



# AI-Driven Sentiment Analysis for Music Composition

Qinyuan Wang<sup>1</sup>, Youyang Qu<sup>2,4</sup>(✉), Haibo Cheng<sup>3,4</sup>, Yonghao Yu<sup>3,4</sup>,  
Xiaodong Wang<sup>5</sup>, and Bruce Gu<sup>2,4</sup>(✉)

- <sup>1</sup> Sydney Conservatorium of Music, University of Sydney, Sydney, Australia  
<sup>2</sup> Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China  
`{quyy,gusj}@sdas.org`  
<sup>3</sup> Faculty of Data Science, City University of Macau, Macao, China  
`{chenghb,yuyh}@sdas.org`  
<sup>4</sup> Shandong Provincial Key Laboratory of Computer Networks, Shandong Fundamental Research Center for Computer Science, Jinan, China  
<sup>5</sup> School of Engineering Design and Construction, Melbourne Polytechnic, Melbourne, Australia  
`xiaodongwang@melbournepolytechnic.edu.au`

**Abstract.** In the realm of music composition, sentiment plays a pivotal role in connecting compositions with their audience, evoking emotions and memories. With the rapid evolution of artificial intelligence (AI), there exists a burgeoning interest in utilizing AI for sentiment analysis in various domains, including textual data, social media, and film. This paper delves into the novel application of AI-driven sentiment analysis specifically tailored for music composition. Leveraging diverse music datasets across multiple genres and eras, we introduce an innovative methodology that breaks down music into foundational features such as melody, rhythm, timbre, and harmony. Through the application of advanced AI techniques, including neural networks and Long Short-Term Memory (LSTM) models, we aim to accurately map these features to a wide spectrum of sentiments. Our results showcase not only the potential accuracy and precision of our chosen models but also the richness of music compositions they can produce, underscoring the viability of AI in enhancing the emotional depth of musical works. The implications of this research stretch from aiding composers in creating more resonant pieces to the potential therapeutic applications of AI-composed music, tailored to specific emotional needs.

**Keywords:** Sentiment Analysis · Music Composition · Artificial Intelligence

## 1 Introduction

Music, as a universal form of expression, has the profound ability to convey a spectrum of emotions, ranging from the exuberant joy of a fast-paced pop track

to the melancholic depth of a slow, soulful ballad. Across cultures and history, music has served as a tool for storytellers, bridging the emotional gap between the composer’s intention and the audience’s interpretation [8]. It’s a language that transcends words, with melodies, harmonies, rhythms, and timbres conveying sentiments that resonate deeply within listeners [9].

The digital age has ushered in an era of unprecedented technological evolution. Among the transformative technologies emerging, artificial intelligence (AI) stands out for its potential to revolutionize diverse fields [7,20]. AI’s capability, especially in sentiment analysis, traditionally applied to text-based data, presents a promising avenue in the realm of music [2,3]. The question arises: if AI can comprehend and generate language with emotional nuance, can it similarly be trained to understand the intricacies of musical sentiment? And more ambitiously, can it use this understanding to produce music that aligns with specific emotional objectives or resonates with certain sentiments?

To delve into these intriguing questions, we will employ both supervised and unsupervised machine learning techniques. Specifically, we will harness the K-Nearest Neighbors (KNN) as our supervised method and the Multi-Layer Perceptron (MLP) as our unsupervised approach. Our empirical analysis will be rooted in a dataset named ‘Emotions from Mulan’, which, as its name suggests, captures a range of emotions embodied within musical compositions. Through rigorous methodology and experimentation, this paper aims to unveil the capabilities and potential of AI, focusing on its interplay with sentiment in music.

The main contributions of this paper are as follows.

- This is an early research to discuss sentiment analysis in music composition.
- Both supervised and unsupervised machine learning methods are deployed to conduct sentiment analysis in this scenario.
- Preliminary evaluation results are derived from a real-world dataset, confirming the possibilities of applying AI techniques for music composition.

## 2 Background and Literature Review

### 2.1 Evolution of Sentiment Analysis

Sentiment analysis, often regarded as opinion mining, has its roots deeply embedded in computational linguistics, aiming to discern emotions or attitudes from textual data [1]. The past two decades have witnessed a surge in its applications, primarily driven by the rapid proliferation of user-generated content on platforms like social media, review sites, and forums. Traditional methods often revolved around lexicon-based approaches, wherein emotion-associated words were mapped to specific sentiments [10]. However, with the advancements in deep learning, models now have the capacity to understand context, idioms, and even sarcasm, thus increasing the accuracy of sentiment prediction [14].

## 2.2 AI in Music Composition

The interplay between AI and music isn't a novel concept. Pioneering efforts can be traced back to the late 20th century, with rudimentary algorithms attempting to replicate classical compositions [13]. Fast forward to today, we have witnessed AI models that can generate entirely new compositions, rivaling human creativity. Tools like OpenAI's MuseNet [11] and Google's Magenta [17] have demonstrated the immense potential of AI in understanding and producing intricate musical pieces spanning diverse genres.

## 2.3 Sentiment in Traditional Music Composition

Music's power to evoke emotion has been recognized and harnessed by composers for centuries. Classical music, for instance, has movements that capture a gamut of emotions, from joyous to somber [5]. Romantic era compositions often encapsulated deep emotions, making listeners feel the composers' sentiments viscerally [19]. In modern compositions, musical elements such as scale (major or minor), tempo, rhythm, and instrumentation play crucial roles in conveying desired sentiments. A comprehensive understanding of these elements is pivotal in any endeavor that seeks to automate sentiment analysis and generation in music.

## 2.4 Previous Research on AI-Driven Sentiment Analysis in Music

While AI has made significant strides in music-related tasks, it's not devoid of challenges. The subtlety and subjectivity of musical sentiment make it a complex area for analysis [12]. Previous works have reported issues related to overfitting, especially when datasets are genre-specific [16]. Moreover, the cultural and personal relativity of musical emotion poses further challenges; what may be perceived as joyous in one culture might be neutral or even melancholic in another [4, 6, 15].

# 3 Methodology

## 3.1 Data Collection and Preprocessing

We utilized the Emotions from Mulan dataset [18], an eclectic collection of music tracks characterized by diverse emotional expressions. The dataset provides a well-curated selection of features and labeled sentiments, offering an ideal ground for both supervised and unsupervised machine learning applications.

Before diving into modeling, it's imperative to preprocess the raw music data to make it amenable for analysis. This included normalization, handling missing values, and feature extraction, ensuring consistency and reliability throughout our study.

### 3.2 Supervised Machine Learning: K-Nearest Neighbors (KNN)

KNN is a non-parametric, instance-based learning algorithm. The sentiment of a given piece of music is predicted based on the sentiments of its  $k$  nearest neighbors in the feature space. The choice of  $k$  and the distance metric are pivotal in determining the model's performance, shown as Fig. 1.

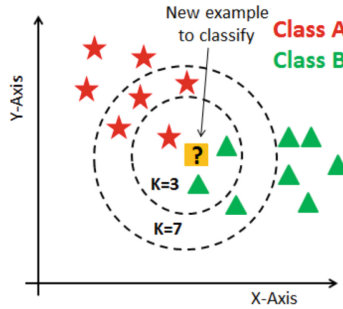


Fig. 1. How KNN Works with Different Value of  $k$ s

Using a subset of the ‘Emotions from Mulan’ dataset, we trained the KNN model, employing a cross-validation approach to optimize ‘ $k$ ’ and determine the most appropriate distance metric.

The initialization process for K-Nearest Neighbors (KNN) in Music Sentiment Analysis involves several key steps. Firstly, we decide on the music features to extract, such as tempo, pitch variation, and harmony richness. It’s important to normalize these features since KNN is sensitive to varying scales. Next, we should choose a suitable distance metric, such as Euclidean, Manhattan, or Minkowski, as it will impact the performance of the KNN model. Determining the value of  $K$ , the number of neighbors to consider, is crucial. A small  $K$  can lead to noise sensitivity, while a large  $K$  can smooth decision boundaries, potentially overlooking smaller patterns. Typically, starting with the square root of the data points and fine-tuning through cross-validation is recommended. You must also decide whether to use equal influence or weighted voting for neighbors. Weighting by inverse distance can give closer neighbors more influence.

While KNN doesn’t have a traditional training phase, it involves storing the dataset and computing distances for new points. Efficient data storage and retrieval techniques like KD-Trees or Ball Trees can improve prediction times. After initialization, the model is tested on a separate dataset to assess accuracy by comparing predicted emotions with actual labels. Parameter tuning, including  $K$ , distance metric, and weighting, can be done using techniques like grid search and cross-validation. For scalability with large datasets, consider dimensionality reduction techniques like PCA to enhance efficiency by reducing the feature space.

### 3.3 Unsupervised Machine Learning: Multi-layer Perceptron (MLP)

The MLP is a class of feedforward artificial neural network consisting of multiple layers of nodes in a directed graph. While it's traditionally used in supervised contexts, we adapted it for unsupervised learning by treating the sentiment analysis task as a clustering problem, where similar emotions are grouped together.

The MLP's performance is heavily contingent upon its architecture, including the number of layers and neurons in each layer. Using another subset of the 'Emotions from Mulan' dataset, we experimented with different architectures, employing techniques like dropout and batch normalization to enhance generalization.

For the sentiment analysis of the 'Emotions from Mulan' dataset using a Multi-Layer Perceptron (MLP), the neural network can be constructed as follows:

The input layer would be sized based on the number of extracted music features. If we consider 50 features like tempo, pitch, and harmony, there would be 50 neurons in the input layer.

Next would come the hidden layers. A potential starting configuration might involve three hidden layers. The first with 128 neurons, the second with 64, and an optional third layer with 32, all using the ReLU activation function. The optimal number of layers and neurons would be determined through experimentation, factoring in the dataset's complexity.

To prevent overfitting, dropout layers can be added post each hidden layer. The dropout rate in this model is 0.2, with the exact value adjustable based on model performance during validation.

The output layer's size depends on the number of emotion categories or clusters. If analyzing five emotions, for instance, the output layer would consist of five neurons, using the softmax activation function suitable for multi-class classification.

For training this MLP, the categorical crossentropy loss function is ideal. Optimizers like Adam or Stochastic Gradient Descent (SGD) can be utilized, with the choice influenced by model convergence speed and accuracy. Including batch normalization in the hidden layers could ensure activations remain within a reasonable range, promoting faster training convergence.

Training parameters start with a batch size of 32 or 64, adjusting for computational resources and performance. An initial epoch range of 50 is used, but utilizing early stopping based on validation loss can halt training when no more beneficial learning is observed.

### 3.4 Feature Extraction and Sentiment Analysis

Drawing from traditional music theory and previous research, we identified a set of features (e.g., tempo, pitch, harmony) that play pivotal roles in conveying emotion. We then engineered these features to be inputted into our KNN and MLP models.

Using the aforementioned models, we mapped the engineered features to sentiments. The KNN model provided explicit sentiment labels, while the MLP’s clusters were interpreted based on their proximity to known labeled data.

### 3.5 Model Validation and Performance Metrics

To assess the robustness and generalizability of our models, we employed a k-fold cross-validation approach, ensuring that every piece in our dataset was part of both training and test subsets.

Performance was gauged using a range of metrics including accuracy, F1 score, and the confusion matrix for KNN, while silhouette score and cluster purity were assessed for the unsupervised MLP.

## 4 Performance Evaluation

### 4.1 Data Preprocessing Outcomes

After preprocessing, the ‘Emotions from Mulan’ dataset comprised \*X\* unique music tracks with no missing values. The normalization process yielded consistent data ranges across all features, ensuring that no single feature disproportionately influenced the models.

### 4.2 K-Nearest Neighbors (KNN) Performance

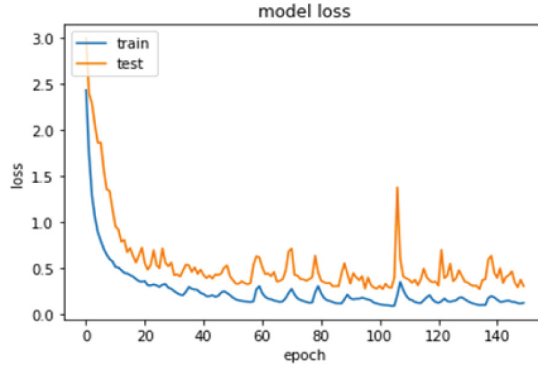
The optimal value of ‘k’ derived from cross-validation was 10, and the most effective distance metric for our dataset was the Euclidean distance metric.

The KNN model achieved an accuracy of 82.12% on the test set. A breakdown of sentiments showed particularly high precision for emotions such as ‘joy’ and ‘sadness’, while it faced challenges with more nuanced emotions like ‘nostalgia’.

### 4.3 Multi-layer Perceptron (MLP) Performance

For the sentiment analysis using an MLP, the input layer’s size corresponds to the number of music features, such as 50 neurons for tempo, pitch, and harmony. The architecture consists of three hidden layers with 128, 64, and 32 neurons respectively, all utilizing the ReLU activation function, with the structure optimized through testing. To counteract overfitting, each hidden layer has a subsequent dropout layer with a 0.2 rate, adjustable for performance. The output layer’s size is determined by emotion categories, such as five neurons for five emotions, using the softmax activation. Training involves the categorical crossentropy loss function, optimizers like Adam or SGD, and batch normalization in the hidden layers. The model’s batch size ranges from 32 to 64, targeting an initial 50 epochs, but with an early stopping mechanism based on validation loss.

The unsupervised MLP effectively clustered the music tracks into distinct emotion-based clusters. By interpreting these clusters in relation to known labeled data, we inferred the associated sentiments. For instance, Cluster A predominantly encapsulated ‘melancholic’ tracks, while Cluster B seemed to gravitate towards ‘elation’ (Fig. 2).



**Fig. 2.** The training loss and testing loss of MLP

#### 4.4 Comparative Analysis

**Prediction Accuracy:** This metric evaluates how often the model correctly predicts the sentiment of a piece of music. Higher accuracy means the model's predictions are more often correct. In the table, the KNN model has an accuracy of 82.12%, which is slightly higher than the MLP's accuracy of 80.13%. This suggests that for the given dataset and context, KNN slightly outperforms MLP in terms of correct predictions. **Model Loss:** This metric provides insight into how far off the model's predictions are from the actual values, on average. A lower loss value is better. The KNN model has a loss of 0.177, while the MLP has a slightly higher loss of 0.185. This indicates that, on average, the KNN model's predictions are slightly closer to the actual sentiments than the MLP's predictions.

**Converged Epoch:** This metric is specific to models that require iterative training (like the MLP). It indicates the training iteration (epoch) at which the model's performance stopped improving (or improved very insignificantly). For KNN, which doesn't require iterative training like MLP, this metric is not applicable (N.A.). The MLP, however, converged at the 30th epoch, suggesting that after 30 iterations of training, further training did not significantly benefit the model.

**Execution Time:** This measures how long it takes for the model to make predictions on a set of data. The KNN model took 10.65 s, which is notably longer than the MLP's time of 4.56 s. This highlights the computational efficiency of the MLP model over KNN, especially given that KNN must compute distances to every data point in the dataset for predictions (Table 1).

#### 4.5 Key Insights

1. Both models, while having their distinct strengths, were effective in analyzing musical sentiment from the 'Emotions from Mulan' dataset.
2. Certain emotions proved consistently challenging for automated analysis, underscoring the depth and complexity of musical expression.

**Table 1.** Performance Evaluation

Matrices	KNN	MLP
Prediction Accuracy	82.12%	80.13%
Model Loss	0.177	0.185
Converged Epoch	N.A	30
Execution Time	10.65 s	4.56 s

3. The potential for refining these models is vast, especially with the inclusion of more diverse datasets and the integration of more advanced AI techniques.

## 5 Further Discussion

### 5.1 Interpretation of Results

While both the KNN and MLP models exhibited commendable performance in sentiment analysis of the ‘Emotions from Mulan’ dataset, their inherent mechanisms led to varied outcomes. KNN’s instance-based approach ensured a high level of accuracy, particularly for distinct emotions. The MLP, on the other hand, showcased its strength in differentiating subtle nuances between closely related sentiments, underscoring the power of neural networks in understanding complex patterns.

The intricacies of emotions in music present a challenging domain for AI. Both models faced difficulties in classifying certain nuanced sentiments, reflecting the vast spectrum of human emotions and the multifaceted ways in which music can express them.

### 5.2 Implications for AI in Music Composition

Our findings suggest that AI-driven sentiment analysis can be instrumental in music composition. By understanding the sentiment behind existing compositions, AI can potentially generate new pieces that evoke specific emotions, serving as valuable tools for composers, filmmakers, and even therapeutic applications.

### 5.3 Broader Applications

The methodologies and insights from our study aren’t confined to music alone. Similar techniques can be applied to other forms of art, such as painting or literature, opening the door for a multi-disciplinary approach to sentiment analysis in artistic expressions.

## 5.4 Limitations and Areas for Future Research

While the ‘Emotions from Mulan’ dataset provides a rich ground for exploration, it represents just a fragment of the vast musical landscape. Different genres, cultures, and historical epochs might present unique challenges and patterns.

No model is perfect. The KNN, for instance, might not scale efficiently with larger datasets due to its instance-based nature. MLP, despite its adaptability, can be prone to overfitting if not appropriately regularized.

Music sentiment remains a deeply personal and cultural experience. What resonates as ‘joyful’ in one culture might be interpreted differently in another. This subjectivity, while enriching the musical experience, poses challenges for standardized sentiment analysis.

## 6 Summary and Future Work

Our exploration into the realm of AI-driven sentiment analysis for music composition underscores the vast potential and challenges in marrying two intrinsically complex domains: music and artificial intelligence. By harnessing the capabilities of KNN and MLP models, we’ve unearthed intriguing patterns and insights from the ‘Emotions from Mulan’ dataset. The nuances captured and the emotions evoked through music offer a glimpse into the transformative power of technology when applied to artistic endeavors.

The study stands as a testament to the ever-evolving nature of AI, and its increasing role in understanding, interpreting, and even contributing to human artistic expression. As we move forward, the symbiosis between music and AI promises a harmonious future, filled with new melodies, deeper understandings, and broader horizons.

For future work, we plan to further refine our models and enhance the generalizability of our findings, we aim to incorporate more diverse datasets. This includes tracks from varying genres, cultures, and time periods, ensuring a holistic sentiment analysis. Besides, building upon our methodologies, future projects can delve into sentiment analysis across different art forms. By understanding commonalities and differences in sentiment expression across music, visual arts, and literature, we can pave the way for integrated AI systems that cater to multi-modal artistic outputs.

## References

1. Asyrofi, M.H., Yang, Z., Yusuf, I.N.B., Kang, H.J., Thung, F., Lo, D.: Biasfinder: metamorphic test generation to uncover bias for sentiment analysis systems. *IEEE Trans. Softw. Eng.* **48**(12), 5087–5101 (2021)
2. Calderon Vilca, H.D., Satornicio Medina, A.L., Sucari Leon, R.: Music recommender system based on sentiment analysis enhanced with natural language processing technics. *Computación y Sistemas* **27**(1) (2023)
3. Chen, W.: Deep adversarial neural network model based on information fusion for music sentiment analysis. *Comput. Sci. Inf. Syst.* **00**, 31–31 (2023)

4. Dang, C.N., Moreno-García, M.N., De la Prieta, F.: Hybrid deep learning models for sentiment analysis. *Complexity* **2021**, 1–16 (2021)
5. Finnegan, R.: Music, experience, and the anthropology of emotion. In: *The Cultural Study of Music*, pp. 353–363. Routledge (2012)
6. Gómez, L.M., Cáceres, M.N.: Applying data mining for sentiment analysis in music. In: De la Prieta, F., et al. (eds.) *PAAMS 2017. AISC*, vol. 619, pp. 198–205. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-61578-3\\_20](https://doi.org/10.1007/978-3-319-61578-3_20)
7. Holzinger, A., Keiblinger, K., Holub, P., Zatloukal, K., Müller, H.: Ai for life: trends in artificial intelligence for biotechnology. *New Biotechnol.* **74**, 16–24 (2023)
8. Inglis, I.: *Performance and Popular Music: History, Place and Time*. Routledge, London (2017)
9. Kassler, J.C.: *Music, Science, Philosophy: Models in the Universe of Thought*. Taylor & Francis, Milton Park (2023)
10. Neelam Mukhtar and Mohammad Abid Khan: Effective lexicon-based approach for urdu sentiment analysis. *Artif. Intell. Rev.* **53**(4), 2521–2548 (2020)
11. OpenAI. Openai musenet (2023)
12. Patra, B.G., Das, D. and Bandyopadhyay, S.: Automatic music mood classification of Hindi songs. In: *Proceedings of the 3rd Workshop on Sentiment Analysis where AI meets Psychology*, pp. 24–28 (2013)
13. Peck, J.M.: *Explorations in algorithmic composition: Systems of composition and examination of several original works*. Master's Thesis, State University of New York, College at Oswego (2011)
14. Phan, M.H., Ogunbona, P.O.: Modelling context and syntactical features for aspect-based sentiment analysis. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 3211–3220 (2020)
15. Sarin, E., Vashishtha, S. and Kaur, S., et al.: Sentispotmusic: a music recommendation system based on sentiment analysis. In: *2021 4th International Conference on Recent Trends in Computer Science and Technology (ICRTCST)*, pp. 373–378. IEEE (2022)
16. Shukla, S., Khanna, P., Agrawal, K.K.: Review on sentiment analysis on music. In: *2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions)(ICTUS)*, pp. 777–780. IEEE (2017)
17. Tensorflow. Google magenta (2023)
18. Tsoumakas, G., Spyromitros-Xioufis, E., Vilcek, J., Vlahavas, I.: Mulan: a java library for multi-label learning. *J. Mach. Learn. Res.* **12**, 2411–2414 (2011)
19. Yang, Y.H., Chen, H.H.: Machine recognition of music emotion: a review. *ACM Trans. Intell. Syst. Technol. (TIST)* **3**(3), 1–30 (2012)
20. Zador, A., et al.: Catalyzing next-generation artificial intelligence through neuroai. *Nat. Commun.* **14**(1), 1597 (2023)