# 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.


## Objectives

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

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

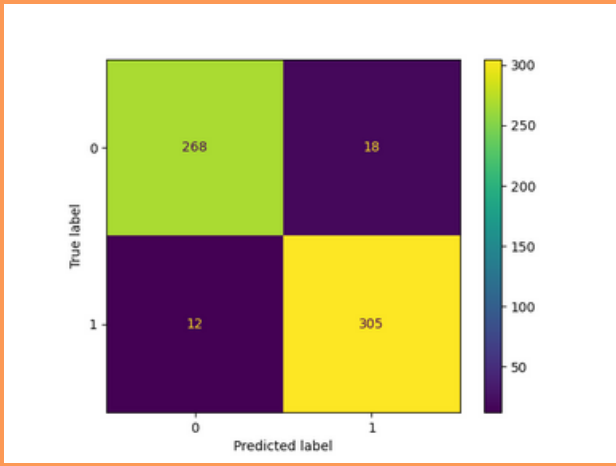
## Methodology

| | |
|---|---|
| Data Collection | Genius.com |
| Data Preprocessing | Lowercase conversion \| Punctuation marks removal \| Tokenization \| Stop words removal \| Lemmatization |
| Feature Extraction | *N-grams (tri-grams)* \| TF-IDF |
| Classification | Multinomial Naive Bayes (Training and Testing) |
| Performance Evaluation | Confusion Matrix \| Accuracy \| Precision \| Recall \| F-measure |

## The Author

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

## Results




Out of the 603 song lyrics for testing, 323 songs for the positive class and 280 songs for the negative 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% |
| F-measure | 95.31% |

## Conclusion

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.

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.