Classifying Music Genres using Topic Modeling of Lyrics

Frederick Nicholas Su
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
frederick.nicholas@binus.ac.id

Philbert
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
philbert001@binus.ac.id

Izhar Octarfilian Susilo
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
izhar.susilo@binus.ac.id

***Abstract*—Music genre classification traditionally depends on musical and audio characteristics such as chord progressions, rhythm, and tempo. However, lyrical content is often unused, even though a high amount of semantic information can be extracted to reveal thematic and stylistic differences between genres. This study explores the relationship between lyrical content and different music genres using topic modeling techniques such as Non-negative Matrix Factorization (NMF).**

***Index Terms*—Topic Modeling, Music Genre Classification, Lyrics Analysis, Natural Language Processing, Non-negative Matrix Factorization (NMF), Machine Learning.**

I. Introduction

Music genres are used to group distinct styles and themes of music perceived by the listener. While traditional methods of genre classification rely on thematic chord progressions, rhythmic changes, and tempo, classification using lyrical content remains an underexplored niche which offers highly informative content relating to the music genre it could be in. This study aims to uncover the relation of lyrical content with certain music genres and classify them accordingly through topic modeling.

By using various topic modeling techniques such as Non-negative Matrix Factorization (NMF), this study analyzes and takes lyrical content with semantic significance, using them to identify unique patterns and relations between certain words and genres. Comparing these relations across a broad range of genres clues into better understanding of the difference between one genre to another, the relations between genres based on lyrical content, and recurring patterns that define a genre.

II. Related Works

Natural Language Processing (NLP) has been widely used to analyze song lyrics in recent years. Previous research has shown that lyrics often reflect cultural and social values. Another study about linguistic patterns in hip-hop lyrics found that lyric themes evolved over time, revealing changes in social expression and complexity [1]. Large lyric datasets such as the Genius Song Lyrics dataset have also been used in lyric analysis projects to support NLP research and topic discovery in music [2].

Several works focus on discovering hidden themes and patterns in lyrics. Chen et al. applied topic modeling to analyze over 500,000 song lyrics and found recurring themes across music genres, showing that topic modeling can effectively uncover latent meaning in lyrics [3]. Earlier work by Grachten et al. explored NLP challenges in lyric processing, such as informal language and repetition, which makes lyric analysis different from normal text analysis [4]. Mohammed et al. combined sentiment analysis with topic classification to detect emotional themes in lyrics, showing how NLP can reveal both emotional tones and thematic structure in music [5].

These studies demonstrate that NLP is a powerful tool for lyric analysis. However, most previous research focused on sentiment or bias analysis rather than purely exploring topics or themes in a simple and interpretable way. Therefore, this project continues previous work by applying topic modeling to discover and analyze recurring themes in song lyrics in a more focused manner.

III. Methodology

The methodology for this study is designed to classify music genres from lyrical content, drawing upon established techniques in natural language processing and machine learning, similar to the approach benchmarked by Zubiaga et al. (2025) [6]. The process follows a sequential pipeline consisting of four main stages: (1) Data Acquisition and Preprocessing, (2) Thematic Feature Extraction using Topic Modeling, (3) Genre Classification, and (4) Model Evaluation.

A. Data Acquisition and Preprocessing

The foundation of this experiment is the Genius Song Lyrics dataset sourced from Kaggle [2], which contains over 5,000,000 song entries. Each entry includes metadata such as artist, song title, genre, lyrics, and language.. The raw dataset was refined through the following preprocessing steps:

1) **Chunk-Based Processing:** To prevent memory overflow, the csv file is iterate in chunks of 20,000 songs. The processing stops once it reaches 500,000 songs.
2) **Language and Data Filtering:** The dataset is filtered to include only English ('en') lyrics. Entries with missing values in the 'lyrics' or 'tag' (genre) columns are dropped.

3) **Fast Preprocessing Pipleline:** The lyrical text for each song was cleaned and standardized using a quick pipeline:
   - **Lowercasing and Character Removal:** Text is converted to lowercase and all non-alphabetic characters are removed using regular expressions.
   - **Custom Stop Word Removal:** in addition to standard English stop words, specific lyrical filler words such as ”ooh,” ”yeah,” ”ah,” ”verse,” ”chorus,” ”intro,” and ”outro” are removed.
4) **Dataset Balancing:** To prevent the model from biasing toward more frequent genres (like Rap or Pop), a balancing function was implemented. This ensured that each genre was represented by exactly 10,000 samples, resulting in a balanced dataset of 60,000 songs for the final training phase.

B. Thematic Feature Extraction using Topic Modeling

To convert the unstructured lyrical text into a quantitative feature set, we employed a topic modeling approach. The goal was to represent each song not as a collection of words, but as a distribution of underlying lyrical themes.

1) **Vectorization**: The preprocessed lyrics were first converted into a numerical format using a Term Frequency-Inverse Document Frequency (TF-IDF) Vectorizer. This creates a document-term matrix where each song is represented by the weighted importance of the words it contains.
2) **Topic Modeling with NMF**: Non-negative Matrix Factorization (NMF) was chosen as the topic modeling algorithm. NMF decomposes the TF-IDF matrix into a document-topic matrix and a topic-word matrix. The resulting document-topic matrix was used as the primary feature set for our classification models. Each song is now represented as a vector, where each element corresponds to the song’s association with a latent topic. The number of topics was treated as a hyperparameter and set to a value (e.g., 50) determined through preliminary experimentation.

C. Genre Classification

With each song represented as a topic vector, the task was framed as a multiclass classification problem. The objective is to train a model that can predict a song’s genre based on its thematic composition.

1) Data Splitting: The dataset was partitioned into a training set (80%) and a testing set (20%). A stratified split was used to ensure that the proportion of each genre was the same in both the training and testing sets, which is crucial for handling class imbalance.
2) Model Selection: To comprehensively evaluate the effectiveness of the topic features, three distinct classification models were trained and compared:
   a) **Logistic Regression**: A robust and interpretable linear model serving as a baseline.
   b) **Linear Support Vector Machine (SVM)**: A linear classifier that performs well in high-dimensional feature spaces but is limited to linearly separable decision boundaries.
   c) **Random Forest**: An ensemble model to capture more complex, non-linear relationships between lyrical themes and genres.

D. Evaluation

The performance of the trained classifiers on the unseen test set was measured using a suite of standard evaluation metrics to provide a holistic view of their effectiveness.

1) **Accuracy**: The overall percentage of correctly classified songs.
2) **Precision, Recall, and F1-Score**: These metrics were calculated on a per-class basis to assess the model’s performance for each individual genre.
3) **Confusion Matrix**: A confusion matrix was generated for the best-performing model to provide qualitative insights into which genres were most easily confused with one another.

IV. Results and Analysis

The three models used for analysis yielded results with variable success which can be seen in Table I. Logistic Regression and linear SVM showed modest performance across all four metrics with accuracies close to random chance.

TABLE I
Classification Report

	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.4983	0.4776	0.4983	0.4827
SVM (Linear)	0.5129	0.4906	0.5129	0.4924
Random Forest	0.5613	0.5484	0.5613	0.5511

The Random Forest model performed a lot better than Logistic Regression and linear SVM achieving an accuracy of 56.13%. This suggests that the relationship between lyrical themes and music genres is non-linear. While certain topics are indicative of a genre, it is the complex combination and weighting of multiple themes that provides the best predictive power.

To give reveal further insights to the varying model performance, a per-genre performance of the best performing model (Random Forest) was generated. The specific performance scores across different metrics can be seen in Table II. The Rap and Misc genres achieve the best results, with F1-scores of 0.80 and 0.79 respectively, indicating that their lyrical topics are relatively distinctive and consistently captured by the topic modeling representation. On the other hand, performance on genres such as Rock and Pop showed much lower scores, not even reaching 0.50 on most metrics.

TABLE II
PER-GENRE CLASSIFICATION PERFORMANCE

Genre	Precision	Recall	F1-score	Support
Country	0.47	0.57	0.51	2000
Misc	0.74	0.80	0.77	2000
Pop	0.41	0.28	0.33	2000
Rap	0.76	0.82	0.78	2000
R&B	0.53	0.57	0.55	2000
Rock	0.40	0.35	0.37	2000
Accuracy	0.56 (12000 samples)			
Macro Avg	0.55	0.56	0.55	12000
Weighted Avg	0.55	0.56	0.55	12000

Most likely, model performance on the Rap genre were much higher because of its linguistically diverse and distinctive lyrics, often with high-level vocabulary and thematic patterns. In contrast, lyrics present in Pop and Rock genres are much more similar and broad, which could've attributed to the poor performance of the model in classifying and differentiating between the two alongside other genres.

V. CONCLUSION

While this shows the feasibility of using topic modeling on lyrics to classify song genres, it also shows how limited that approach would be to simple models such as Logistic Regression, linear SVM, and Random Forest. The per-genre analysis shows that linguistically distinctive genres such as rap benefit the most from this approach, whereas pop and rock genres remain challenging due to shared lyrical themes and generic vocabulary.

Future work should aim to explore models better suited to linguistic nuance and non-linear structure, such as transformer-based text encoders like BERT-styled embeddings, hierarchical neural networks, or graph-based topic representations. Performance may also improve by incorporating stylistic features, contextual word embeddings, or multimodal signals such as audio features. These approaches could likely capture subtle nuances in genre distinctions that are limited to the current approach.

REFERENCES

[1] A. Bansal, R. Agarwal, and K. Jain, "Linguistic Complexity and Socio-cultural Patterns in Hip-Hop Lyrics," arXiv preprint arXiv:2505.00035, 2025.

[2] C. G. de Castro Junior, "Genius Song Lyrics with Language Information," Kaggle, Dataset, 2022. Available: https://www.kaggle.com/datasets/carlosgdcj/genius-song-lyrics-with-language-information

[3] D. Chen, A. Satish, R. Khanbayov, C. M. Schuster, and G. Groh, "Beats of Bias: Analyzing Lyrics with Topic Modeling and Gender Bias Measurements," in Proc. SBP-BRiMS, 2024.

[4] M. Grachten, F. Gouyon, and P. Herrera, "Natural language processing of lyrics," in Proc. 5th Int. Conf. Music Information Retrieval (ISMIR), 2004.

[5] M. A. Mohammed, A. H. Ali, and A. Alsaedi, "Sentiment Analysis and Lyrics Theme Recognition of Music Lyrics Based on Natural Language Processing," Journal of Engineering Science, vol. 17, no. 2, pp. 135–145, 2023

[6] O. P. Suthar, A. Mishra, and S. Singhal, "Exploring the Landscape of Natural Language Processing for Text Analytics: A comprehensive Review," Procedia Comput Sci, vol. 259, pp. 453–462, Jan. 2025, doi: 10.1016/J.PROCS.2025.03.347