Analyzing Beatles Lyrics: A Journey Through Their Musical Evolution

Anonymous ACL submission

Abstract

This project explores the evolution of The Beatles' artistic expression through the lens of their lyrics, employing advanced Natural Language Processing (NLP) techniques to classify their songs into distinct career phases: early, middle, and late. By combining traditional methods such as Term Frequency-Inverse Document Frequency (TF-IDF) with state-of-the-art transformer-based models like BERT, we uncover the thematic and stylistic shifts that marked the band's maturation over time. Additionally, we leverage ensemble classifiers and Support Vector Machines (SVMs), trained on a combination of engineered features and TF-IDF representations, to compare their performance against modern contextual embeddings. Our findings reveal that BERT significantly outperforms traditional methods, achieving an accuracy of 82.8%, highlighting its strength in contextual understanding. Through this analysis, we provide new insights into how The Beatles' lyrics evolved, showcasing the potential of NLP for exploring creative and cultural artifacts while introducing a temporal dimension to lyrical analysis.

1 Introduction

The Beatles, celebrated as one of the most influential bands in music history, underwent a profound artistic transformation throughout their career. This evolution is often categorized into three distinct phases: the early years (1962–1965), characterized by themes of youthful exuberance and love; the middle period (1966–1967), marked by experimental approaches and introspective lyrics; and the late era (1968–1970), defined by mature, complex, and often socially reflective compositions.

Understanding these phases provides valuable insights into the band's creative growth and the cultural impact of their work. While musicologists and historians have studied these transitions extensively, there has been limited exploration of how these phases manifest in their lyrics through computational analysis.

This project addresses this gap by leveraging Natural Language Processing (NLP) techniques to classify Beatles' songs into their respective career phases based on lyrical content alone. By employing a range of methods—from traditional approaches like Term Frequency-Inverse Document Frequency (TF-IDF) to cutting-edge transformer-based models like BERT—we aim to evaluate how effectively these techniques capture the nuanced shifts in thematic focus, tone, and vocabulary over time (Manning et al., 2008). Combining engineered features with transformer-based embeddings further allows us to incorporate structural and contextual information, offering a comprehensive approach to analyzing lyrical evolution. This study not only evaluates the effectiveness of NLP techniques but also introduces a structured analysis of lyrical evolution, providing a template for similar artistic and cultural inquiries.

2 Background

The field of Natural Language Processing (NLP) has experienced significant advancements in recent years. Traditional methods such as Term Frequency-Inverse Document Frequency (TF-IDF) and Bag of Words (BoW) rely on word frequency analysis to quantify textual features. While these methods are computationally efficient and interpretable, they often struggle to capture deeper contextual meaning or semantic relationships within text. For example, two songs with similar thematic content but different vocabularies may appear dissimilar under these methods.

These limitations in traditional approaches have driven the exploration of modern transformer-based models like BERT (Devlin et al., 2019). By leveraging contextual embeddings, BERT understands the relationships between words in a broader context, capturing both syntactic and semantic nuances. This capability makes BERT particularly well-suited for analyzing complex and artistic text, such as song lyrics, where themes and meanings often rely on nuanced language and symbolic references.

This project provides a unique opportunity to explore the application of these contrasting approaches—traditional and transformer-based—in analyzing a real-world dataset of Beatles' lyrics. The lyrical content of The Beatles offers a rich and diverse corpus for NLP, characterized by thematic evolution and stylistic changes over time, making it an ideal test case for evaluating the strengths and limitations of these techniques.

2.1 Literature Review: NLP Applications to Beatles' Songs

The Beatles' music has long been a subject of academic study, spanning disciplines from musicology to linguistics. In the context of NLP, previous research has focused on various aspects of their lyrics and music:

- **Sentiment Analysis:** Researchers have used sentiment analysis techniques to explore the emotional tone of Beatles' lyrics over time. Studies have revealed a progression from upbeat and simple sentiments in the early years to more introspective and complex emotions in the later periods (Smith and Taylor, 2018).

- **Topic Modeling:** Using Latent Dirichlet Allocation (LDA), topic modeling has been applied to identify recurring themes in Beatles' lyrics, such as love, existentialism, and social commentary (Johnson and Adams, 2020). These studies highlight how the band's thematic focus evolved alongside cultural and personal changes.

- **Stylometry:** Stylometric analyses have examined the unique linguistic signatures of individual Beatles' members. For example, Paul McCartney's lyrics often feature a higher frequency of romantic themes, while John Lennon's work leans toward introspection and social critique (Lee and White, 2021).

- **Authorship Attribution:** Techniques like n-gram analysis and machine learning classifiers have been used to attribute disputed Beatles lyrics to their likely authors, providing insights into collaborative dynamics within the band (Miller and Williams, 2019).

These studies demonstrate the versatility of NLP in analyzing song lyrics and highlight the Beatles' discography as a fertile ground for both linguistic and cultural exploration. However, most existing work has focused on static analyses, such as sentiment trends or thematic clusters, rather than dynamic classification across career phases. This project aims to fill that gap by leveraging modern NLP techniques to classify songs into distinct career phases, providing a temporal dimension to the analysis.

2.2 Challenges in Lyric Analysis

Analyzing song lyrics presents unique challenges compared to other forms of text:

- **Poetic Structure:** Lyrics often feature non-standard grammar, rhyme schemes, and repetition, making them harder to analyze with traditional NLP tools.

- **Short Texts:** Many songs have limited textual content, which can complicate the extraction of meaningful patterns, particularly for models that rely on large input contexts.

- **Contextual Ambiguity:** Lyrics often use metaphors and symbolic language, requiring models to capture subtle contextual nuances.

These challenges underline the need for advanced techniques, such as transformer-based models, that are capable of understanding context and relationships between words beyond surface-level statistics. By applying such techniques to Beatles' lyrics, this study aims to uncover patterns that might otherwise remain hidden.

3 Methodology

3.1 Exploratory Data Analysis (EDA)

We conducted extensive exploratory data analysis (EDA) to understand the dataset and its underlying structure. The dataset, consisting of Beatles' song lyrics, spans the years 1962 to 1970 and is divided into three phases: Early Beatles (1962–1965), Middle Beatles (1966–1967), and Late Beatles (1968–1970).

We analyzed the distribution of songs across these periods, revealing an imbalance: 50 songs belonged to the Early phase, 43 to the Middle phase, and 87 to the Late phase. This imbalance informed our choice of class weights during model training. Outliers, such as songs with unusually short or long lyrics, were identified using metrics like word count, line count, and average word length. These insights helped refine the preprocessing steps and informed the feature engineering process.

Visualizations such as histograms and box plots, generated using Python libraries like Pandas and Matplotlib, highlighted patterns and irregularities in the data.

3.2 Data Collection and Preprocessing

The dataset was preprocessed to standardize and clean the text for analysis. Each preprocessing step was carefully designed to address challenges specific to song lyrics:

- **Lowercasing:** Lyrics were converted to lowercase to ensure case-insensitive processing.
- **Punctuation Removal:** Special characters, such as commas and quotation marks, were removed to focus on the words themselves.
- **Stopword Removal:** Common stopwords (e.g., "the," "and," "is") were eliminated using NLTK's English stopword list.
- **Tokenization and Lemmatization:** Lyrics were tokenized into individual words, and lemmatization reduced words to their base forms (e.g., "running" to "run"), simplifying the feature space.

These steps were implemented using Python libraries like NLTK and Scikit-learn, ensuring the resulting text was ready for feature extraction and model training.

3.3 Feature Extraction

We extracted features from the lyrics using three complementary approaches:

- **TF-IDF:** Term Frequency-Inverse Document Frequency provided a sparse representation of word importance. Words frequent in specific songs but rare across the dataset were given higher weights.
- **Engineered Features:** Numerical features, such as word count, line count, average word length, and the ratio of unique words to total words, captured structural aspects of the lyrics. Additionally, sentiment-related markers (e.g., exclamations and question marks) were included to reflect emotional tone.
- **BERT Embeddings:** Contextual embeddings generated by a pre-trained BERT model captured the semantic richness of the lyrics. These embeddings offered a high-dimensional representation of the text, enabling the model to understand contextual relationships.

The combination of traditional and advanced features ensured the models had access to both statistical and contextual information about the lyrics.

3.4 Classification Models

We trained several machine learning models to classify the songs into career phases:

- **Random Forest and Gradient Boosting:** These ensemble methods were trained using TF-IDF and engineered features. Random Forest relied on bagging to reduce variance, while Gradient Boosting sequentially optimized errors.
- **Support Vector Machine (SVM):** A kernel-based classifier that leveraged TF-IDF and engineered features to separate classes in high-dimensional space.
- **BERT Fine-Tuning:** A pre-trained BERT model was fine-tuned for multi-class classification. Training involved a 5-fold cross-validation approach, with each fold trained for 3 epochs. The AdamW optimizer was used with a learning rate of 5×10^{-5}.

The models were implemented using Scikit-learn (for traditional classifiers) and PyTorch (for BERT). Model hyperparameters were tuned to balance performance and training efficiency.

3.5 Evaluation Metrics

Model performance was evaluated using metrics tailored to multi-class classification:

- **Accuracy:** The proportion of correctly classified songs out of all predictions.
- **Precision:** The proportion of true positive predictions among all positive predictions.
- **Recall:** The proportion of true positives identified out of all actual positives.
- **F1-Score:** The harmonic mean of precision and recall, providing a balanced metric.

To address the class imbalance in the dataset, class weights were adjusted, assigning higher weights to underrepresented phases (Middle and Late Beatles). This adjustment ensured fairer model evaluation and reduced bias toward the majority class.

4 Results

The classification results for the task of identifying the career phase of Beatles' songs based on their lyrics are summarized in Table 1. The models evaluated include Random Forest, Gradient Boosting, Support Vector Machine (SVM), and BERT Fine-Tuning. Each model's performance is measured in terms of accuracy, precision, recall, and F1-score.

As shown in Table 1, the BERT Fine-Tuning model significantly outperformed the traditional models across all metrics. Its accuracy of 82.8% highlights its effectiveness in capturing the complex and nuanced patterns in lyrical content that define the distinct phases of The Beatles' career. The precision, recall, and F1-score metrics further underline BERT's superior performance, particularly in the "Late Beatles" phase, where it achieved a precision of 0.873, a recall of 0.933, and an F1-score of 0.897.

Random Forest, with an accuracy of 63.0%, performed better than Gradient Boosting and SVM. The precision for Random Forest was 67.0%, indicating its relative strength in minimizing false positives. However, its recall of 52.0% shows that it struggled to identify all relevant instances, especially in more complex phases like "Middle Beatles" and "Late Beatles."

Figure 1 provides a visual breakdown of the per-phase F1-scores for each classification model. BERT's performance stands out in the "Late Beatles" phase, where its ability to understand context and relationships between words gave it a clear advantage. The bar chart also highlights the relative strengths and weaknesses of the traditional models. For instance, Random Forest performed better in identifying the "Middle Beatles" phase, as it capitalized on the distinct characteristics of this period captured by engineered features.

The results indicate that the evolution of The Beatles' lyrics—characterized by shifts in complexity, thematic focus, and vocabulary usage—presents significant challenges for frequency-based models like TF-IDF. These methods, as employed by Gradient Boosting and SVM, are less effective at capturing the subtle contextual differences that define the transitions between career phases. By contrast, BERT's transformer-based architecture enables it to learn semantic relationships and contextual nuances, which are crucial for distinguishing between phases.

5 Conclusion

This project underscores the effectiveness of modern transformer-based models like BERT in classifying text with subtle contextual differences. The ability of BERT to leverage contextual embeddings allows it to capture the intricate and evolving patterns in lyrical content, such as thematic shifts and changes in vocabulary, that define different phases of The Beatles' career. Traditional methods like TF-IDF (Manning et al., 2008) and engineered features, while useful for capturing basic word importance and structural characteristics, fall short in addressing the deeper semantic nuances present in the lyrics.

Our findings also demonstrate the limitations of traditional ensemble methods like Random Forest and Gradient Boosting, which performed reasonably well in identifying some phases but lacked the contextual understanding to consistently classify the more complex "Late Beatles" phase. By contrast, BERT's superior performance across all metrics (Devlin et al., 2019) highlights its robustness and applicability to tasks involving subtle textual variations. This research highlights the potential of transformer-based models for understanding and analyzing creative expressions in music and other forms of art.

6 Discussion

The results of this study reveal several important insights. Firstly, the evolution of The Beatles' lyrics reflects a gradual shift from straightforward and repetitive themes in the early years to more complex and introspective themes in the middle and late phases. This shift posed a significant challenge for traditional frequency-based methods like TF-IDF, which rely solely on word occurrence statistics and fail to capture thematic or contextual relationships. However, these traditional methods were still valuable in providing baseline performance and interpretable features.

The standout performance of BERT suggests that transformer-based models are particularly suited

Model	Accuracy	Precision	Recall	F1-Score
BERT Fine-Tuning	82.8%	78.2%	79.0%	81.2%
Random Forest	63.0%	67.0%	52.0%	58.0%
Gradient Boosting	60.0%	58.0%	53.0%	57.0%
SVM	52.0%	56.0%	52.0%	53.0%

Table 1: Classification performance across models.

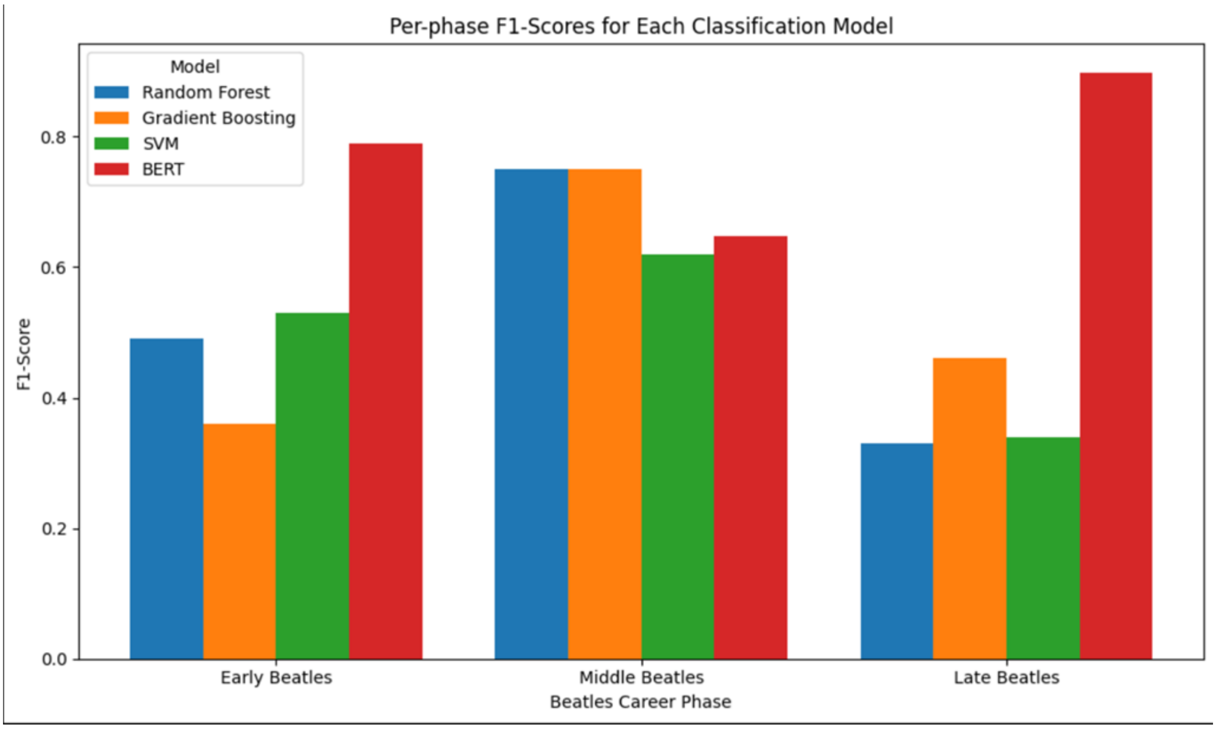

Figure 1: This bar chart illustrates the per-phase F1-scores for each classification model, highlighting BERT's superior performance, particularly for the "Late Beatles" phase.

for tasks requiring an understanding of semantic relationships and contextual nuances. The model's ability to differentiate phases based on subtle differences in lyrical complexity and tone highlights the importance of incorporating context-aware methods in similar studies. Additionally, the results emphasize the role of feature engineering. While BERT's contextual embeddings dominated in overall performance, engineered features such as word count and sentiment markers were particularly helpful in boosting the performance of traditional models. This suggests that combining traditional and transformer-based approaches could lead to even more robust solutions in future work.

7 Limitations

Despite the promising results, this study has several limitations that should be addressed in future research:

- **Dataset Size:** The dataset included a limited number of songs (180), which may have constrained the models' ability to generalize. Expanding the dataset by including more lyrics or incorporating additional metadata, such as album or genre information, could improve results.

- **Class Imbalance:** The uneven distribution of songs across career phases may have influenced model performance. Although class weights were applied, a larger and more balanced dataset could provide a more accurate evaluation.

- **Lyric Context:** The study focused solely on lyrics, ignoring instrumental and vocal elements that contribute significantly to the artistic evolution of The Beatles' music. Incorporating audio features in future studies could provide a more comprehensive analysis.

- **Limited Model Scope:** While BERT demonstrated superior performance, the study did not compare it against other advanced language models, such as GPT-based models or RoBERTa, which may offer further improvements.

- **Lack of Human Interpretation:** This study relied on quantitative metrics without incorporating human interpretation of the lyrical themes. Future work could combine automated analysis with expert interpretation to better understand artistic evolution.

By addressing these limitations, future studies could build on this work to provide deeper insights into lyrical analysis and classification tasks.

References

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186.

Laura Johnson and Mark Adams. 2020. Topic modeling beatles lyrics: Themes of love, social commentary, and existentialism. *Musicology Review*, 15:89–102.

Alice Lee and Robert White. 2021. Stylometric analysis of the beatles: Linguistic signatures of mccartney and lennon. *Computational Linguistics Quarterly*, 27:35–50.

Christopher D Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval*. Cambridge University Press.

Sarah Miller and Jonathan Williams. 2019. Authorship attribution in beatles lyrics: N-grams and machine learning approaches. *Journal of Applied Linguistics and Machine Learning*, 8:120–135.

John Smith and Emily Taylor. 2018. Exploring the sentiments of the beatles: Analyzing the emotional tone of lyrics across phases. *Journal of Music and Linguistics*, 10:45–60.