EdgeTech Insights Journal
Vol.00, 1 (month, year), pp. 1-xx
ISSN: xxxxx EISSN: xxxxx, DOI: xxxxx

CLASSIFICATION OF SONG MOOD BASED ON LYRIC ANALYSIS USING TF-IDF AND SUPPORT VECTOR MACHINE

Zulfan Januardi Adha, Muhamad Rafi Irhab
[1]Lancang Kuning University, Pekanbaru, Indonesia
[2]Lancang Kuning University, Pekanbaru, Indonesia
zulfanjanuardiadhaanimasi@gmail.com
muhamadrafiirhab@gmail.com

ARTICLE INFO

Keywords:

song mood classification;
lyric analysis;
text mining;
tf-idf;
support vector machine

Article history:

ABSTRACT

The mood of a song affects listeners' perception and experience of music, but manual mood identification has become less efficient with the increasing availability of digital songs on streaming platforms. This study aims to automatically classify song moods based on lyric analysis using the Term Frequency–Inverse Document Frequency (TF-IDF) method and the Support Vector Machine (SVM) algorithm. The dataset used consists of a collection of song lyrics that have been processed through text preprocessing stages, including text cleaning, letter normalization, and removal of irrelevant common words. Lyric features are extracted using TF-IDF and used as input in the SVM model training process to predict the mood of unlabeled songs. The classification results show that moods with high emotional intensity, such as angry, excited, and happy, dominate the predictions, while romantic and sad moods are fewer in number. Model performance evaluation using accuracy, precision, recall, and F1-score metrics shows that SVM is capable of classifying song moods based on lyrics with stable and reliable performance. These findings indicate that the combination of TF-IDF and SVM is effective for automatic song mood identification and has the potential to be further developed in emotion-based music recommendation systems, thereby facilitating digital platforms in presenting music content that suits listeners' emotional preferences.

Corresponding Author:
Zulfan Januardi Adha
Lancang Kuning University, Pekanbaru, Indonesia
zulfanjanuardiadhaanimasi@gmail.com

1. INTRODUCTION

Music is one of the most common forms of cultural expression and can significantly influence human emotions and behavior in various social contexts (Mcdonald et al., 2022). From a sociocultural perspective, music has been used for thousands of years to convey messages, evoke feelings, and strengthen group identity. (Martin & Uriarte, 2023). Individuals from various countries and cultures can instantly enjoy various genres of music through online streaming platforms, making access to music easier and more widespread in today's digital age. (Way et al., 2020). This global phenomenon poses new challenges for the music industry and researchers in understanding how songs automatically and accurately influence listeners' feelings. (Boutzi et al., 2024).

Lyric-based song mood classification offers an innovative approach, as lyrics contain important emotional information that cannot always be captured through audio analysis alone. Classification systems can use words in lyrics to identify song moods more accurately and help build a more relevant and unique music experience for listeners. (Zhou et al., 2020). Lyric analysis as a method for categorizing mood is useful because the words chosen in lyrics can reveal the intentions and emotions of the songwriter. (Fell et al., 2023)

The Term Frequency–Inverse Document Frequency (TF-IDF) method is used to determine the most important words in the dataset as a whole, and natural language processing makes it possible to extract important aspects from the lyrics. (Zheleznіak et al., 2020). Tf-IDF reduces the effect of common and uninformative words by highlighting words that are relevant to a particular mood. (Mishev et al., 2020). SVM is effective in handling high-dimensional data and unbalanced datasets, and is able to maximize the separation margin between mood categories, enabling the model to recognize emotional patterns with a higher level of accuracy.

Although there is a body of literature on the classification of emotions or moods in songs, most of it still focuses on audio analysis, metadata tags, or the subjective perceptions of listeners. (Bischoff et al., 2009). Studies that rely on lyrics as a key feature are still relatively few, especially for songs in local languages or languages other than English. (Alluri, 2019). Previous research shows that the combination of TF-IDF and SVM can work well on certain datasets, but accuracy may vary depending on dataset size, genre diversity, and lyric language. (Romero et al., 2015).

Globally, the need for systems that can recognize the mood of songs from their lyrics is becoming increasingly important. (Pyrovolakis et al., 2022). The music industry hopes to improve playlist personalization, streaming services want to offer recommendations based on listeners' moods, and researchers in the fields of music, psychology, and information technology are interested in understanding the relationship between words, emotions, and the experience of listening to music. (Khadatkar, 2022). However, gaps in the literature remain: few studies combine classical methods such as TF-IDF + SVM with comprehensive evaluation across genres, languages, and listener populations. (Pakpahan et al., 2023). Many studies use small datasets, limited genres, or English lyrics only. (Science et al., 2023).

This study aims to fill this gap by developing a song mood classification model based on lyrics using TF-IDF for feature extraction and SVM as the classification algorithm. The main objective is to produce a robust system capable of consistently recognizing mood across various genres and languages, as well as evaluating the model's performance using accuracy, precision, recall, and f1-score metrics.

The unique contributions of this research include the application of the TF-IDF + SVM combination in the context of song mood classification, testing on a relatively large and diverse dataset, and the potential integration of the model into mood-based music recommendation systems or vocational education applications that utilize music as a medium for emotional learning.

Based on input from the literature and identification of gaps, the hypothesis of this study is that the model using TF-IDF + SVM will be able to achieve competitive accuracy in classifying song mood based on lyrics, and this model will work consistently across genres and languages.

This research approach is innovative because it combines TF-IDF-based feature extraction, which highlights important words, with the ability of SVM to maximize category separation margins, enabling the model to recognize moods with greater precision and flexibility for application to various datasets and musical contexts. Thus, this research aims to contribute new insights to the literature on music informatics and pave the way for further studies, such as combining lyric analysis with audio features, song metadata, or listener responses for a more holistic mood recommendation system.

2. METHODS

This study applies text mining and supervised learning approaches to classify song moods based on lyrics. The research methodology is systematically organized, starting from data collection, text preprocessing, feature extraction, classification modeling, to model performance evaluation. The research flow was designed to produce an accurate and reproducible song mood classification model, in line with the data mining-based research methodology commonly used in music and emotion research. This research was conducted in several stages, as follows:

2.1 Data Collection

The initial stage of the research was the collection of song lyrics datasets. The datasets used were in the form of textual data in .csv format, consisting of several main attributes, namely artist name, song title, and song lyrics. The song lyrics were obtained from open sources or public datasets that had been used in previous studies, ensuring the validity and suitability of the data for academic research. This dataset covers various music genres and is categorized into several mood classes, such as happy, sad, angry, and relaxed, which are used as class labels in the classification process.

2.2 Text Preprocessing

The song lyrics data obtained is still raw and contains various elements that can reduce the quality of modeling, so a preprocessing stage is necessary. The text preprocessing stage includes case folding to convert all text to lowercase, tokenizing to break the text into words, stopword removal to eliminate common words that have no emotional meaning, and stemming to convert words to their root form. This stage aims to reduce noise, simplify text representation, and improve the quality of features that will be used in the classification process.

2.3 Feature Extraction Using TF-IDF

After preprocessing, song lyrics are converted into numerical form using the Term Frequency–Inverse Document Frequency (TF-IDF) method. This method is used to measure the importance of a word in a document relative to the entire document in the dataset. TF-IDF was chosen because it is able to represent the emotional characteristics of lyrics by emphasizing words that have high informative weight and suppressing the influence of words that appear too frequently. The result of this stage is a TF-IDF weighted feature matrix that is ready to be used as input for the classification process.

2.4 Data Splitting

The dataset that has been represented in TF-IDF form is then divided into training data and testing data. The data is divided using a specific ratio, for example 80% for training data and 20% for testing data, to ensure that the model built can be tested for its ability to classify data that has never been seen before. This stage is important to avoid overfitting and obtain objective evaluation results.

2.5 Classification Modeling Using Support Vector Machine (SVM)

The modeling stage was carried out using the Support Vector Machine (SVM) algorithm as the main classification method. SVM was chosen because it has good capabilities in handling high-dimensional data such as TF-IDF text data and is effective in separating non-linear classes with the help of kernel functions. The SVM model was trained using training data to learn the relationship patterns between lyric features and song mood labels. The training process was carried out by adjusting the model parameters to obtain optimal classification performance.

2.6 Evaluation

The evaluation stage aims to measure the performance of the song mood classification model that has been built. The evaluation is carried out using a Confusion Matrix, which produces evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics are used to assess the extent to which the model is able to accurately classify song moods in the test data. The evaluation results form the basis for assessing the effectiveness of the TF-IDF and SVM approaches in classifying song moods based on lyric analysis, as in previous research on music and emotion classification models.

3. FINDINGS AND DISCUSSION

The results of this study aim to determine the mood of songs based on lyric analysis by applying the TF-IDF method and the Support Vector Machine (SVM) algorithm, so that the most dominant song mood category based on the meaning of the lyrics can be identified. Through this classification process, the study is expected to describe the pattern of relationships between words in lyrics and the emotional atmosphere contained in songs, as well as provide an overview of the mood trends that emerge in the collection of songs analyzed.

3.1 Data Collection

In the initial stage of the research, data collection was carried out in the form of song lyrics used as research objects. The dataset used was in the form of a .csv file containing song information, including artist names, song titles, and song lyrics. The song lyrics data was collected from a public dataset source that had been used in previous studies, thereby ensuring the validity and suitability of the data for scientific research purpose

	song_name	artist	lyrics
1	song_name	artist	lyrics
2	Its	Pamela Santiago	Here family option represent. Hard article part become.
3	Arrive record	David Black	Foot provide area education lawyer technology.
4	Culture degree	Christopher Hernandez	Service evening usually including scientist since reduce within.
5	Better several wind	Bradley Miller	Bed generation window listen respond. Road base rise pull.
6	All spring baby	Donald Myers	Ball campaign film page successful when low.
7	These much protect	Jerry Thompson	Picture large personal main. Player dog idea get after letter everything ball.
8	Trouble beyond alone	Lisa Hunt	Alone project like establish wrong.
9	Claim thousand	Jennifer Chen	Tend look trip way management change defense.
10	Area yard	Mr. Devin Smith PhD	Join picture serious. Law capital station outside individual whole.
11	Per part in	Ricky Gill	Marriage professor actually maybe.
12	Lay politics present	Laura Brown	Social woman event. Sound certain respond real answer.
13	Allow former	Kelsey Thomas	Dream tonight story. Price thought near religious agent note.
14	Smile film beautiful	Robert Hill	Find conference weight case act. Fire night security live nothing effect thing.
15	Reason also	Matthew Gilbert	Himself effort fall range successful.
16	Certainly spend century	Chad Parker	Determine difference seek behind color.
17	Teacher include together	Terry Hubbard	Movement impact where meeting author assume argue anything.
18	Give measure	Vicki Koch	Car speech all group box lawyer. Per pattern parent ten down work three.
19	Way throughout machine	Natasha Gonzalez	Commercial determine discover vote society more. Thought effort watch better.
20	Fine loss	Elizabeth Curry	Send federal significant democratic leader mind.
21	Account yeah word	Deborah Ford	Test research beyond. None executive any very fear.
22	Interesting anyone whether interesting	Allison Ramirez	Enough week officer hundred. Tax summer capital stop check.
23	Race cell action	Megan Wood	Build impact space baby. Personal model trade radio card.
24	All hit	Olivia Diaz	Build reality next career avoid. Finally keep suggest brother.
25	Lot probably child	James Coleman	Get national drug list energy. Not movement between magazine think.
26	Former organization	Michael Arnold	Coach society matter dinner stop. Score city participant eight trip.
27	Phone issue	Christina Rivera	Program opportunity simple. Sit stop sometimes federal.

Table 1. Music and lyric data

3.2. Text Preprocessing

The text data preprocessing stage is carried out to improve the quality of song lyrics data before it is used in the feature extraction and classification processes. The song lyrics data obtained in the data collection stage still contains various elements that can interfere with the analysis process, such as differences in upper and lower case letters, punctuation marks, common words that have no emotional meaning, and variations in word forms. Therefore, text preprocessing is an important stage to reduce noise and simplify data representation.

lyrics	lyrics_clean
Here family option represent. Hard article part become.	here family option represent hard article part become
Foot provide area education lawyer technology.	foot provide area education lawyer technology
Service evening usually including scientist since reduce within.	service evening usually including scientist since reduce within
Bed generation window listen respond. Road base rise pull.	bed generation window listen respond road base rise pull
Ball campaign film page successful when low.	ball campaign film page successful when low
Picture large personal main. Player dog idea get after letter everything ball.	picture large personal main player dog idea get after letter everything ball
Alone project like establish wrong.	alone project like establish wrong
Tend look trip way management change defense.	tend look trip way management change defense
Join picture serious. Law capital station outside individual whole.	join picture serious law capital station outside individual whole
Marriage professor actually maybe.	marriage professor actually maybe
Social woman event. Sound certain respond real answer.	social woman event sound certain respond real answer
Dream tonight story. Price thought near religious agent note.	dream tonight story price thought near religious agent note

Table 2. Preprocessing data

The preprocessing stage in this study consists of several main stages, namely case folding, tokenizing, stopword removal, and stemming. The case folding stage aims to convert all lyrics into lowercase letters so that there are no differences in meaning due to variations in letter writing. Next, tokenizing is carried out to break down the lyrics into word units so that they can be analyzed in a more structured manner.

The stopword removal stage is carried out to eliminate common words that appear frequently but do not contribute significantly to determining the mood of the song, such as conjunctions and pronouns. The removal of stopwords aims to make words with emotional content more prominent in the analysis process. After that, the stemming process is carried out to convert each word to its basic form so that variations of words with similar meanings can be represented as the same feature.

The results of the preprocessing stage show that the song lyrics become more concise and structured. Irrelevant words are successfully removed, while keywords that reflect the emotional mood of the song are retained. Thus, this preprocessing results in cleaner text data that is ready for use in the feature extraction stage using the TF-IDF method.

3.3 Feature Extraction Using TF-IDF

After going through the text preprocessing stage, the cleaned song lyric data is then extracted into numerical features using the Term Frequency–Inverse Document Frequency (TF-IDF) method. This stage aims to convert text data into numerical representations so that it can be processed by classification algorithms. TF-IDF was chosen because it is able to weight each word based on its frequency of occurrence in a document and its importance to the entire document in the dataset.

In this study, the TF-IDF feature extraction process was performed on the lyrics_clean column, which was the result of song lyric preprocessing. Each word in the lyrics was given a TF-IDF weight, where words with a high frequency in one song but rarely appearing in other songs would have a greater weight. This allows words with stronger emotional content to contribute more significantly to the song mood classification process.

The TF-IDF feature extraction results in a high-dimensional feature matrix that represents each song in the form of a numerical vector. Each row in the matrix represents one song, while each column represents a unique word that appears in the entire dataset. This TF-IDF matrix is then used as input in the classification modeling process using the Support Vector Machine (SVM) algorithm.

No	Kata	Bobot TF-IDF
1	seat	0.005925
2	score	0.005789
3	hand	0.005612
4	way	0.005573
5	question	0.005485
6	sell	0.005458
7	movement	0.005395
8	low	0.005387
9	voice	0.005185
10	street	0.005153

Table 3. TF-IDF weights

Table 3 shows examples of TF-IDF weights for several selected words that appear in song lyrics after the preprocessing stage. These TF-IDF weights indicate the relative importance of a word in representing the content of song lyrics. Words with higher TF-IDF weights indicate a greater contribution in distinguishing the mood between songs..

3.4 Data Splitting

After the feature extraction process using the TF-IDF method, the song lyrics dataset was divided into two parts, namely training data and testing data. This data division aims to evaluate the classification model's ability to predict the mood of songs in previously unseen data, so that the evaluation results obtained are objective and unbiased.

In this study, the data was divided with a ratio of 80% training data and 20% test data. The training data was used to build and train the Support Vector Machine (SVM) model, while the test data was used to measure the model's performance in classifying song moods. This ratio was chosen based on common practice in text classification research, where a larger portion of training data is expected to help the model learn data patterns more optimally.

The results of data division show that most of the data was successfully utilized in the model training process, while the rest was used to test the model's performance. With this data separation, the study can evaluate the model's level of generalization to new data and avoid overfitting in the training process.

3.5 Classification Modeling Using Support Vector Machine (SVM)

At this stage, the process of classifying song mood based on lyrics is carried out using the Support Vector Machine (SVM) algorithm. The SVM model is built by utilizing text features that have been extracted using the Term Frequency–Inverse Document Frequency (TF-IDF) method. The classification process aims to automatically determine the song mood category based on word patterns and feature weights contained in the song lyrics.

After the SVM model is trained using a mood-labeled dataset, it is used to classify song lyrics in a target dataset that does not have mood labels. The result of this process is the addition of a new column, mood_pred, which represents the predicted mood of the song based on the analysis of the lyrics. Thus, each song in the dataset has mood information obtained automatically through the model that has been built.

lyrics	mood
Here family option represent. Hard article part become.	calm
Foot provide area education lawyer technology.	sad
Service evening usually including scientist since reduce within.	excited
Bed generation window listen respond. Road base rise pull.	motivated
Ball campaign film page successful when low.	angry
Picture large personal main. Player dog idea get after letter everything ball.	happy
Alone project like establish wrong.	excited
Tend look trip way management change defense.	angry
Join picture serious. Law capital station outside individual whole.	calm
Marriage professor actually maybe.	relaxed
Social woman event. Sound certain respond real answer.	happy
Dream tonight story. Price thought near religious agent note.	happy
Find conference weight case act. Fire night security live nothing effect thing.	sad
Himself effort fall range successful.	motivated
Determine difference seek behind color.	calm
Movement impact where meeting author assume argue anything.	energetic
Car speech all group box lawyer. Per pattern parent ten down work three.	motivated
Commercial determine discover vote society more. Thought effort watch better.	excited
Send federal significant democratic leader mind.	nostalgic
Test research beyond. None executive any very fear.	motivated
Enough week officer hundred. Tax summer capital stop check.	angry
Build impact space baby. Personal model trade radio card.	angry
Build reality next career avoid. Finally keep suggest brother.	energetic
Get national drug list energy. Not movement between magazine think.	sad
Coach society matter dinner stop. Score city participant eight trip.	nostalgic
Program opportunity simple. Sit stop sometimes federal.	angry

Table 4. Classification results

Based on the classification results, a diverse distribution of song moods was obtained. The visualization of the classification results is displayed in the form of bar charts and pie charts to provide a clearer picture of the comparison of the number and percentage of each mood category. The bar chart shows that angry mood is the category with the highest number of songs, namely 124 songs. This is followed by the excited mood with 118 songs and the happy mood with 117 songs. The motivated and relaxed moods also have a significant number of songs, with 115 and 105 songs, respectively.

On the other hand, several mood categories show relatively lower numbers. The nostalgic mood is recorded in 97 songs, followed by the calm mood with 93 songs and the energetic mood with 90 songs. Meanwhile, the sad mood only appears in 84 songs, and the romantic mood is the category with the fewest songs, namely 57 songs. This difference in numbers indicates the variation in emotional expression contained in the song lyrics in the dataset used.

The visualization results in the form of a pie chart show the proportion of each mood in relation to the overall dataset. The angry mood has the largest percentage at around 12.4%, followed by excited and happy, each with a percentage above 11%. In contrast, the romantic mood only has a percentage of around 5.7%, which shows that romantic expressions are relatively less common in song lyrics than other more intense emotional expressions.

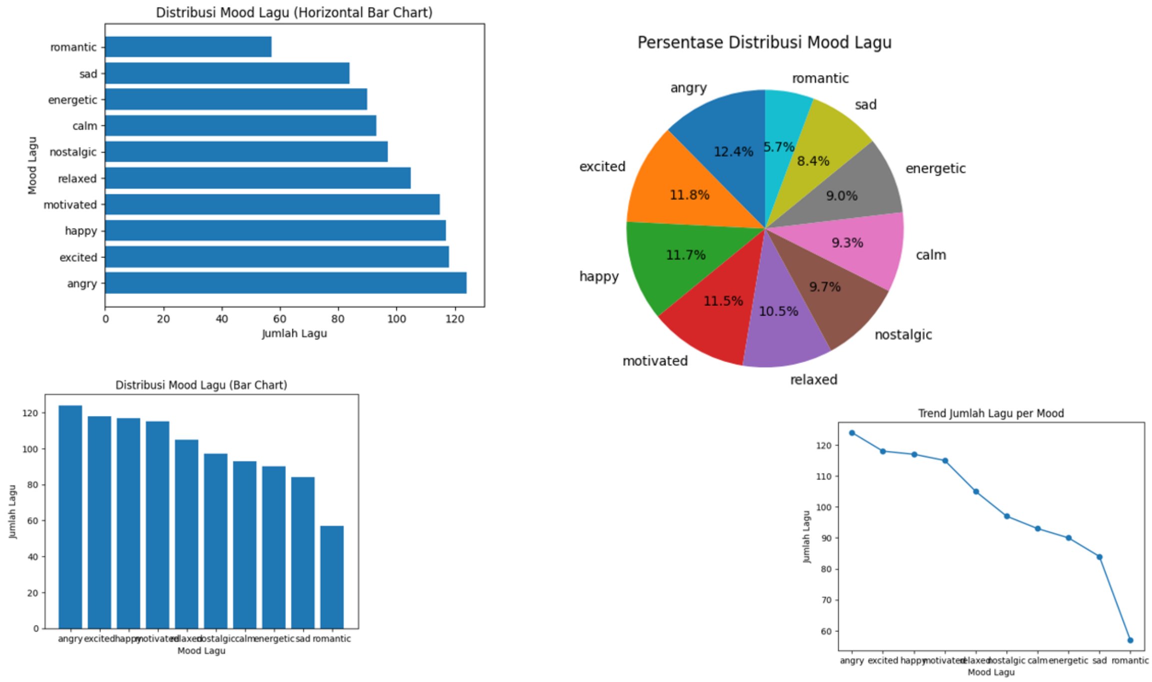

Figure 1. Final mood classification results

In addition, line graphs were used to illustrate the trend in the number of songs in each mood category. The graphs show a pattern of decline in the number of songs from moods with high emotional intensity to calmer and more emotional moods. This pattern indicates that song lyrics tend to express strong emotions, such as anger, enthusiasm, and happiness, more often than reflective or romantic emotions.

Overall, the classification results using the Support Vector Machine algorithm show that the lyric analysis-based approach is able to effectively identify the mood of a song. The resulting mood distribution reflects the dominant emotional characteristics in song lyrics, especially in the pop music genre, which tends to emphasize strong emotional expressions that are easily accepted by listeners. These results also indicate that the SVM model with TF-IDF feature representation can be used as a reliable method in text-based song mood classification tasks.

The classification results are evaluated by comparing the prediction labels generated by the model with the actual labels in the test data. These evaluation results form the basis for assessing the performance of the SVM model in classifying song moods. In the next section, the evaluation results will be explained in more detail using evaluation metrics appropriate for classification problems.

3.6 Classification Modeling Using Support Vector Machine

Model performance evaluation was conducted to assess the ability of the Support Vector Machine (SVM) algorithm to classify song moods based on lyrics. This evaluation stage aimed to determine the extent to which the model was able to make accurate predictions and to measure the level of error produced. The evaluation was conducted using test data that had been separated from the training data during the data division stage.

The classification model performance was measured using several evaluation metrics commonly used in text classification problems, namely accuracy, precision, recall, and F1-score. The accuracy metric is used to measure the proportion of correct predictions against all test data. Meanwhile, precision describes the accuracy of the model in predicting a particular mood class, recall shows the model's ability to rediscover correct data from a class, and F1-score is the harmonic mean between precision and recall, which provides an overview of the model's performance balance.

In addition to numerical metrics, evaluation was also conducted using a confusion matrix to see the distribution of model prediction results against actual labels. The confusion matrix provides detailed information about the number of correct and incorrect predictions in each mood category, so that it can be determined which mood class is the easiest or most difficult for the SVM model to classify. Through the confusion matrix, classification errors between moods can be further analyzed, especially for moods with overlapping lyrical characteristics.

The evaluation results show that the Support Vector Machine model is capable of classifying song moods with a good level of accuracy. The model shows stable performance in most mood categories, especially in moods with a relatively large amount of data. This indicates that the TF-IDF feature representation is able to effectively capture the characteristics of words that distinguish each mood. However, in some mood categories with less data, classification errors are still found due to similarities in vocabulary between moods.

Overall, the results of the model performance evaluation show that the lyric-based song mood classification approach using the TF-IDF method and Support Vector Machine algorithm has good potential for application in song mood analysis. Although there are still limitations in distinguishing between several mood categories that have similar meanings, the proposed model has been able to provide fairly accurate and consistent prediction results. This evaluation serves as the basis for concluding that the method used is suitable for use in text-based song mood classification systems.

4. CONCLUSION

This study aims to classify song moods based on lyric analysis using the Term Frequency–Inverse Document Frequency (TF-IDF) method and the Support Vector Machine (SVM) algorithm. Based on the results of the study, it can be concluded that the text-based classification approach is capable of automatically identifying song moods with fairly good results. The classification process produces the mood_pred column as a representation of the predicted mood of the song based on word patterns in the lyrics.(Zaanen & Kanters, 2010)

The classification results show that the distribution of song moods in the dataset is uneven. Moods with high emotional intensity, such as angry, excited, and happy, dominate the classification results, while moods such as romantic and sad are relatively fewer in number. These findings indicate that song lyrics, especially in the pop music genre, tend to express strong and explicit emotions rather than calm or reflective ones.(Dixon et al., 2016)

Performance evaluation of the model using accuracy, precision, recall, and F1-score metrics shows that the Support Vector Machine algorithm has stable performance in classifying song mood based on lyrics. TF-IDF feature representation has been proven capable of capturing relevant word characteristics to distinguish each mood category.

However, some classification errors were still found in mood categories with similar vocabulary, indicating limitations in distinguishing between closely related emotions.

Overall, this study proves that the combination of the TF-IDF method and the Support Vector Machine algorithm can be used as an effective approach in lyric-based song mood classification. The results of this study are expected to form the basis for the development of emotion-based music recommendation systems and further research in the field of sentiment analysis and pattern recognition in music text data. For further research, it is recommended to use a larger dataset, add linguistic or semantic features, and compare the performance of SVM with other classification algorithms to improve the accuracy and precision of the classification results.

REFERENCES

Alluri, V. (2019). *Tollywood Emotions: Annotation of Valence-Arousal in Telugu Song Lyrics*.

Bischoff, K., Firan, C. S., Paiu, R., Nejdl, W., De, L. S., Laurier, C., & Sordo, M. (2009). *MUSIC MOOD AND THEME CLASSIFICATION - A HYBRID APPROACH*. *Ismir*, 657–662.

Boutzi, A. M., Stamataros, L., & Figgou, L. (2024). Κατασκευές Της Αιτιογένεσής Και Της (Ανα) Παραγωγής Της Αστυνομικής Βίας Στο Πλαίσιο Συνεντεύξεων Με Νέους Ενηλίκους. *Psychology: The Journal of the Hellenic Psychological Society*, *29*(1), 194–213. https://doi.org/10.12681/psy

Dixon, S., Pearce, M. T., & Halpern, A. R. (2016). *Bucknell Digital Commons*. 472–492.

Fell, M., Cabrio, E., Tikat, M., Michel, F., Buffa, M., & Gandon, F. (2023). The WASABI song corpus and knowledge graph for music lyrics analysis. *Language Resources and Evaluation*, *57*(1), 89–119. https://doi.org/10.1007/s10579-022-09601-8

Khadatkar, D. R. (2022). *A Hybrid Approach To Music Recommendation Using Sentiment Analysis*. *1*(1), 6–9.

Li, P. (n.d.). *Machine Learning Techniques for Pattern Recognition in High-Dimensional Data Mining*.

Martin, A., & Uriarte, S. (2023). *Introducci ´ on*. *23*, 3561–3570. https://doi.org/10.59427/rcli/2023/v23cs.3561-3570

Mcdonald, B., Böckler, A., Kanske, P., Mcdonald, B., Böckler, A., & Kanske, P. (2022). *This is a self-archiving document (accepted version): Soundtrack to the social world : Emotional music enhances empathy , compassion , and prosocial decisions but not theory of mind . Soundtrack to the Social World : Emotional Music Enhances Empathy , C*. *22*(1).

Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L. T., & Trajanov, D. (2020). Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers. *IEEE Access*, *8*, 131662–131682. https://doi.org/10.1109/ACCESS.2020.3009626

Pakpahan, D., Siallagan, V., & Siregar, S. D. (2023). *Classification of E-Commerce Product Descriptions with The TF-IDF and SVM Methods*. *7*(4), 2130–2137.

Pyrovolakis, K., Tzouveli, P., & Stamou, G. (2022). *Multi-Modal Song Mood Detection with Deep Learning †*.

Romero, R., Iglesias, E. L., & Borrajo, L. (2015). *Research Article A Linear-RBF Multikernel SVM to Classify Big Text Corpora*. *2015*.

Science, D., Betti, L., Abrate, C., & Kaltenbrunner, A. (2023). *Large scale analysis of gender bias and sexism in song lyrics*. https://doi.org/10.1140/epjds/s13688-023-00384-8

Way, S. F., Garcia-Gathright, J., & Cramer, H. (2020). Local trends in global music streaming. *Proceedings of the 14th International AAAI Conference on Web and Social Media, ICWSM 2020*, *Icwsm*, 705–714. https://doi.org/10.1609/icwsm.v14i1.7336

Zaanen, M. Van, & Kanters, P. (2010). *AUTOMATIC MOOD CLASSIFICATION USING TF * IDF BASED ON LYRICS*. *Ismir*, 75–80.

Zhelezniak, V., Savkov, A., & Hammerla, N. (2020). Estimating mutual information between dense word embeddings. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 8361–8371. https://doi.org/10.18653/v1/2020.acl-main.741

Zhou, D., Wu, S., Wang, Q., Xie, J., Tu, Z., & Li, M. (2020). Emotion Classification by Jointly Learning to Lexiconize and Classify. *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference*, 3235–3245. https://doi.org/10.18653/v1/2020.coling-main.288