

Mood-Based Song Recommendation System Using TF-IDF and Cosine Similarity

Andri Riyan Dwi Putra Siregar, Rizaldy Khair

Muhammadiyah University of North Sumatra Medan, Medan North Sumatra

ABSTRACT

This study created a mood-based music recommendation system using a content-based filtering approach with TF-IDF and Cosine Similarity to handle the huge amount of digital music data. Using about 15,000 songs from Kaggle, the system processes metadata and audio features like valence, energy, danceability, and tempo, all implemented in a Python and Flask web app. The mix of text similarity and audio features works really well for giving spot-on recommendations, as shown by performance scores of 0.97 accuracy, 0.90 precision, 0.82 recall, and 0.86 F1-score. Plus, with a System Usability Scale (SUS) rating in the good range, the system is not just technically accurate but also super easy to use and well-liked by users.

Keywords: Recommendation System, Mood, TF-IDF, Cosine Similarity, Content-Based Filtering

Corresponding author

Name: Andri Riyan Dwi Putra Siregar

Email: andrisiregarrr1212@gmail.com

INTRODUCTION

Music has become an important part of human life, not just as entertainment but also as a way to express emotions. Many studies mention that music can affect a person's mood, from boosting happiness, calming the mind, to motivating when feeling tired. With the development of digital technology, there are now millions of song collections that can be accessed online through various platforms like Spotify, Apple Music, and Joox. The huge availability of songs, however, actually creates a new problem: users find it difficult to find songs that match their moods. Recommendation systems have been widely applied in various fields, such as movies, e-commerce, and job matching according to user preferences or needs. Some commonly used approaches include collaborative filtering and content-based filtering.

In this research, the chosen approach is content-based filtering, where the recommendation process is based on the content or attributes of the items themselves, such as genre, title, or song descriptions that represent a certain mood (Aziz, 2023). By using TF-IDF to analyze song data, the system can identify attributes from the songs that reflect specific emotional themes. TF-IDF gives weight to words that appear significantly in the attributes, allowing the system to recognize songs that match the desired mood. The main problem in this study is the difficulty users have in finding songs that fit their mood. To

address this, the proposed solution is to build a content-based song recommendation system. This system uses the features and information contained in songs to provide more personal and relevant recommendations.

Several previous studies have applied the Term Frequency–Inverse Document Frequency (TF-IDF) and Cosine Similarity methods in various search and content-based recommendation systems. This approach has been proven effective in classifying thesis documents with 98% accuracy (Wahyuni et al., 2017), online news recommendation systems with 76% precision (Noorhidayah et al., 2019), job vacancy recommendation systems with 80% accuracy (Hersianty et al., 2025), and even movie recommendation systems based on synopses (Azmi et al., 2024). However, most recommendation systems on music platforms today still mainly rely on collaborative filtering based on user interaction history. That approach has limitations like the cold-start problem for new users and is not very responsive to momentary emotional states. Also, most previous studies only focused on pure text weighting without fully integrating audio features to measure a song's emotional fit.

Based on that gap, the main problem this research aims to solve is how to design and implement a content-based filtering song recommendation system that can provide a list of songs relevant to the user's mood. Specifically, this study aims to implement the TF-IDF method to extract and weight words in song metadata, as well as to use the Cosine Similarity method to calculate the similarity between the user's mood preferences and the song data. In addition, this research also aims to evaluate the technical performance of the system using accuracy, precision, recall, and F1-score metrics, while measuring usability and user satisfaction through the System Usability Scale (SUS) method.

This research provides both scientific and practical contributions by developing a song recommendation system that combines text weighting (TF-IDF) from emotional metadata with Spotify audio feature extraction such as valence, energy, danceability, and tempo. Merging this textual information with musical characteristics can mathematically produce recommendations that aren't just homogeneous in terms of category, but also consistent in audio vibe and reduce the risk of overspecialization. The results of this study are expected to add to the academic literature in text mining and music recommendation systems, as well as offer insights for system developers to optimize personalization of digital entertainment content based on users' psychological states.

RESEARCH METHODOLOGY

The type of research used in this study is quantitative research with an experimental approach. An experimental design was chosen because it allows researchers to measure results objectively in numerical form and analyze them statistically to draw valid conclusions. Through this approach, the researchers designed, built, and tested a song recommendation system using the Term Frequency–Inverse Document Frequency (TF-IDF) weighting method and Cosine Similarity measurement within a content-based filtering framework. Besides evaluating the technical performance quantitatively, this study also

measures user satisfaction (usability) using the System Usability Scale (SUS) instrument, which is converted into measurable numerical scores.

Subjek/Populasi dan Sampel

1. Population and Research Objects: The main object of study is a digital song dataset that was openly obtained from the Kaggle platform, sourced from Spotify metadata. After going through data filtering and cleaning (preprocessing), the total population of valid songs used in the system is 14,871 songs.
2. Usability Test Respondents (Sample): For testing the system's usability, the subjects involved as respondents are general users who have an interest in music, with a sample size of 20 respondents from the productive age group.

DATA COLLECTION PROCEDURE

The data collection method in this study uses secondary data obtained online by downloading a song dataset from the Kaggle platform. The dataset includes text metadata information (such as song title, artist name, and genre) as well as numerical audio features calculated by Spotify (such as valence, energy, danceability, tempo, acousticness, loudness, speechiness, instrumentalness, and liveness). The data collection and preparation procedures include:

1. Downloading the raw dataset (ClassicHit.csv) from Kaggle.
2. Selecting and cleaning the data (data cleaning) to remove
3. Mapping audio feature values like valence and energy to specific mood categories.
4. Collecting qualitative and quantitative data by distributing the System Usability Scale (SUS) questionnaire, which consists of 10 statements on a 1 to 5 Likert scale, to 20 respondents after they tried the system.

DATA ANALYSIS

Data analysis in this study is divided into two main stages, namely system computational analysis and statistical evaluation analysis:

1. Text Analysis and Recommendation Computing (Text Mining & Similarity): Automatic Preprocessing: Using the TfidfVectorizer library for tokenizing (splitting words) and case folding (standardizing to lowercase) on the combined mood and genre text attributes. TF-IDF Weighting: Calculating the importance level of words using the formula that multiplies term frequency by the inverse document frequency to form a vector matrix sized $14,871 \times 26$
2. System Performance Evaluation Analysis (Quantitative Metrics): Measuring the technical performance of recommendation classification using a Confusion Matrix broken down into Accuracy, Precision, Recall, and F1-score metrics.
3. Usability Questionnaire (SUS) Analysis: Calculating the final score of the system usability questionnaire by adding up the contribution scores of each statement (converted based on odd/positive and even/negative numbers), then multiplying by 2.5 to get a score range of 0–100, and finally averaging it to determine the system acceptance category.

Table 1: Detailed Calculation of Similarity Scores for the Top 10 Recommended Songs
(Input: Mood 'Happy', Genre 'Metal')

Song Title	Artist	Score TF-IDF
Space Truckin'	Deep Purple	1,0000
I'll Wait	Van Halen	1,0000
All Night Long	Rainbow	1,0000
Christine Sixteen	KISS	1,0000
Heaven's On Fire	KISS	1,0000
The Beautiful People	Marilyn Manson	1,0000
Too Hot to Handle - 2008 Remaster	UFO	1,0000
We're Not Gonna Take It	Twisted Sister	1,0000
Piece of Me	Skid Row	1,0000
Sinful Love	Blue Öyster Cult	1,0000

Source: 10.000

Table 2 Hasil Evaluasi Kinerja Teknis Sistem

Evaluation Metrics	Calculation Formula	Result Value
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN} = \frac{9 + 88}{100}$	0,97 (97%)
Precision	$\frac{TP}{TP + FP} = \frac{9}{9 + 1}$	0,90 (90%)
Recal	$\frac{TP}{TP + FN} = \frac{9}{9 + 2}$	0,82 (82%)
F1-Score	$2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{0,90 \times 0,82}{0,90 + 0,82}$	0,86 (86%)

Sumber: Hasil Perhitungan Metrik Evaluasi (2026)

MAIN FINDING

Based on experiments and testing conducted on the mood-based song recommendation system using the Term Frequency–Inverse Document Frequency (TF-IDF) method and Cosine Similarity, the main findings are as follows: Results of Data Preprocessing and Weighting: From the initial data, a total of 14,871 song data were considered valid after going through cleaning, duplicate removal, and normalization stages. Text weighting using TF-IDF on the combined mood and genre attributes produced a matrix with dimensions of 14,871×26 unique word features. Results of Similarity Calculation (Cosine Similarity): Combining text-based similarity (TF-IDF) and audio feature similarity

(valence, energy, tempo, danceability) was able to produce similarity scores (hybrid similarity score) ranging from 0 to 1.

For example, in testing the input mood "Happy" and genre "Metal", the song "Space Truckin'" by Deep Purple got the highest similarity score of 0.9729 (TF-IDF: 1.0; Audio: 0.9458). System Technical Performance: The system performance evaluation using 100 test data based on the Confusion Matrix resulted in an Accuracy of 0.97, Precision of 0.90, Recall of 0.82, and an F1-Score of 0.86. User Usability Level: The system usability test involving 20 respondents using the System Usability Scale (SUS) instrument gave an average score of 78.4. This score places the system in the Acceptable category with a Good rating.

DISCUSSION

INTERPRETATION OF FINDINGS

Based on the analysis and data testing that has been done, the mood-based song recommendation system using the Term Frequency–Inverse Document Frequency (TF-IDF) and Cosine Similarity methods has proven to provide accurate and relevant results. The system's accuracy value reaching 0.97 (97%) shows that the computational model built successfully maps text characteristics (mood and genre categories) and audio features (valence, energy, danceability, tempo) with high precision.

CONCLUSION

A precision score of 0.90 and a recall score of 0.82 reflect that the system can minimize the appearance of inappropriate song recommendations (false positives), while also being effective in filtering relevant songs from a total of 14,871 song data. From the user experience side, a System Usability Scale (SUS) score of 78.4, which falls into the Acceptable and Good category, indicates that the web-based interface developed is very intuitive, easy to use without requiring technical training, and provides high comfort for users in exploring music based on their emotional states.

RELATED TO LITERATURE

The findings in this study strengthen and expand on previous literature regarding the application of the TF-IDF and Cosine Similarity methods beyond pure text domains. Compared to earlier studies by Wahyuni et al. (2017) on thesis document classification, which achieved 98% accuracy, Noorhidayah et al. (2019) on online news recommendations with 76% precision, and Hersianty et al. (2025) on job vacancy recommendations with 80% accuracy, applying a similar method in the music domain in this study managed to record a competitive accuracy performance of 97%.

The advantage of this research lies in combining a content-based text approach with quantitative audio features from Spotify, creating a hybrid approach. This successfully tackles fundamental weaknesses of conventional recommendation systems like collaborative filtering, which often struggle with the cold-start problem for new users, and also avoids the limitations of purely text-based methods that overlook the rhythm and musical energy dynamics of a song.

RESEARCH LIMITATIONS

Even though the system has shown good technical performance and usability, there are several limitations that can affect the results of this study, including: Dependence on Dataset Quality The system heavily relies on the completeness of the initial metadata and the accuracy of mapping audio feature values (valence and energy) into mood categories in the Kaggle dataset. If there are initial labeling errors in the dataset, this can affect the final similarity scores. Characteristics of Content-Based Filtering (Overspecialization) This approach tends to recommend items that have high similarity with the initial query, which limits the user's exploration of new music genres or variations outside their main preferences. Limitations of Mood Categories The mood options in the system are still rigid and limited to the categories pre-set by the developers, so users can't add more specific personal emotion descriptions.

IMPLICATION

Based on the results and limitations of the research, the implications of this study are divided into practical implications and directions for future development. Practical Implications: The recommendation system that has been deployed online through the PythonAnywhere platform and directly integrated with Spotify playback links can be used by the digital music industry or general users as a quick alternative for finding songs based on current psychological or emotional states.

Suggestions for Research

The next developer is advised to combine content-based filtering methods with a collaborative filtering approach (hybrid method) to increase the variety and diversification of song recommendations. Expand the scope of data analysis by involving in-depth song lyrics extraction or using more complex Natural Language Processing (NLP) techniques. Add customization features for users to input free-text descriptions of their emotions so the system can analyze moods more dynamically and flexibly.

CONCLUSION

This study successfully designed, built, implemented, and evaluated a song recommendation system based on mood using the Term Frequency–Inverse Document Frequency (TF-IDF) method and Cosine Similarity within a content-based filtering framework. Through data preprocessing steps that included cleaning, removing duplicates, and mapping audio features like valence and energy from the Spotify dataset on Kaggle, a total of 14,871 valid song data points were obtained. Text weighting using TF-IDF produced a vector representation with dimensions of $14,871 \times 26$ unique word features, which were then matched with user queries through Cosine Similarity calculations. The combination of text-based and audio feature approaches proved effective in producing an accurate hybrid similarity score, with the system consistently able to display the top 10 recommended songs most relevant to user preferences.

In terms of technical performance and usability, a quantitative evaluation using the Confusion Matrix on 100 test data showed very satisfying results, with an accuracy of 0.97 (97%), precision of 0.90 (90%), recall of 0.82 (82%), and an F1-score of 0.86 (86%). Additionally, a qualitative evaluation using the System Usability Scale (SUS) involving 20 respondents resulted in an average score of 78.4, indicating that the system falls into the Acceptable category with a Good rating. This proves that the web-based application, which has been deployed online through PythonAnywhere and integrated with Spotify playback links, is very intuitive, easy to use, and provides a positive user experience.

Even so, there are a few limitations, like the potential for overspecialization due to the nature of content-based filtering and the restricted choices of rigid mood categories. So, for future development, the system can be improved by applying a hybrid method that combines it with collaborative filtering to enrich the variety of recommendations. Also, further research is recommended to expand free-text input features using Natural Language Processing (NLP) so users can type in emotion descriptions more flexibly, and to enhance dataset complexity through in-depth analysis of song lyrics for broader and more dynamic recommendation coverage.

Suggestions for Implementation and Further Research

The findings, performance analysis, and some limitations encountered during the development of this song recommendation system lead to several important recommendations that can be divided into practical application aspects in the industry and future research directions. These suggestions are expected to optimize the system's functionality and expand the use of mood-based recommendation technology.

In terms of practical application, this web-based system, which has been successfully deployed online and integrated with Spotify playback links, has huge potential to be further developed on a commercial scale. Developers or digital music industry players can optimize the system integration by fully leveraging the Spotify Web Playback SDK. With this integration, users aren't just redirected through external links—they can directly play, pause, or skip recommended songs right within the main app interface, creating a much more seamless listening experience. Plus, since the current interface is responsive and accessible across devices like computers and smartphones, the system could be repackaged as a native mobile app to make it easier for users to access on the go.

Meanwhile, for future research development, there are several methodological and technical aspects that can be improved. First, to address the fundamental weakness of the content-based filtering approach, which tends to experience overspecialization—where the system only recommends songs very similar to the initial preferences, thus limiting the exploration of new songs—later studies are suggested to apply a hybrid method. Combining content-based filtering and collaborative filtering will help enrich the variety and diversity of recommendations based on the behavior of other users.

Second, the scope and flexibility of system inputs (query input) can be improved by using advanced Natural Language Processing (NLP) techniques or free-text-based sentiment analysis. Instead of limiting mood options to rigid standard categories, future

systems could be designed to read and analyze sentences describing emotions or daily complaints typed directly by users, translating them into more dynamic recommendation parameters. Finally, the complexity of the dataset can be expanded by including in-depth song lyric extraction to capture the emotional and poetic nuances of a piece of music more comprehensively beyond standard metadata and audio features.

REFERENCES

- Abidin, A. Z. Z., & Sukmadinata, A. (2020). Damage detection system in operating systems using TF-IDF and cosine similarity methods. *Journal of Informatics Science*, 8(2), 108–112. <https://doi.org/10.33884/jif.v8i2.2155>
- Apriani, A., Zakiyudin, H., & Marzuki, K. (2021). Application of cosine similarity algorithm and TF-IDF weighting in a new student admission system at a private campus. *Bumigora Information Technology Journal (BITe)*, 3(1), 19–27. <https://doi.org/10.30812/bite.v3i1.1110>
- Ardhana, V. Y. P. (2021). Usability Testing of the Halodoc Application Using the System Usability Scale (SUS) Method, *J. Kesehatan. Qamarul Huda*, vol. 9, no. 2, pp. 132–136.
- Aziz, M. A. (2023). Job recommendation system using content-based filtering algorithm (TF-IDF and cosine similarity)
- Azmi, A. H., Mahardika, N. I., Prasetya, R. S. A., & Sari, A. P. (2024). Content-based movie recommendation system using cosine similarity and TF-IDF techniques. *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research*, 1(4), 2135–2145. <https://doi.org/10.32672/mister.v1i4.2206>
- Binangkit, C. A., Voutama, A., & Heryana, N. (2023). The use of UML (Unified Modeling Language) in planning a web-based musical instrument rental management system. *JATI (Journal of Informatics Engineering Students)*, 7(2), 1429–1436. <https://doi.org/10.36040/jati.v7i2.6858>
- Hersianty, V. M., Amalia, E. L., Puspitasari, D., & Wibowo, D. W. (2025). The application of the TF-IDF algorithm and cosine similarity in job vacancy recommendation systems. *JATI (Journal for Informatics Engineering Students)*, 9(1), 1619–1625.
- Kesuma, H. W. A. (2016). The application of the TF-IDF method and cosine similarity in the Commercial Law Code book app.
- Melita, R. (2018). Application of the Term Frequency Inverse Document Frequency (TF-IDF) method and cosine similarity in an information retrieval system to understand hadith explanations based on the web (Case Study: Sahih Bukhari-Muslim Hadith)
- Noorhidayah, N., Indriani, F., & Faisal, M. R. (2019). Online news recommendation system using TF-IDF weighting and cosine similarity. *National Seminar on Computer Science (SOLITER)*, 2, 85–93.
- Rasyid, R. A., & Ningsih, D. H. U. (2024). Application of TF-IDF Algorithm and Cosine Similarity for Search Queries on a Tourism Destination Dataset. *JTIK Journal (Journal of Information and Communication Technology)*, 8(1), 170–178. <https://doi.org/10.35870/jtik.v8i1.1416>

- Sari, T. (2020). The Effect of Music on Anger and Stress in Elderly. International Journal of Scientific and Technological Research. <https://doi.org/10.7176/jstr/6-03-30>
- Siswidiyanto, S., Munif, A., Wijayanti, D., & Haryadi, E. (2020). Web-based house rental information system using the prototype method. Interkom Journal, 15(1), 18–25. <https://doi.org/10.35969/interkom.v15i1.64>
- Voutama, A., & Novalia, E. (2021). Designing an Android-based M-Magazine app as a medium for high school bulletin boards. TEKNO KOMPAK Journal, 15(1), 104–115. <https://doi.org/10.35969/interkom.v15i1.64>
- Wahyuni, R. T., Prastiyanto, D., & Suprpto, E. (2017). The application of the cosine similarity algorithm and TF-IDF weighting in a thesis document classification system. Journal of Electrical Engineering, 9(1), 18–23.