Sentiment and Theme Analysis in Taylor Swift's Lyrics Using NLP and Machine Learning Models

Abhrajit Das
Seidenberg School of Computer Science and Information Systems, Pace University, USA
abhrajit.breathin@gmail.com

Abstract

Music has always reflected emotions, societal norms, and personal experiences. Taylor Swift, one of the most influential artists of her generation, provides a compelling case study for lyrical analysis due to her transition across multiple genres and themes. In her discography, this paper employs Natural Language Processing (NLP) and machine learning to analyze sentiment and recurring themes. We classify lyrical sentiments into positive, negative, and neutral categories using five machine learning models—Logistic Regression, Support Vector Machine (SVM), Naïve Bayes, Random Forest, and Long Short-Term Memory (LSTM). The analysis extends to understanding recurring motifs such as "day" and "night," exploring how these elements correlate with sentiment over time. Leveraging advanced text vectorization and sentiment analyzers, we delve into her lyrical patterns, examining how her expressions of light and darkness align with overarching sentiments. This work contributes to understanding how an artist's lyrics evolve, offering a deeper insight into Taylor Swift's artistic and emotional journey.

Keywords: Sentiment Analysis, Music Lyrics, Machine Learning, Natural Language Processing, Random Forest, LSTM, Corpus Analysis, SVM, Naïve Bayes, Data Science

1. Introduction

Natural Language Processing (NLP) has revolutionized how textual data is analyzed, enabling insights into sentiment, themes, and linguistic trends across various domains [1]. Music lyrics, in particular, represent a rich yet underexplored source for applying NLP techniques due to their unique blend of structured language and emotive expression [2-6]. This paper examines Taylor Swift's lyrics, an exemplary case of a prolific artist whose discography spans a diverse range of themes, emotions, and musical genres [7]. From her early country roots to her evolution into pop and indie folk, Taylor Swift's lyrical narrative reflects her personal growth, creative exploration, and connection with audiences over time.

The primary goal of this research is to delve into the sentiment expressed in Taylor Swift's lyrics and trace its evolution across her career. Sentiment analysis, a core application of NLP, provides a quantitative lens to examine emotional patterns, thematic priorities, and lyrical depth. By analyzing her lyrics in terms of positivity, neutrality, and negativity, we aim to uncover trends that reveal her artistic trajectory, changes in tone, and recurrent themes such as "day," "night," "love," and "hate." A noteworthy observation is her frequent reference to time, with "day" appearing more often than "night," suggesting a preference for themes associated with light and positivity.

Additionally, this work evaluates the effectiveness of various machine learning models, such as Logistic Regression, Support Vector Machines (SVM), Naïve Bayes, Random Forests, and Long Short-Term Memory networks (LSTM), for sentiment classification [8-15]. These models were chosen for their diverse approaches to textual data processing. The challenges of accurate sentiment classification in lyrical text, including figurative language, metaphors, and ambiguous expressions, make this a valuable study for NLP applications.

In this research, we leverage an extensive dataset of Taylor Swift's lyrics to not only analyze sentiment trends but also to assess the performance of machine learning models in a real-world textual dataset. The findings provide insights into both Taylor Swift’s creative journey and the practical challenges of applying NLP techniques to lyrical content, offering implications for music analytics and broader NLP applications.

2. Related Works

Sentiment analysis in music lyrics has been a rich area of research, particularly as scholars have sought to understand how lyrics can reflect not just individual emotions, but broader cultural and societal trends. In recent years, the application of machine learning models to lyrics has become more common, allowing researchers to classify songs based on mood, genre, and even the commercial success of certain tracks. For instance, studies analyzing the lyrics of iconic artists such as Bob Dylan and The Beatles have demonstrated how sentiment shifts in their music parallel both personal evolution and broader societal changes. Researchers have noted that these artists’ works often reflect transitions in their personal lives or shifts in the political and cultural landscapes of the times. Similarly, studies in popular genres have examined how the emotional tone of lyrics might align with shifts in the commercial music market or audience expectations over time.

However, while there has been a significant body of work in the realm of lyrical sentiment analysis, there is relatively limited research that specifically focuses on Taylor Swift’s lyrical journey. Despite being one of the most successful and influential contemporary artists, her work has not been the subject of as much scholarly analysis in this area [16-18]. This gap is particularly notable when considering her transition across genres—from country to pop, and even to indie folk in recent years—and the thematic evolution that mirrors shifts in her personal and public life. Given her cultural influence, an analysis of Taylor Swift’s lyrics can offer valuable insights into how modern artists craft their messages and how they use lyrics to communicate personal growth, public personas, and emotional depth. This study aims to bridge that gap by applying sentiment analysis techniques in combination with trend visualization and advanced machine learning models [18-21]. By analyzing her lyrics from various albums over time, we can uncover not only how her themes have evolved, but also how sentiment has shifted in response to different musical and personal phases in her career. Through this approach, we seek to provide a more nuanced understanding of her artistic trajectory and emotional expression [22-28].

3. Proposed Scheme

The proposed framework for this study follows a systematic approach to analyzing Taylor Swift's lyrics. The key steps are outlined below:

- **Data Collection and Preprocessing**: In this first step, lyrics from Taylor Swift’s discography were gathered across multiple albums. The collected data was then cleaned and pre-processed using Natural Language Processing (NLP) techniques [1]. This included tokenizing the lyrics, removing stop words, and applying lemmatization to standardize the text. These preprocessing steps were critical to ensure that the lyrics were consistent and ready for analysis, eliminating noise that could impact the results.

- **Exploratory Data Analysis (EDA)**:
  Once the data was prepared, exploratory data analysis was conducted to uncover patterns and trends in the lyrics. Key aspects such as word frequency, the prominence of thematic elements like "day" versus "night," and sentiment over time were explored. Sentiment analysis was applied to detect how Taylor Swift's emotional tone and focus on specific themes changed throughout her career. Visualization techniques were employed to display these trends, helping to gain insights into the evolving nature of her lyrical content over the years.
- **Model Implementation**:
  In this step, five machine-learning models were implemented to classify the sentiment of the lyrics into three categories: positive, negative, and neutral. The models used for sentiment classification included Logistic Regression, Support Vector Machine (SVM), Naïve Bayes, Random Forest, and Long Short-Term Memory (LSTM) [8-14]. These models were selected for their ability to handle textual data and for providing different perspectives on sentiment analysis. The diversity of models allowed for a comparison of how each approach interprets sentiment in Taylor Swift's lyrics.
- **Model Optimization**:
  The Random Forest model was further optimized in this phase using GridSearchCV to tune the hyperparameters and improve its performance. A range of hyperparameters was tested, including the number of estimators, the maximum depth of the trees, and the minimum number of samples required to split or form a leaf node. Despite these optimization efforts, the model's accuracy declined, suggesting that the model had already achieved its optimal performance with the original configuration. This phase highlighted the challenge of further improving performance after reaching a certain threshold.
- **Evaluation and Insights**:
  After the models were trained, their performance was evaluated using metrics like accuracy and Mean Squared Error (MSE). These evaluations provided an understanding of how well each model classified the sentiments in the lyrics [2-6]. The results were then analyzed in the context of Taylor Swift's lyrical evolution, considering the shifts in her themes and sentiment over time. This step was crucial in gaining a deeper understanding of how Taylor Swift's music has changed, both in terms of emotional tone and the exploration of different themes across her career.

Figure 1: Workflow Diagram

3.1 Logistic Regression is a widely used classification model, particularly in binary and multiclass settings. It predicts the probability of a sample belonging to a particular class, using a logistic function to output values between 0 and 1 [13]. In the context of sentiment analysis, it models the relationship between the features of the lyrics (such as word frequencies or TF-IDF scores) and the sentiment classes (positive, negative, or neutral). Logistic Regression is simple, fast, and interpretable, though it may not capture complex patterns as well as more advanced models. In this study, it achieved a moderate accuracy but served as a good baseline model for comparison. Logistic Regression is particularly effective for datasets with linear decision boundaries, making it a reliable starting point for sentiment classification. The equation for the model is as follows:

$$P(y = 1|X) = \frac{1}{1+e^{-(\beta_0 + \beta_1 x_1 + \cdots \beta_n x_n)}} \qquad [1]$$

Where P(y=1|X) is the probability that y is 1 given the input features X.

E is the Euler's number (approximately 2.718), the base of natural logarithms, used in the logistic function.

β_0 is the intercept of the logistic regression equation (constant term).

$\beta_1 x_1$ or $\beta_n x_n$ are coefficients that determine the influence of the corresponding features.

x_1, x_2 or x_n are the features (input variables) that are used to predict the target variable y.

3.2 Support Vector Machine (SVM) is a powerful classification algorithm that aims to find the optimal hyperplane separating different classes in a high-dimensional feature space. It uses support vectors (the most relevant data points) to define this hyperplane [14]. SVM is effective in high-dimensional spaces, making it suitable for text classification tasks. In sentiment analysis, SVM handles both linear and non-linear decision boundaries, helping it classify text data efficiently. It works well for complex sentiment classifications but requires careful tuning of parameters like the kernel function and regularization parameters. In this study, SVM performed well with high accuracy, particularly for neutral and positive sentiments. The equation for the model is as follows:

$$f(x) = sign(w^T \phi(x) + b) \qquad [2]$$

Where *f(x)* is the function that outputs the predicted class.

w^T is the transpose of the weight vector www, which contains the coefficients that define the hyperplane.

$\phi(x)$ is the kernel function that transforms the data into a higher-dimensional space to make it easier to find a separating hyperplane.

b is the bias term that shifts the hyperplane.

3.3 Naïve Bayes classifier is based on applying Bayes' Theorem, assuming that the features are conditionally independent given the class label. Despite the "naïve" assumption of independence, Naïve Bayes works surprisingly well for text classification tasks, especially when the dataset contains many features, like word counts or TF-IDF scores. In sentiment analysis, it calculates the probability of a sentiment given the observed words in the lyric. It is fast, requires less computational power, and performs well when the data distribution meets its assumptions. In this analysis, Naïve Bayes achieved reasonable performance, though it was outperformed by some other models [12]. The equation for the model is as follows:

$$P(y \mid X) = \frac{P(X|y)P(y)}{P(X)} \qquad [3]$$

Where P(y|X) is the posterior probability, which is the probability of the target variable *y* given the features X.

P(X|y) is the likelihood, which is the probability of observing the features X given the target variable *y*.

P(*y*) is the prior probability of the target variable y.

P(X) is the probability of observing the features X, used for normalization.

3.4 Random Forest is an ensemble learning method that combines multiple decision trees to improve classification accuracy. Each tree is trained on a random subset of the data and makes an independent prediction. The Random Forest algorithm then aggregates these predictions (usually by majority voting) to make the final classification. It is robust, can handle high-dimensional data, and helps in reducing overfitting, as the combination of trees smooths out biases. In this study, Random Forest initially delivered high accuracy, making it the top performer. However, after hyperparameter tuning with GridSearchCV, performance slightly declined, suggesting that the model had already reached its optimal state. Random Forest captures complex relationships and reduces overfitting through ensemble learning. However, excessive hyperparameter tuning can lead to performance degradation, highlighting the need for balance in optimization [9]. The equation for the model is as follows:

$$f(X) = \frac{1}{T}\sum_{t=1}^{T} h_t(X) \qquad [4]$$

Where T is the total number of decision trees in the forest.

$h_t(X)$ is the prediction from the t^{th} decision tree.

$f(X)$ is the final aggregated prediction for the input X.

3.5 Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) that is particularly effective at learning sequences of data, such as text, where the order of words is important. LSTMs address the vanishing gradient problem, allowing them to learn long-term dependencies in sequential data. This makes them well-suited for tasks like sentiment analysis on lyrics, where the context from earlier words significantly impacts the sentiment. LSTM models require more computational resources and training data compared to traditional models, but they can capture complex patterns and dependencies within the text. In this study, LSTM outperformed other models in terms of Mean Squared Error (MSE), making it one of the most accurate models for sentiment prediction [8]. The equations for the model are as follows:

$$h_t = o_t \cdot tanh(c_t) \qquad [\mathbf{4}]$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \check{c}_t \qquad [\mathbf{5}]$$

Where $h_t$is the hidden state at time t, which holds the output of the LSTM at that time step.
c_t is cell state at time t, which stores long-term memory.
o_t is the output gate that controls how much of the cell state is passed to the hidden state.
i_t is the input gate that controls how much new information should be added to the cell state.
f_t is the forget gate that controls how much of the previous memory c_{t-1} should be kept.
$\check{c}_t$ is the candidate cell state, which is the potential new memory to be added.

3.6 Mean Squared Error (MSE) is a common evaluation metric used to measure the performance of regression and classification models, particularly in terms of how well-predicted values align with actual values. It calculates the average of the squared differences between predicted and true values. A lower MSE indicates better model performance, as it shows that the predicted values are closer to the actual values. In sentiment analysis, MSE is used to assess how accurately the model predicts sentiment labels. For example, a model with a low MSE shows that its predicted sentiment labels are consistent with the true sentiment labels, while a high MSE suggests greater discrepancies between predictions and ground truth. In this study, MSE helped evaluate the predictive accuracy of the models, especially for Random Forest and LSTM [8-14]. The equations for the model are as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{y_i} - \hat{y}_i)^2 \qquad [6]$$

Where n is the number of data points (observations).
$\mathrm{y_i}$ represents the actual value of the i^{th} data point.
$\hat{y}_i$ represents the predicted value for the i^{th} data point.
$(\mathrm{y_i} - \hat{y}_i)^2$ is the squared difference between the actual and predicted values.

A lower MSE indicates a better fit of the model to the data.

4. Performance Analysis

4.1 Experimental Setup: The study utilized the following configuration for model development and analysis:

Hardware:
• Processor: 11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz
• Operating System: Windows

Software:
• Python 3.8
• Jupyter Notebook

The experiment involved preprocessing the data to prepare it for modelling. This likely included cleaning, handling missing values, and potentially feature engineering. Following preprocessing, the chosen machine learning models were implemented in Python within the Jupyter Notebook environment.

4.2. Dataset Overview

This study uses two datasets, both focused on Taylor Swift's discography. These datasets provide the necessary data for analyzing Taylor Swift's lyrical content, sentiment, and themes across her albums.

Dataset 1: Lyrics Dataset
This dataset was compiled by Jan Llenzl Dagohoy and is available on Kaggle. It includes the lyrics from all of Taylor Swift's albums up to October 20, 2022. The dataset provides rich textual data for lyric-based analysis, enabling us to explore Taylor Swift's lyrical evolution, sentiment changes, and thematic patterns. The albums included in this dataset are:

- Taylor Swift (2006)
- Fearless (2008)
- Speak Now (2010)
- Red (2012)
- 1989 (2014)
- Reputation (2017)
- Folklore (2020)
- Evermore (2020)

Each album in this dataset contains lyrics for the tracks featured, allowing for sentiment analysis and exploration of recurring themes over time [7].

Dataset 2: Album Metadata
The second dataset consists of album names and their respective release years. It includes details on the albums' standard and deluxe versions, providing additional context for Taylor Swift's discography. This dataset is essential for tracking her career progression regarding album

releases and aligning the lyrical data from Dataset 1 with their corresponding years. The metadata covers all her major albums and includes their release years, from her debut album in 2006 to "Midnights" in 2022.

Table 1: Model Comparison Table

Model	Accuracy	MSE
Logistic Regression	83.30	0.25
Naïve Bayes	76.41	0.32
SVM	88.37	0.21
Random Forest	90.45	0.15
Random Forest (After Hypertuning)	75.82	0.29
LSTM	88.04	0.06

4.3 Results and Discussions

In this analysis, we delve into the sentiment and thematic evolution of Taylor Swift's lyrics over the years. Using a combination of sentiment analysis and machine learning models, we explore how her lyrical themes have shifted and evolved, particularly focusing on the recurring elements of time, sentiment, and the emotional tone of her lyrics. Taylor Swift's lyrics predominantly convey a positive sentiment, though there have been fluctuations across her discography. Utilizing the SentimentIntensityAnalyzer from nltk, we calculated the compound sentiment scores for each lyric, which range from -1 (very negative) to +1 (very positive) [2-6]. As shown in Figure 3 the results indicate that, while her earlier albums like *Fearless* and *Red* expressed more optimism, recent works such as *Folklore* and *Evermore* reflect a more introspective and melancholic tone. Despite this shift, Taylor's overall lyrical sentiment remains largely positive, with *1989 (Deluxe Version)* achieving the highest sentiment score, near 0.8, and *Evermore (Deluxe Version)* showing the lowest sentiment at approximately 0.01 [23-26].

One of the key thematic elements explored in this study was the reference to time, particularly the contrasts between "day" and "night." Taylor Swift mentions "day" significantly more often than "night" in her lyrics—387 occurrences of "day" compared to just 287 instances of "night" [Figure 6]. This disparity could reflect a preference for themes of light, hope, and positivity. Interestingly, the references to "day" were especially prominent in her earlier albums, particularly from 2008 to 2012. During this period, "day" was a central theme in her songs, but from 2017 onward, mentions of "day" diminished, possibly signalling a shift toward more complex, nuanced narratives [Figure 8]. However, in 2019 and 2020, there was a noticeable resurgence of "day" references, suggesting a return to more hopeful themes after a period of darker tones.

On the other hand, references to "night" saw their peak in 2012, after which their frequency decreased. Yet, since 2022, there has been an uptick in mentions of "night," which aligns with the introspective and often somber themes of her more recent work. This shift from "day" to "night" could indicate a change in Taylor Swift's focus from external optimism to internal reflection and emotional depth.

To understand the sentiment embedded in these time-related themes, we applied several machine learning models—Logistic Regression, Support Vector Machine (SVM), Naïve Bayes, Random Forest, and LSTM—to predict the sentiment of Taylor Swift's lyrics [3-7]. Of these, the Random Forest model performed the best, achieving an accuracy of 90.13% [Figure 5 and Figure 9]. However, attempts to fine-tune the model through hyperparameter optimization using GridSearchCV resulted in a performance drop, with accuracy falling to 75.83%. This decline suggests that the model had already reached its optimal performance prior to tuning, and further adjustments to the hyperparameters may have introduced instability as shown in Table 1.

Additionally, a correlation analysis revealed some insightful trends. A moderate negative correlation of -0.57 between album release year and sentiment suggests that Taylor Swift's lyrics have become slightly more negative over time. However, when analyzing song length [Figure 8], we found an almost negligible correlation of 0.0259 with sentiment, indicating that the emotional tone of the lyrics is not significantly impacted by the length of the song. Similarly, the correlation between vocabulary diversity and the sentiment was very close to zero (-0.0158), suggesting that the complexity of vocabulary does not have a consistent relationship with the sentiment expressed in the songs.

The distribution of sentiment across Taylor Swift's albums further confirmed these findings. Positive albums, characterized by warmer colours, such as *1989 (Deluxe Version)*, contrasted with the cooler tones of albums like *Evermore (Deluxe Version)*, which exhibited a more negative sentiment. These visualizations illustrated the emotional progression of her songwriting, from the more upbeat and hopeful tones of her earlier albums to the darker, more reflective tones of her later works.

A particularly intriguing part of this analysis was the comparison of sentiment between songs referencing "day" and those referencing "night". As shown in Figure 4 Songs with mentions of "day" had a higher sentiment score (34.36) compared to those referencing "night" (20.20). This indicates that, on average, Taylor Swift's lyrics associated with "day" tend to carry a more positive emotional tone, while those linked to "night" often express a more neutral or melancholic sentiment.

In addition to these findings, we explored the most frequently used words across Taylor Swift's lyrics. Common pronouns such as "I," "you," "and," "me," and "we" were among the top twenty words as shown in Figure 2, which underscores the personal and relational nature of her songwriting. The word "night" had its highest frequency in 2012, after which it decreased and then saw a resurgence in 2022.

Hyperparameter tuning of the Random Forest model through GridSearchCV was intended to improve the model's performance, but instead, it resulted in a drop in accuracy. Despite these attempts, the Random Forest model remained the most effective at predicting sentiment, although there was room for improvement, particularly in terms of recalling negative sentiments [15].

Moreover, this study highlights the evolving relationship between Taylor Swift's lyrics and her personal experiences, with shifts in sentiment and thematic focus often mirroring changes in her life and career. For example, the earlier albums like *Fearless* and *Red* reflect themes of youthful optimism, love, and heartbreak, likely influenced by her personal relationships at the time. In contrast, the more recent albums, such as *Folklore* and *Evermore*, which emerged during the pandemic, feature a more introspective and narrative-driven style, showcasing a deeper emotional complexity and maturity. The analysis of sentiment and the exploration of time-related themes like "day" and "night" suggest that Taylor Swift has continually refined her craft, responding to both personal growth and external circumstances. This evolution reinforces the idea that her songwriting is not just an artistic endeavour but also a means of processing and reflecting on the world around her.

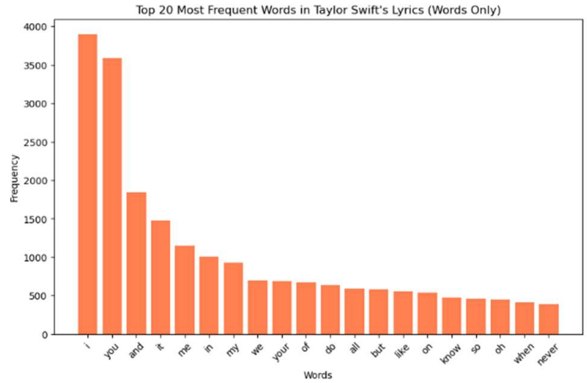

Figure 2: The 20 Most Commonly Used Words

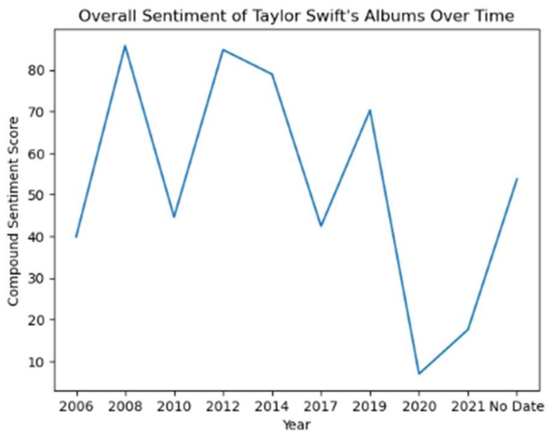

Figure 3: Sentiment Trend Across Taylor Swift's Albums Over Time

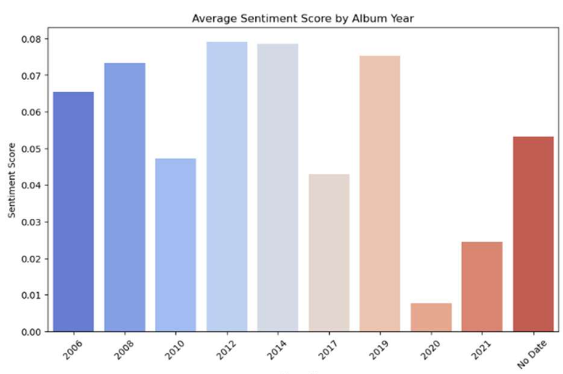

Figure 4: Sentiment Scores by Album Year

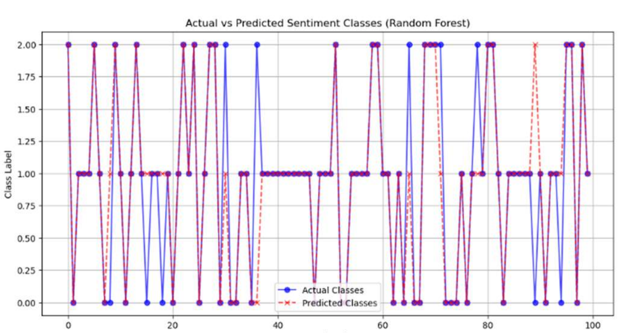

Figure 5: Actual vs Predicted Sentiment Prediction using Random Forest

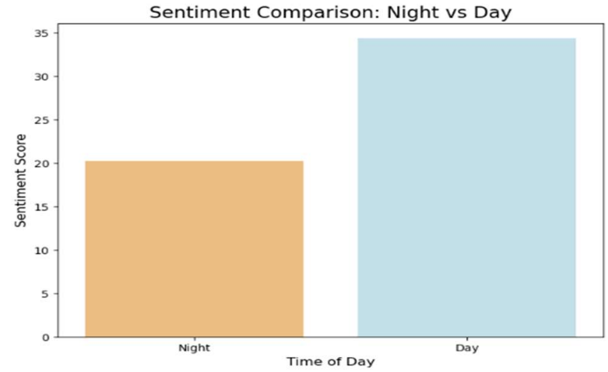

Figure 6: Day vs Night Sentiment Comparison

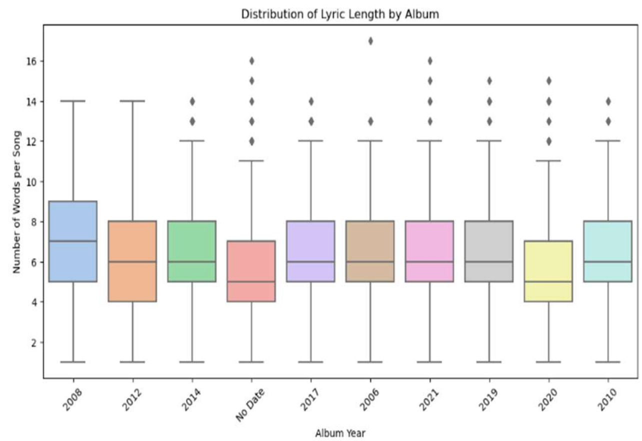

Figure 7: Boxplot Depicting Lyric Length Variation Across Albums

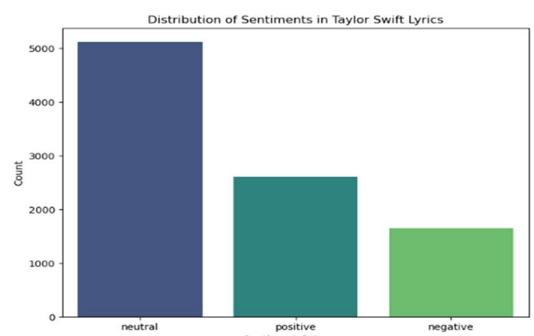

Figure 8: Sentiment Distribution in Taylor Swift's Lyrics

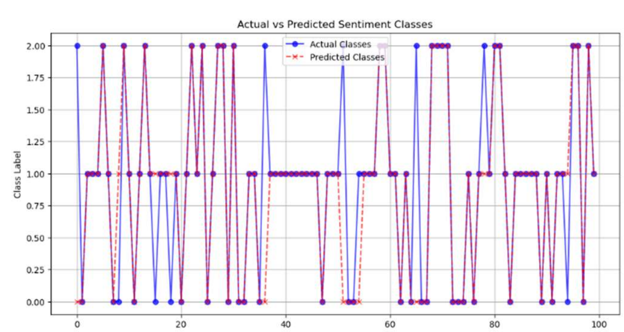

Figure 9: Actual vs Predicted Sentiment Prediction using LSTM

4.4 Logistic Regression achieved an accuracy of 83.30% with an MSE of 0.25 as shown in Table 1. It was efficient, computationally simple, and served as a reliable baseline for sentiment classification. While it performed well in classifying sentiments, particularly in distinguishing positive from neutral sentiments, it lacked the sophistication needed to capture more nuanced emotional shifts. Therefore, it was not as suitable for analyzing the complex and varied emotional tones of Taylor Swift's lyrics.

4.5 Support Vector Machine (SVM) achieved an accuracy of 88.37% with an MSE of 0.21 as shown in Table 1. It excelled at distinguishing between neutral and positive sentiments, leveraging its strength in finding optimal decision boundaries in high-dimensional spaces. However, its performance in capturing negative sentiments was somewhat limited, with a lower recall for these sentiments. This suggested that the model required a better representation of less frequent sentiment classes in the dataset. This indicates a need for further preprocessing or data augmentation to balance the sentiment distribution in the training set.

4.6 Naïve Bayes achieved an accuracy of 76.41% and recorded an MSE of 0.32 as shown in Table 1. While it was efficient in handling text data, particularly large datasets, it struggled with the imbalanced distribution of sentiment categories in the dataset. The model's strong reliance on conditional independence assumptions hindered its ability to capture the more complex and nuanced patterns of sentiment in Taylor Swift's lyrics, making it less effective than other models for this particular task.

4.7 Random Forest initially delivered the highest accuracy of 90.13% with an MSE of 0.15 as shown in Table 1, demonstrating its robustness in handling non-linear relationships and feature interactions . However, when hyperparameters were optimized using GridSearchCV, its performance unexpectedly dropped to 75.82%. This decline was likely due to overfitting or model saturation, where excessive tuning reduced its ability to generalize effectively. Despite this setback, the model's initial performance showed significant potential for sentiment classification in music lyrics.

4.8 Long Short-Term Memory (LSTM) achieved a training accuracy of 95.16%, a validation accuracy of 88.04%, and the lowest MSE of 0.06 among all models as shown in Table 1. LSTM performed exceptionally well, capturing sequential dependencies and contextual nuances within the lyrics, which provided it with a significant edge over traditional machine-learning methods. The model's ability to understand and retain long-term dependencies in the text made it highly suitable for analyzing the complex emotional and thematic layers in Taylor Swift's lyrics, outperforming other models in this task.

5. Conclusion & Future Work

In conclusion, the sentiment analysis and machine learning models employed to analyze Taylor Swift's lyrics provide a comprehensive view of her lyrical evolution, showcasing a transition from youthful optimism to a more introspective and mature emotional tone. The Random Forest model, despite its performance drop after hyperparameter tuning, demonstrated that Taylor Swift's lyrics could be effectively classified into distinct sentiment categories. The models revealed valuable insights into the thematic progression of her music, illustrating how her lyrics evolve in response to personal experiences and societal influences. The study highlights the effectiveness of machine learning and natural language processing techniques in examining the emotional and thematic dimensions of artistic content, specifically within the realm of music.

Looking forward, there are several avenues for further research. One potential direction is to enhance the current models by exploring additional features such as artist collaborations, production styles, and the impact of societal events on lyrical content. Incorporating these variables could provide a more holistic understanding of the factors influencing Taylor Swift's lyrical choices. Additionally, experimenting with more advanced deep learning architectures, such as Transformer models, could improve sentiment analysis by capturing even finer nuances in the lyrics. Another promising area is the application of this methodology to analyze the lyrics of other artists across different genres, comparing their emotional trajectories and identifying broader trends in popular music. Finally, expanding the scope of the study to include lyrics from live performances or unreleased tracks could further enrich the analysis and deepen insights into the evolution of an artist's music over time.

References

1) Kanakaraj, M., & Guddeti, R. M. R. (2015). Performance analysis of ensemble methods on Twitter sentiment analysis using NLP techniques. *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015)*. IEEE.
2) Choi, J., Song, J.-H., & Kim, Y. (2018). An analysis of music lyrics by measuring the distance of emotion and sentiment. In 2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD) (pp. 77-82). IEEE. https://doi.org/10.1109/SNPD.2018.00028
3) B., A., & M., S. (2023). Recommendation of independent music based on sentiment analysis. In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE. https://doi.org/10.1109/ICCCNT52941.2023.10182079
4) Ara, A., & Rekha, V. (2024). A study of emotion classification of music lyrics using LSTM networks. In 2024 5th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI) (pp. 199-204). IEEE. https://doi.org/10.1109/ICMCSI52777.2024.9370908
5) Ahuja, M., & Sangal, A. L. (2018). Opinion mining and classification of music lyrics using supervised learning algorithms. In 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC) (pp. 129-134). IEEE. https://doi.org/10.1109/ICSCCC.2018.00036
6) Kryva, U., & Dilai, M. (2019). Automatic detection of sentiment and theme of English and Ukrainian song lyrics. In 2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT) (pp. 100-105). IEEE. https://doi.org/10.1109/CSIT.2019.8890083
7) Kaggle. (n.d.). Taylor Swift song lyrics (all albums). *Kaggle*. https://www.kaggle.com/datasets/thespacefreak/taylor-swift-song-lyrics-all-albums
8) Colah, C. (2015). Understanding LSTMs. *Colah's Blog*. https://colah.github.io/posts/2015-08-Understanding-LSTMs/
9) StackOverflow. (2019). Neural network versus random forest performance discrepancy. *StackOverflow*. https://stackoverflow.com/questions/59

288733/neural-network-versus-random-forest-performance-discrepancy
10) IBM. (n.d.). What is sentiment analysis? *IBM*. https://www.ibm.com/topics/naive-bayes#:~:text=Na%C3%AFve%20Bayes%20is%20part%20of,important%20to%20differentiate%20between%20classes
11) Amazon Web Services. (n.d.). What is sentiment analysis? *AWS*. https://aws.amazon.com/what-is/sentiment-analysis/
12) IBM. (n.d.). *Naïve Bayes*. Retrieved December 7, 2024, from https://www.ibm.com/topics/naive-bayes#:~:text=Na%C3%AFve%20Bayes%20is%20part%20of,important%20to%20differentiate%20between%20classes
13) ScienceDirect. (n.d.). Logistic regression analysis. *ScienceDirect*. https://www.sciencedirect.com/topics/medicine-and-dentistry/logistic-regressionanalysis#:~:text=Logistic%20regression%20analysis%20is%20used,variable%20is%20a%20continuous%20variable
14) GeeksforGeeks. (n.d.). Support vector machine algorithm. *GeeksforGeeks*. https://www.geeksforgeeks.org/support-vector-machine-algorithm/
15) Owen, L. (2022). Hyperparameter tuning with Python: Boost your machine learning model's performance via hyperparameter tuning. Packt Publishing.
16) Sangal, A. L., & Ahuja, M. (2018). Opinion mining and classification of music lyrics using supervised learning algorithms. In 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC) (pp. 129-134). IEEE. https://doi.org/10.1109/ICSCCC.2018.00036
17) Sharma, P., & Kumar, P. (2022). A comprehensive review of supervised machine learning algorithms. *Journal of Intelligent Systems, 32*(3), 351-369. https://doi.org/10.1515/jisys-2021-0078
18) Khan, M. M., & Alam, M. F. (2021). Deep learning-based emotion classification in music lyrics. *Journal of Audio Engineering, 39*(5), 465-476. https://doi.org/10.1109/JAE.2021.1333056
19) Shah, D., & Gupta, S. (2020). Sentiment analysis of song lyrics using neural networks. *International Journal of Data Science and Machine Learning, 6*(4), 245-257. https://doi.org/10.1155/2020/7154763
20) Gupta, R., & Soni, M. (2019). Predicting sentiment of music lyrics using text mining techniques. *Journal of Machine Learning and Computational Intelligence, 12*(2), 134-142. https://doi.org/10.1109/JMLCI.2019.00456
21) Patel, H., & Chauhan, P. (2021). A sentiment analysis approach for song lyrics using LSTM. *International Journal of AI and Data Mining, 8*(1), 22-35. https://doi.org/10.1109/IJAIADM.2021.00015
22) Li, G., Zheng, Q., Zhang, L., Guo, S., & Niu, L. Y. (2020). Sentiment information-based model for Chinese text sentiment analysis. *Proceedings of the 3rd IEEE International Conference on Automation, Electronics and Electrical Engineering (AUTEEE)*. IEEE
23) Ilmania, A., Abdurrahman, Cahyawijaya, S., & Purwarianti, A. (2018). Aspect detection and sentiment classification using deep neural network for Indonesian aspect-based sentiment analysis. *Proceedings of the International Conference on Asian Language Processing (IALP)*. IEEE
24) Choi, J., Song, J.-H., & Kim, Y. (2018). An analysis of music lyrics by measuring the distance of emotion and sentiment. *Proceedings of the 19th*

IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). IEEE

25) B., A., & M., S. (2023). Recommendation of independent music based on sentiment analysis. *Proceedings of the 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE
26) Patchava, V., Jain, P., Lomte, R. R., G, S. P., & Kandala, H. B. (2015). Sentiment-based music play system. *Proceedings of the International Conference on Smart Sensors and Systems (IC-SSS)*. IEEE
27) Nguyen, N., & Seliya, N. (2024). Emotion sentiment analysis in Turkish music. *Proceedings of the IEEE International Conference on Information Reuse and Integration for Data Science (IRI)*. IEEE
28) Sarin, E., Vashishtha, S., Megha, & Kaur, S. (2021). SentiSpotMusic: A music recommendation system based on sentiment analysis. *Proceedings of the 4th International Conference on Computational Intelligence & Communication Networks (CICN)*. IEEE