition: A state of the art review. In *Proc. ismir* (Vol. 86, pp. 937-952).
2. Downie, J. S. (2003). Music information retrieval. *Annual review of information science and technology*, *37*(1), 295-340.
3. Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and*

social psychology, *39*(6), 1161.

4. Malheiro, R., Panda, R., Gomes, P., & Paiva, R. P. (2016). Emotionally-relevant features for classification and regression of music lyrics. *IEEE Transactions on Affective Computing*, *9*(2), 240-254.
5. Çano, E., & Morisio, M. (2017, March). Moodylyrics: A sentiment annotated lyrics dataset. In *Proceedings of the 2017 international conference on intelligent systems, metaheuristics & swarm intelligence* (pp. 118-124).
6. Rajendran, R. V., Pillai, A. S., & Daneshfar, F. (2022). LyBERT: Multi-class classification of lyrics using Bidirectional Encoder Representations from Transformers (BERT).
7. Tan, K. R., Villarino, M. L., & Maderazo, C. (2019, February). Automatic music mood recognition using Russell's twodimensional valence-arousal space from audio and lyrical data as classified using SVM and Naïve Bayes. In *IOP Conference Series: Materials Science and Engineering* (Vol. 482, No. 1, p. 012019). IOP Publishing.
8. Nalini, N. J., & Palanivel, S. (2016). Music emotion recognition: The combined evidence of MFCC and residual phase. *Egyptian Informatics Journal*, *17*(1), 1-10.
9. MODEL, R. S. C., & TO, V. D. C. International Journal of Modern Pharmaceutical Research. *Psychology*, *11*, 14.Jamdar, A., Abraham, J., Khanna, K., & Dubey, R. (2015). Emotion analysis of songs based on lyrical and audio features. *arXiv preprint arXiv:1506.05012*.
10. Dalida, M. R., Aquino, L. B., Hod, W. C., Agapor, R. A., Huyo-a, S. L., & Sampedro, G. A. (2022, December). Music Mood Prediction Based on Spotify's Audio Features Using Logistic Regression. In *2022 IEEE 14th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)* (pp. 1-5). IEEE.
11. Jamdar, A., Abraham, J., Khanna, K., & Dubey, R. (2015). Emotion analysis of songs based on lyrical and audio features. *arXiv preprint arXiv:1506.05012*.
12. Fersini, E., Messina, E., & Pozzi, F. A. (2014). Sentiment analysis: Bayesian ensemble learning. *Decision support systems*, *68*, 26-38.
13. Poria, S., Cambria, E., Howard, N., Huang, G. B., & Hussain, A. (2016). Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing*, *174*, 50-59.