Sentiment Analysis on Songs based on Song Lyrics using Naïve Bayes Algorithm

Ayessa Amor N. Hernandez and Concepcion L. Khan

***Abstract*— Music induces basic to complex emotions such as happiness, sadness, and nostalgia. These emotions can be classified into categories like positive or negative using sentiment analysis. Existing studies on mood classification mostly focus on the audio features of a song while the lyric features are ignored. A few studies on lyrics mood classification, on the other hand, pointed out the need to explore other classifiers like Naïve Bayes and improve its performance, using a larger dataset. In this study, a Naïve Bayes classifier model was created to identify whether a song is positive or negative based on its lyrics. The model which produced exceptional results with 95.02% accuracy and 94.42% precision was trained and tested using a dataset containing 1,810 song lyrics. Feature extraction techniques such as N-grams (tri-grams) and TF-IDF were applied after preprocessing the data.**

***Index Terms*— Naïve Bayes, NLP, sentiment analysis, TF-IDF, N-grams**

I. Introduction

A. Background of the Study

Music makes people feel different emotions which affect their mood. It may induce basic to complex emotions such as happiness, sadness, or nostalgia [1]. These emotions or moods can be classified using sentiment analysis.

The three experiments conducted by [2] show that lyrics "appear to have greater power to direct mood change than music alone and can imbue a particular melody with affective qualities".

Sentiment analysis is a Natural Language Processing (NLP) technique to extract sentiments from texts like lyrics. According to [3], sentiment analysis can be used in fields involving brand monitoring, integrated analysis, public relations, marketing, data mining, and political analysis.

B. Statement of the Problem

Existing studies on mood classification mostly focus on the audio features of a song while the lyric features are ignored. However, the lyric features of a song have proved its usefulness in mood classification. Other researchers who conducted studies in lyrics mood classification pointed out the need to explore other classifiers such as Naïve Bayes since this classifier is one of the top classifiers for text classification. When modeling the Naïve Bayes classifier, there is an observation of a need to use a larger dataset when training and testing while producing higher accuracy scores.

Presented to the Faculty of the Institute of Computer Science, University of the Philippines Los Baños in partial fulfillment of the requirements for the Degree of Bachelor of Science in Computer Science

Lyrics are hard to classify because lyrics are mostly abstract, and emotions are conveyed indirectly. There is a need to discover a combination of feature extraction techniques that will help to produce higher accuracy scores.

There are no existing open-source sentiment-classified datasets with readily available lyrics that can be used for training and testing mood classifiers due to copyright issues thus requiring the need to collect song lyrics to use.

C. Significance of the Study

Sentiment analysis on song lyrics using Naïve Bayes can be used in different music and technology fields such as in Music Information Retrieval (MIR) systems, Music Emotion Recognition (MER) systems, and music psychology.

MIR is a multidisciplinary research field that aims to develop systems and processes to retrieve information from music [4]. Perception of music is studied in the field of MIR and is usually applied in music recommendation systems and music search engines [5].

Music Emotion Recognition (MER) is a computational task that aims to automatically recognize emotions from music. With the advancement of brain science, music emotion recognition (MER) has received a lot of attention in academia and business. For example, recommendation systems, automatic music composition, psychotherapy, and music visualization have all made extensive use of MER [6].

Music psychology is a field of study that aims to understand and explain human behavior and experience toward music. Research in music psychology has applications in a variety of fields, including music composition, performance, education, criticism, therapy, and studies of human attitude, performance, intelligence, creativity, and social behavior.

D. Objectives of the Study

The main objective of this study is to evaluate the performance of the Naive Bayes classifier algorithm on sentiment analysis based on song lyrics, thus the specific objectives are as follows:

1) Collect song lyrics from the internet
2) Apply data cleaning and data preprocessing (tokenization, feature extraction techniques, etc.) to the dataset
3) Train and test the dataset
4) Evaluate the performance of the classifier

E. Scope and Limitations

This study focuses on the evaluation of the Naïve Bayes classifier algorithm and thus other classifier algorithms (e.g. SVM) are not used. Also, the dataset used in this study is limited to English song lyrics only.

II. Review of Related Literature

Sentiment analysis is commonly used in analyzing and understanding product or service reviews, or sentiments of a group of people regarding an issue. [7] did a sentiment analysis of the public's COVID-19 sentiments using Tweets. The study used the Naive Bayes and the logistic regression classification methods, and a strong accuracy of 91% was observed with the Naive Bayes method.

[8] proved in their study the usefulness of lyrics in mood classification by utilizing linguistic lyric features and text stylistic features of song lyrics. They pointed out the need to explore other classification models in mood classifications, other than using SVM or Support Vector Machine.

A study by [9] showed the success of Naïve Bayes in lyrics classification when they classified if a song was by Metallica or Nirvana. The model got a precision score of 93% with a Laplace estimator of 0.06.

In a study by [10], they manually labeled their 1200 dataset (1000 for training and 200 for testing) into happy and sad categories. Their Naïve Bayes model got an 80% accuracy in training and 54.5% accuracy in testing. In another study by [11] on music mood recognition using lyrics, they used the Naïve Bayes algorithm to train 120 songs and test 60 songs and their model got an 85% accuracy score. These studies showed the need to (1) produce higher accuracy scores in testing while (2) using a larger dataset.

III. Methodology

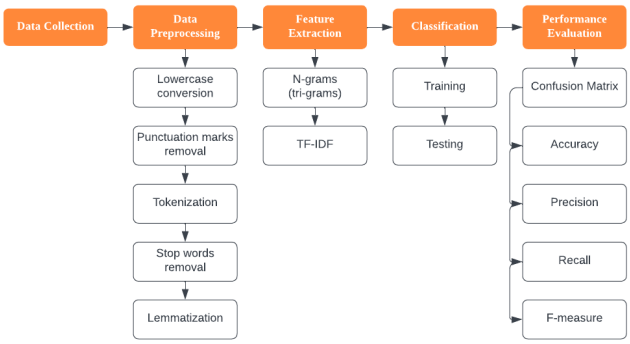

Fig. 1: Flowchart

1) Data Collection: *LyricsGenius* is an easy-to-use *python* library for searching lyrics from Genius.com, a site containing a collection of song lyrics around the globe [12].

A sentiment-classified dataset from MoodyLyrics [13] which consists of 2000 songs with information such as the artist and the song title is used to search for a song.

However, while collecting the data, there are song lyrics not found and song lyrics that are not in English resulting in an unbalanced dataset with 1810 song lyrics.

2) Data Preprocessing: In this study, Python's *NLTK* is used for preprocessing. These are the steps done to clean and preprocess the data:

a) Lowercase conversion
b) Punctuation marks removal
c) Tokenization
d) Stop words removal
e) Lemmatization

Figure 2 shows a sample of the data before cleaning and preprocessing, while **Figure 3** shows the changes after cleaning, and **Figure 4** shows the changes after preprocessing.

I walked through the door with you, the air was cold

But something 'bout it felt like home somehow

And I left my scarf there at your sister's house

And you've still got it in your drawer, even now

Fig. 2: Before cleaning and preprocessing

i walked through the door with you the air was cold

but something bout it felt like home somehow

and i left my scarf there at your sister s house

and you ve still got it in your drawer even now

Fig. 3: After cleaning

walked door air cold something bout felt like home somehow left scarf sister house still got drawer

Fig. 4: After preprocessing

3) Feature Extraction: Feature extraction techniques like *n-grams* (tri-grams) and TF-IDF are used in this study. It is done using the *CountVectorizer* and *TfidfVectorizer* of *sklearn*

4) Classification: The Naïve Bayes classifier is used to train and test the data. The dataset with 1810 song lyrics is split, 66.67% for training and 33.33% for testing.

5) Performance Evaluation: To evaluate the performance of the model, the following metrics were used:

a) Confusion Matrix

Actual Class / Predicted Class	1	0
1	True Positive	False Positive
0	False Negative	True Negative

Fig. 5: Confusion Matrix

b) Accuracy

$$\frac{(TP + TN)}{(TP + FP + TN + FN)}$$

c) Precision

$$\frac{TP}{(TP + FP)}$$

d) Recall

$$\frac{TP}{(TP + FN)}$$

e) F-measure

$$\frac{(2 * precision * recall)}{(precision + recall)}$$

IV. Results and Discussion

The following visualizations are presented to get an idea of what the data tells us:

Fig. 6: A word cloud of the songs with a positive mood

Figure 6 shows that words like *love* and *baby* are often used in positive songs while **Figure 7** shows that *fire* and *know* are often used in negative songs. With this, the difference between the lyrics of each mood is shown.

The following evaluation shows the performance of the Naïve Bayes classifier in identifying the mood of a song based on its lyrics.

Figure 8 shows that out of the 603 song lyrics for testing, 323 songs for the positive class and 280 songs for the negative

Fig. 7: A word cloud of the songs with a negative mood

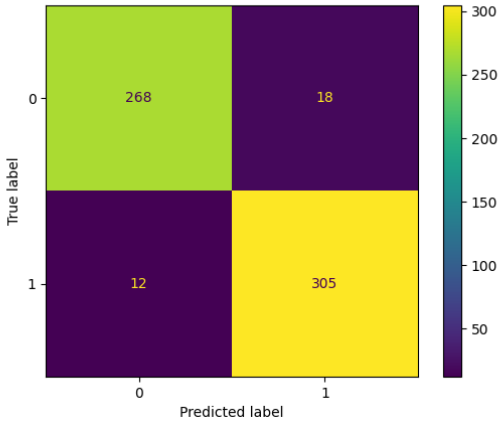

Fig. 8: Confusion matrix of the NB classifier

class, the classifier correctly classified the positive class with 94.42% precision while getting a 95.71% precision for the negative class. This implies that the classifier is effective for classifying both positive and negative classes which is what is needed for mood classification.

Accuracy	95.02%
Precision	94.42%
Recall	96.21%
F1-score	95.31%

Fig. 9: Evaluation scores of the classifier

In **Figure 9**, the F1-score shows that the classifier, despite being trained with an unbalanced dataset, obtained incredible results. **Figure 9** shows that the overall performance of the classifier is exceptional but should always take into account that the results are dependent on the sentiment-annotated dataset, MoodyLyrics, which is used in this study.

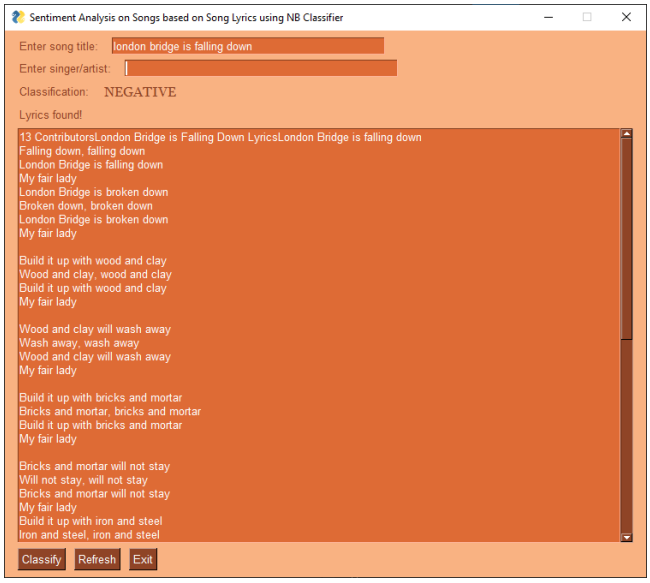

Fig. 10: A snapshot of the sample run

V. Conclusion and Future Work

A Naïve Bayes classifier was created to evaluate its performance on mood classification of songs based on the song lyrics. The classifier showed high evaluation scores after being trained and tested on a dataset consisting of 1,810 song lyrics.

While collecting song lyrics from Genius.com using the *lyricsgenius* library, there are song search requests made by the library that appears to be successful but do not qualify as correct, for example, returning a book passage or a list of songs instead of the song lyrics. Because of these errors, data cleaning can be challenging for much larger datasets.

With proper data pre-processing and feature extraction techniques, it can be concluded from this study that the Naïve Bayes algorithm serves as a powerful tool in song lyrics mood classification. However, despite observing exceptional results in this study, this work can be improved by using a much larger dataset to see the stability of the effectiveness of the classifier.

Acknowledgment

Many thanks to Assoc Prof. Khan for her guidance in the making of this study.

References

[1] P. N. Juslin, L. Harmat, and T. Eerola, "What makes music emotionally significant? Exploring the underlying mechanisms," *Psychology of Music*, vol. 42, no. 4, pp. 599–623, 7 2014.

[2] V. N. Stratton and A. H. Zalanowski, "Affective Impact of Music Vs. Lyrics," *Empirical Studies of The Arts*, vol. 12, no. 2, pp. 173–184, 7 1994.

[3] N. Kapoor, "Types of Sentiment Analysis and Its Uses - The Startup - Medium," 12 2021. [Online]. Available: https://medium.com/swlh/types-of-sentiment-analysis-and-its-uses-ad733535c895

[4] J. A. Burgoyne, I. Fujinaga, and J. S. Downie, *Music Information Retrieval*. Wiley, 11 2015.

[5] M. Fell, E. Cabrio, and F. Gandon, "Natural language processing for music information retrieval : Deep analysis of lyrics structure and content," 2020.

[6] X. Cui, Y. Wu, J. Wu, Z. You, J. Xiahou, and M. Ouyang, "A review: Music-emotion recognition and analysis based on EEG signals," *Frontiers in Neuroinformatics*, vol. 16, 10 2022. [Online]. Available: https://doi.org/10.3389/fninf.2022.997282

[7] J. Samuel, G. G. M. N. Ali, M. Rahman, E. Esawi, and Y. Samuel, "COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification," *Information*, vol. 11, no. 6, p. 314, 5 2020. [Online]. Available: https://doi.org/10.3390/info11060314

[8] S. Shukla, P. Khanna, and K. C. Agrawal, "Review on sentiment analysis on music," 12 2017. [Online]. Available: https://doi.org/10.1109/ictus.2017.8286111

[9] D. Buzic and J. Dobša, "Lyrics classification using Naive Bayes," 5 2018. [Online]. Available: https://doi.org/10.23919/mipro.2018.8400185

[10] S. Raschka, "MusicMood: Predicting the mood of music from song lyrics using machine learning," 11 2016. [Online]. Available: https://arxiv.org/abs/1611.00138

[11] K. R. Tan, M. L. Villarino, and C. Maderazo, "Automatic music mood recognition using Russell's twodimensional valence-arousal space from audio and lyrical data as classified using SVM and Naïve Bayes," *IOP conference series*, vol. 482, p. 012019, 3 2019. [Online]. Available: https://doi.org/10.1088/1757-899x/482/1/012019

[12] Johnwmillr, "GitHub - johnwmillr/LyricsGenius: Download song lyrics and metadata from Genius.com ," 2020. [Online]. Available: https://github.com/johnwmillr/LyricsGenius

[13] E. Çano and M. Morisio, "MoodyLyrics," *MoodyLyrics: A Sentiment Annotated Lyrics Dataset*, 3 2017. [Online]. Available: https://doi.org/10.1145/3059336.3059340

Ayessa Amor N. Hernandez A BSCS undergraduate student from the University of the Philippines - Los Baños who loves music and finds data science an interesting field.