

Project Proposal

Music Sentiment Analysis: Classification Using Integrated Lyrics and Audio Features

Student Name: Yuchen Zhu

Student ID: 2335100

Supervisor Name: Jizheng Wan

Project Category/Topic: AI

Project Aim:

- To develop an ensemble model integrating lyric text and audio features for enhanced sentiment prediction in songs. This ensemble approach is expected to surpass single-feature-based models, providing a richer understanding of the interplay between lyrics and audio in music's emotional 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 predict the sentiment of songs more accurately.

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] (the two-dimensional circumplex model of valence and arousal), 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. Building upon this methodology, 'LyBERT' [5] integrates these features through transfer learning and the BERT model to conduct a multi-category sentiment classification of lyrics based on Russell's emotion model. In addition, a study [6] found that the classification of music's valence achieved high accuracy when lyrical features were trained using Naïve Bayes. 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 [7] used MFCC and residual phase features and adopted a Support Vector Machine (SVM) to classify music emotion, and the previously mentioned study [6] also employed SVM but focused on classifying music's arousal, demonstrating high accuracy in this specific emotional dimension. In addition, there is research [8] 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. Furthermore, another study [9] also utilized Spotify's audio features, but it evaluated the performance of logistic regression in predicting song emotions, which resulted in 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 [10] and [6] 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 [11], ensemble learning plays a significant role in improving the accuracy of sentiment classification, while [12] explores the application value of feature-level and decision-level fusion methods in sentiment analysis. Based on these studies, the aim of this project is to effectively combine lyric and audio feature models, developing a more comprehensive and accurate ensemble model for music emotion recognition.

Project Objectives/Deliverables:

Objective 1: To Curate a Comprehensive Dataset

- To gather and integrate diverse English-language music tracks, creating a dataset 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 identify, train, and optimize sentiment analysis models for audio features and lyrics text, and then develop an ensemble approach for improved sentiment prediction.

Objective 4: To Exceed Established Benchmarks in Model Evaluation

- To evaluate the developed models and aim to surpass established benchmarks in music sentiment prediction, utilizing 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, 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 combines the strengths of the previously trained optimal models, aiming for improved sentiment prediction.

Deliverable 4: Evaluation and Benchmark Surpassing

- Conduct a rigorous assessment of the developed models, using consistent

evaluation metrics and publicly available test sets for transparent comparison, and demonstrate that these models surpass baseline models in music sentiment prediction.

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, Objective 4 is to evaluate our music sentiment model against a public test set and exceed public standard benchmarks.

Overall, the deliverables ensure a thorough and systematic approach to music sentiment analysis, meticulously addressing each stage from data collection and preprocessing to model training, tuning, and evaluation, thereby aiming to enhance the reliability and accuracy of the results.

Methodologies:

1. Data collection and integration:

Data Collection:

- Collect datasets with audio features, lyric text, and sentiment tags from various public platforms, such as Kaggle. Use Spotify, Genius, and Last.fm APIs to fill in missing audio features or lyric details.

Data integration:

- Standardize and merge datasets from multiple sources to maintain attribute, range, and unit uniformity for data consistency, using tools such as pandas.

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.

Lyric text preprocessing and feature extraction:

- Conduct basic text preprocessing steps such as word segmentation and elimination of stop words.
- Utilize text feature extraction techniques such as BoW and TF-IDF to transform the preprocessed lyrics into numerical features suitable for model training.

Audio feature preprocessing:

- Utilize pre-extracted audio features from databases, such as Spotify, which include attributes like acousticness, danceability, energy, tempo, valence, etc.
- Normalize these audio features to a consistent scale to facilitate comparable analysis across different dimensions.

Data partition:

- Divide the processed dataset into training, validation, and test sets, ensuring a balanced distribution of sentiment labels across each set to prevent model bias.

3. Model building, training, and ensemble:

Audio feature models:

- Test a range of algorithms such as CNN, RNN, LSTM, SVMs, Random Forests, etc., and select the model that performs best in sentiment analysis based on audio features.

Lyric text models:

- Test a range of algorithms such as Naive Bayes Classifiers, SVMs, Logistic Regression with BoW features, Random Forests with TF-IDF features, LSTM, etc., and select the model with the best performance on lyric text sentiment analysis.

Ensemble methods:

- Utilize ensemble learning strategies like Stacking, Bagging, or Voting to merge models for audio features and lyric text, aiming to harness the complementary relationship between them for sentiment prediction.

Model tuning:

- Tune and optimize the model using cross-validation, basing adjustments on validation set performance.

4. Model evaluation and comparison:

Performance Evaluation:

- Evaluation metrics such as precision, recall, and F1 score are employed to assess the predictive capabilities of the model using our own test set.

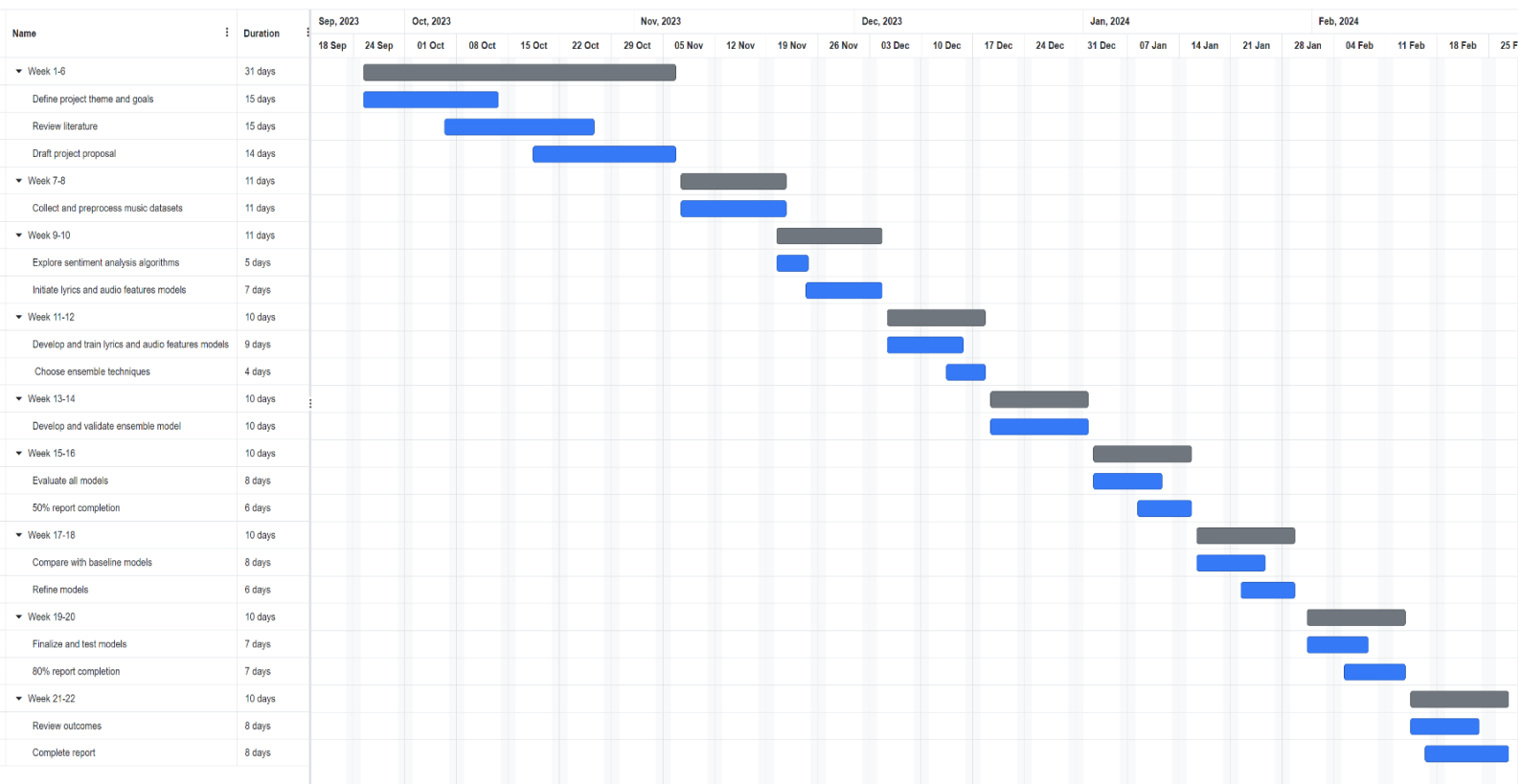
Model Comparison and Optimization:

- Compare the developed model against the benchmark models documented in the literature using the same public test set, aiming to not only align with widely acknowledged standards but also to exceed the established baseline across various evaluation metrics.

Project plan:

Week	Dates	Tasks
Week 1- 6	Sep 25 - Nov 5	<ul style="list-style-type: none">- Define project theme, aims, and requirements.- Review related literature and datasets.- Draft and submit project proposal.
Week 7-8	Nov 6 - Nov 19	<ul style="list-style-type: none">- 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	<ul style="list-style-type: none">- 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	<ul style="list-style-type: none">- Complete development of lyrics and audio features models.- Train models on the dataset.- Select ensemble techniques for model integration.
Week 13-14	Dec 18 - Dec 31	<ul style="list-style-type: none">- Develop an ensemble model by combining lyrics and audio features.- Compare and Validate ensemble model performance against individual models.
Week 15-16	Jan 1 - Jan 14	<ul style="list-style-type: none">- 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	<ul style="list-style-type: none">- Compare model results with existing baseline models.- Refine and optimize models based on feedback.
Week 19-20	Jan 29 - Feb 11	<ul style="list-style-type: none">- 80% completion of the report.- Analyze results, make final adjustments, and conduct final tests on all models.
Week 21-22	Feb 12 - Feb 25	<ul style="list-style-type: none">- 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 a complete dataset is unavailable, exploration of public data sources within the sector will be undertaken, and the use of public music APIs such as Spotify, Genius, Last.fm, etc., will be considered to supplement or replace missing values.
- If the model underperforms or if ensemble learning fails to meet performance expectations, a comprehensive analysis of data quality, model diversity, feature engineering, as well as data provenance and annotation practices will be conducted to identify and address potential bottlenecks. Insights and outcomes from this analysis will be documented to inform future research and model development efforts.

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
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. Rajendran, R. V., Pillai, A. S., & Daneshfar, F. (2022). LyBERT: Multi-class classification of lyrics using Bidirectional Encoder Representations from Transformers (BERT).
6. 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.

7. Nalini, N. J., & Palanivel, S. (2016). Music emotion recognition: The combined evidence of MFCC and residual phase. *Egyptian Informatics Journal*, 17(1), 1-10.
8. MODEL, R. S. C., & TO, V. D. C. International Journal of Modern Pharmaceutical Research. *Psychology*, 11, 14.
9. 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.
10. Jamdar, A., Abraham, J., Khanna, K., & Dubey, R. (2015). Emotion analysis of songs based on lyrical and audio features. *arXiv preprint arXiv:1506.05012*.
11. Fersini, E., Messina, E., & Pozzi, F. A. (2014). Sentiment analysis: Bayesian ensemble learning. *Decision support systems*, 68, 26-38.
12. 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.