



Classification of Explicit Songs Based on Lyrics Using Random Forest Algorithm

Luh Kade Devi Dwiyani¹, I Made Agus Dwi Suarjaya²,
Ni Kadek Dwi Rusjyanthi³

^{1,2,3}Information Technology Department, Udayana University, Bali, Indonesia
Email: ¹devidwiyani52@gmail.com, ²agussuarjaya@it.unud.ac.id, ³dwi.rusjyanthi@unud.ac.id

Abstract

This study focuses on the potential negative impact of explicit songs on children and adolescents. Although an explicit song labeling program is currently in place, its coverage is limited to songs released by artists affiliated with the Recording Industry Association of America (RIAA). Consequently, songs falling outside the program's scope remain inadequately labeled. To address this issue, a machine learning model was developed to effectively classify explicit songs and mitigate mislabeling challenges. A comprehensive dataset of song lyrics was collected using web scraping techniques for the purpose of constructing the classification model. The model was trained using the TF-IDF vectorization method and the random forest algorithm. A meticulous comparison of distribution parameters was conducted between the training and testing data sets to determine the optimal model. This superior model achieved a training-testing data distribution ratio of 90:10, with an impressive accuracy of 96.3%, precision of 99.3%, recall of 93.5%, and an f1-score of 96.3%. The classification results revealed that explicit songs accounted for 39.22% of the dataset, and the visual representation highlighted the fluctuating prevalence of explicit songs over time. Additionally, the hip-hop/rap genre exhibited the highest proportion of explicit songs, reaching a staggering 92%.

Keywords: Classification, Explicit Song, Machine Learning, TF-IDF, Random Forest

1. INTRODUCTION

The current digital era has made various digital contents, including music, easily accessible to everyone. However, children and adolescents often have subjective perceptions and tend to consume information without proper filtering. Therefore, it is crucial to protect them from explicit songs that contain offensive language, depictions of violence, sex, or substance abuse, which are inappropriate for their age. A study [1] highlights one negative impact of such songs, stating that listening to songs with sexual lyrics can influence adolescent boys' sexting behavior in the future.



To address this issue, the Recording Association of America (RIAA) has implemented the Parental Advisory Label Program (PAL). The program aims to identify explicit songs and assign them appropriate labels. However, the PAL program only covers songs published by artists affiliated with the RIAA, leaving songs on digital music streaming platforms outside its scope without proper labeling. Additional efforts are necessary to overcome this problem, such as employing machine learning technology for explicit song classification.

Several studies have utilized machine learning for song classification. Research [2] compared a dictionary-based method with a deep neural network for detecting explicit content in English songs and found that the complex model only slightly outperformed the simpler approach. Another study [3], which used naive Bayes to classify Indonesian songs based on age, reported low accuracy with this algorithm. Conversely, in research [4], an exploration of various vectorization methods and algorithms revealed that TF-IDF combined with the random forest algorithm achieved the highest accuracy. The effectiveness of the random forest algorithm was also demonstrated in a study [5] that classified explicit songs based on toxicity, where it outperformed logistic regression and XGBClassifier. Additionally, a study [6] focused on detecting explicit content in Korean songs using Adaboost and bagging techniques, and it determined that bagging with selective POS and IDF tags produced the most favorable results.

This study introduces a novel approach to classifying explicit songs, departing from previous methodologies that employed complete song lyrics. Instead, the study focuses on utilizing specific excerpts that best capture the explicitness of a song. This approach aims to address potential classification errors observed in prior research. By honing in on relevant lyrical segments, the study aims to enhance accuracy, improve classification effectiveness, and streamline data processing complexity. Additionally, the research explores four distinct scenarios for training-testing data distribution to identify the optimal model.

The classification results are then visualized to gain valuable insights into the annual growth of explicit songs and identify music genres with the highest percentage of explicit content. Employing the random forest algorithm and leveraging song excerpts for classification, this research makes a substantial contribution to the development of more effective explicit song classification methods. Moreover, it provides valuable information on growth trends and the characteristics of musical genres associated with explicit songs. This article contains explicit language and themes. Some readers may find these elements offensive, discriminatory, or vulgar. We encourage readers to approach the article objectively and respond thoughtfully to the language and explicit themes presented.

2. METHODS

The English song lyrics dataset utilized in this study was obtained through a meticulous web scraping process. Prior to the classification phase, the data undergoes several essential steps, including labeling, a split lyric process to extract a representative portion reflecting the song's explicitness, and preprocessing. For the classification task, the random forest algorithm in conjunction with the TF-IDF vectorization method is employed. Figure 1 provides a detailed illustration of the specific methods employed in this study.

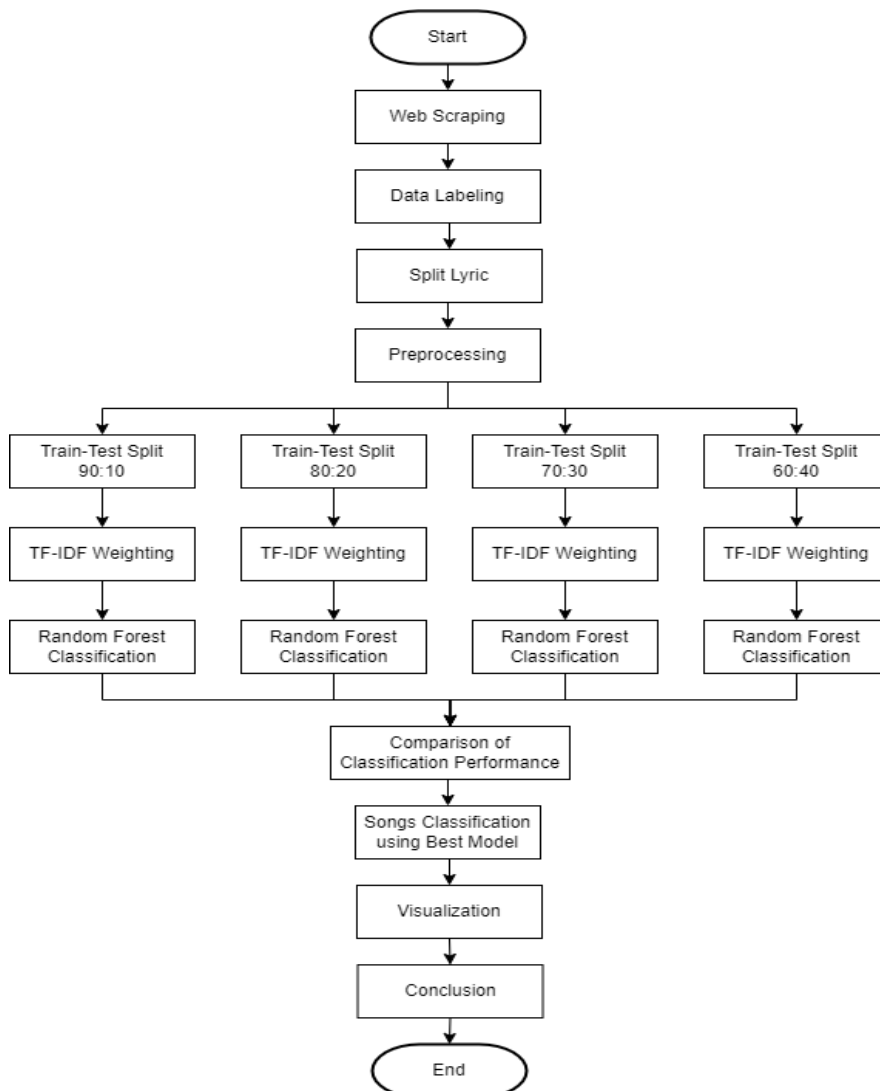


Figure 1. Research flowchart

2.1. Web Scraping

In this study, data collection is facilitated through web scraping, which involves utilizing the Python programming language and the BeautifulSoup library. BeautifulSoup, a Python package renowned for effectively extracting structured data from web pages, proves to be an ideal tool for scraping song lyrics. Notably, it offers a user-friendly interface, ensuring ease of use [7]. The data to be collected encompasses various elements such as artist name, song title, song lyrics, year of release, and genre. To obtain this information, web scraping techniques were employed, extracting data from diverse sources. For artist name, song title, and lyrics, web scraping was performed on the lirik.kapanlagi.com website. Conversely, for details regarding the year of release and song genre, web scraping was conducted on the Google search site. Figure 2 depicts an example website page employed as a data source.

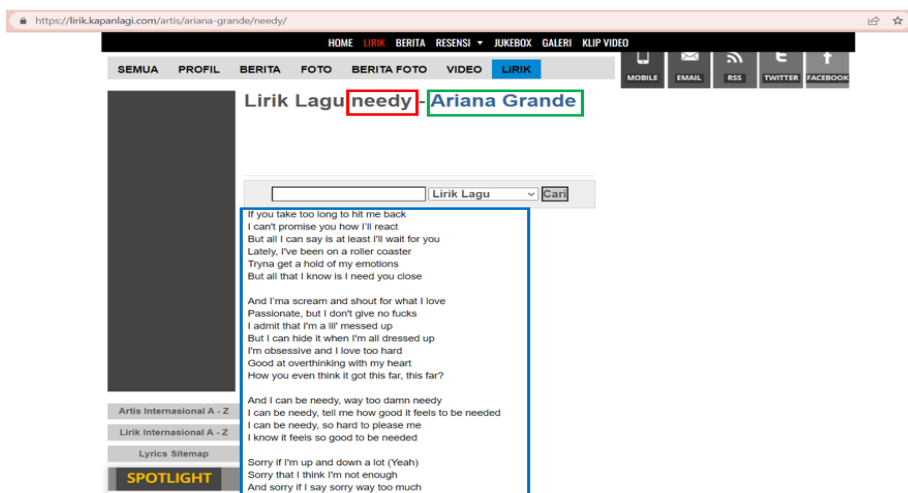


Figure 2. lirik.kapanlagi.com page

Figure 2 illustrates a webpage from the lirik.kapanlagi.com site, showcasing the data collection process for English song lyrics. The collected data comprises the artist name (highlighted by a green box), song title (highlighted by a red box), and song lyrics (highlighted by a blue box). Figure 3 displays a webpage from the Google search site, serving as a source for gathering information on the release year of a song. The web scraping process involves inserting the song title and artist name into the URL (highlighted by a red box) and subsequently extracting the release year data from the Google snippet feature (highlighted by a green box).

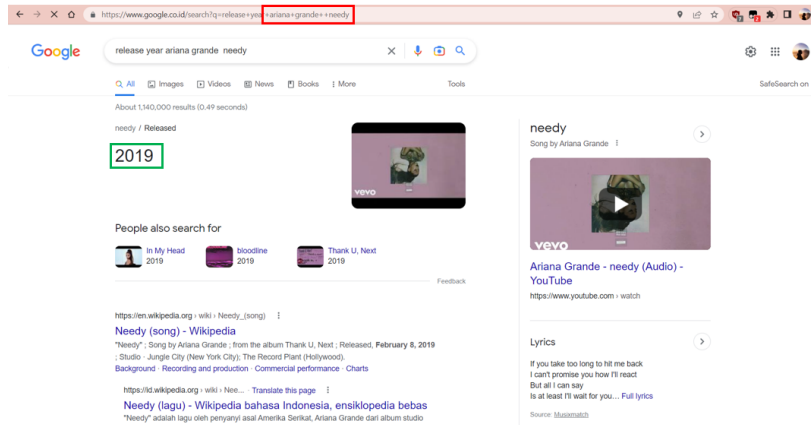


Figure 3. Google search (release year) page

Figure 4 showcases a webpage from the Google search site utilized for acquiring data on the genre of a song. Similar to the previous process, the web scraping involves inserting the song title and artist name into the URL (highlighted by a red box) and subsequently scraping the genre data from the Google snippet feature (highlighted by a green box).

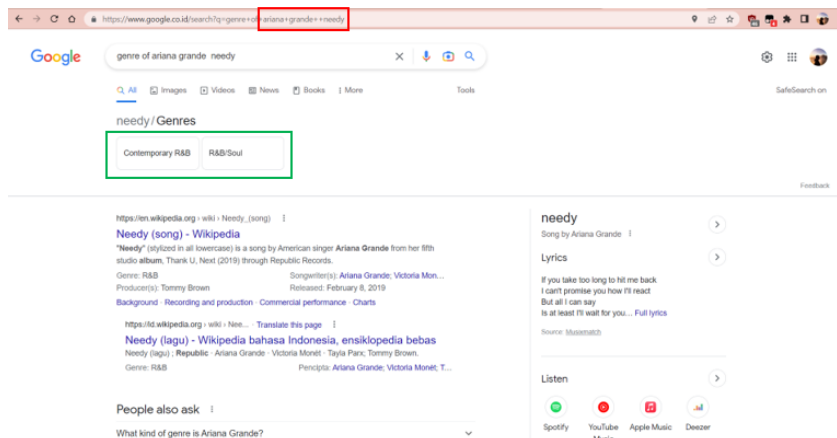


Figure 4. Google search (genre) page

2.2. Data Labeling

The process of data labeling involves assigning explicit or non-explicit labels to songs. This labeling process is carried out manually by referencing the available information on the Spotify music streaming platform. In particular, songs featuring an "E" icon on Spotify are designated as explicit. Figure 5 depicts the specific "E" icon referenced in this process.

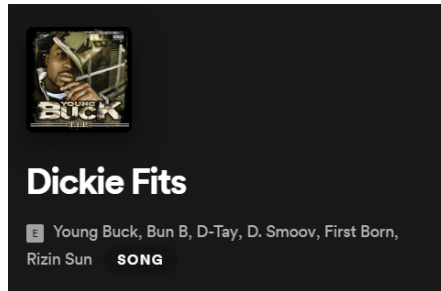


Figure 5. Song label on Spotify

2.3. Split Lyric

The term "split lyric" pertains to the automated process of segmenting lyrics into individual sentences. This split lyric process is executed utilizing the Python programming language. Following the division of lyrics, a meticulous selection procedure ensues to identify sentences that most accurately capture the explicit or non-explicit nature of the song. This selection process is conducted manually, considering explicit song criteria encompassing offensive language, depictions of violence, sexual behavior, substance abuse, offensive and discriminatory terms, as well as instances of physical or mental abuse.

2.4. Preprocessing

Preprocessing is an important step that must be carried out before data is classified, where the goal is to transform unstructured data into a more structured and uniform. In addition, the preprocessing process also aims to reduce data dimensions to make it easier to process. Preprocessing methods can improve dataset quality and text classification results [8]. In this study, several stages of preprocessing were carried out as follows.

- a. Lowercasing is a step-in data preprocessing to convert all letters in the lyrics to lowercase.
- b. Cleaning is a step-in data preprocessing to remove characters such as punctuation marks and numbers in lyrics.
- c. Stopwords removal, is a step-in data preprocessing to eliminate very common words in a language that do not contribute much to the meaning of the text.
- d. Lemmatization is a step-in data preprocessing to convert words into their basic form based on the corpus.

2.5. TF-IDF

TF-IDF is a method to determines term significance through calculating it weight [9]. The TF-IDF calculation in this study aims to select the most informative and

relevant features in distinguishing explicit and non-explicit classes and also avoid features that are less important or irrelevant. The following equation shows the TF-IDF formula.

$$TF = \frac{\text{number of times term } t \text{ appears in document } d}{\text{total number of terms in document } d} \quad (1)$$

TF (Term Frequency) is a method to calculate how often a word appears in a document.

$$IDF = \log \frac{\text{total number of documents in the corpus}}{\text{number of documents that contain term } t} \quad (2)$$

IDF (Inverse Document Frequency) is a method for determining the importance of a word in a collection of documents.

$$TF-IDF = TF \times IDF \quad (3)$$

The TF-IDF assigns the weight of a term to a document by giving a high value when the term appears many times in a number of documents, a low value when the term appears less frequently in a document or occurs in many documents, and the lowest value when the term appears in almost all documents.

2.6. Random Forest

In this study, the random forest algorithm is employed for explicit song classification. The random forest algorithm encompasses three crucial components. Firstly, it utilizes bootstrap sampling to construct multiple decision trees. Secondly, each decision tree independently generates predictions. Lastly, the random forest algorithm combines the outputs from each decision tree, employing a majority voting scheme to determine the final classification result [10].

The Random Forest algorithm is renowned for its simplicity and impressive performance, particularly when handling high-dimensional datasets [4]. To ensure expedited processing and optimize the efficiency of the random forest model, it is crucial to carefully determine the number of trees employed. As the number of decision trees increases, the complexity of the prediction step escalates [11]. However, despite this heightened complexity, the utilization of multiple decision trees endows random forests with the potential to achieve significantly superior accuracy compared to a single classifier [12]. It is precisely this remarkable capability that establishes the random forest algorithm as an ideal choice for classifying explicit songs in this study. Figure 6 illustrates the workflow of the Random Forest algorithm.

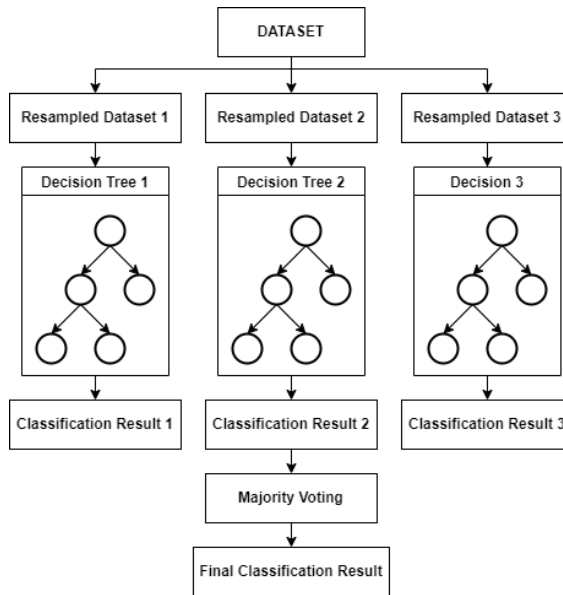


Figure 6. Workflow of random forest

3. RESULTS AND DISCUSSION

3.1 Dataset Results from Web Scraping

The process of web scraping from the sites lirik.kapanlagi.com and Google Search yields a comprehensive dataset of song lyrics, encompassing various attributes such as artist name, song title, song lyrics, year of release, and song genre. The dataset comprises a total of 200,670 English song lyrics, out of which 1,107 are utilized for modeling purposes, while the remaining 199,563 serve as data to be classified. The subsequent visualization provides a comprehensive overview of the dataset composition intended for classification.

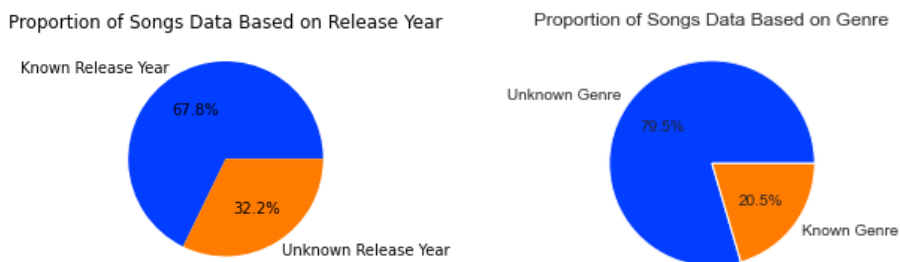


Figure 7. Proportion of songs based on known and unknown year and genre

Figure 7 presents a breakdown of songs in terms of known and unknown years and genres. In terms of the year, approximately 67.8% or 135,286 songs have a known year of release, while the remaining 32.2% or 64,277 songs have an unknown year. As for the genre, approximately 79.5% or 158,745 songs have an unknown genre, while the remaining 20.5% or 40,818 songs have a known genre. It is worth noting that a significant portion of the dataset lacks information on the year of release and genre due to the unavailability of related data on the web pages from which the scraping process was conducted.

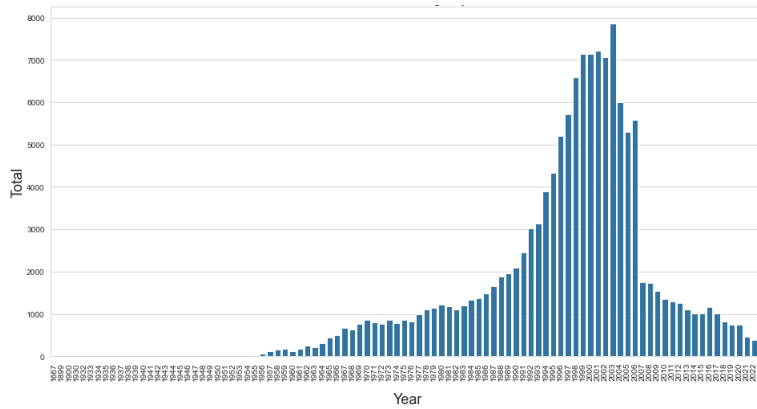


Figure 8. Distribution of songs by year

Figure 8 illustrates the distribution of songs from the year 1667 to 2023, noting that data for 2023 is available up until March. The graph showcases a significant variation between the minimum and maximum amounts of collected data. Specifically, the years 1667, 1936, and 1943 exhibit the lowest data availability, with only one song recorded for each respective year. In contrast, the year 2003 boasts the highest abundance of data, with 7,857 songs available. Analyzing the distribution of songs across different years allows for the examination of explicit song trends over time.

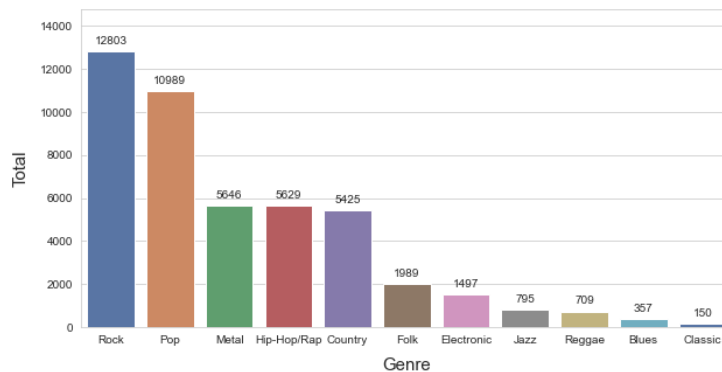


Figure 9. Distribution of songs by genre

Figure 9 showcases the distribution of songs across eleven popular music genres. Notably, there is a substantial disparity in the quantity of songs between the genre with the highest and lowest data count. The rock genre reigns supreme, with a staggering 12,803 songs, while the classic genre lags behind with a modest 150 songs. It is important to acknowledge that some songs may belong to multiple genres, as they possess musical characteristics that align with various genres simultaneously. To provide an insight into the web scraping outcomes, an example of the results is presented in Table 1.

Table 1. Web scraping results

Artist	Title	Release Year	Genre	Lyric
Young Buck	Dickie Fits	2005	Hip-hop /Rap	“See nigga we thugged out for a reason. Niggaz ain't thuggin, because, they like the look nigga. Or they like to be on these....”

3.2 Data Labeling Result

The labeling process is conducted manually by referencing the label information available on the Spotify music streaming platform. This information is indicated by the presence of an "E" logo associated with a song when searched on Spotify. However, considering the potential for certain songs to contain explicit content yet lack the explicit label, songs without the "E" logo undergo manual verification. During this manual verification, the presence of offensive language, depictions of violence, sexual behavior, substance abuse, offensive and discriminatory words, as well as physical or mental abuse within the lyrics, is meticulously examined. To provide a glimpse into the results of the data labeling process, an example is presented in Table 2 below.

Table 2. Data labeling results

Lyric	Label
“See nigga we thugged out for a reason. Niggaz ain't thuggin, because, they like the look nigga. Or they like to be on these....”	Explicit
“It's been a long time coming. This breakdown in our sights. And I just keep on running. Back for the same fight. I'll look...”	Non-Explicit

The process of labeling song data yielded a total of 1,107 labeled songs, which were classified into explicit and non-explicit categories. Figure 10 provides a visual representation of the distribution, showcasing the respective counts for each category.

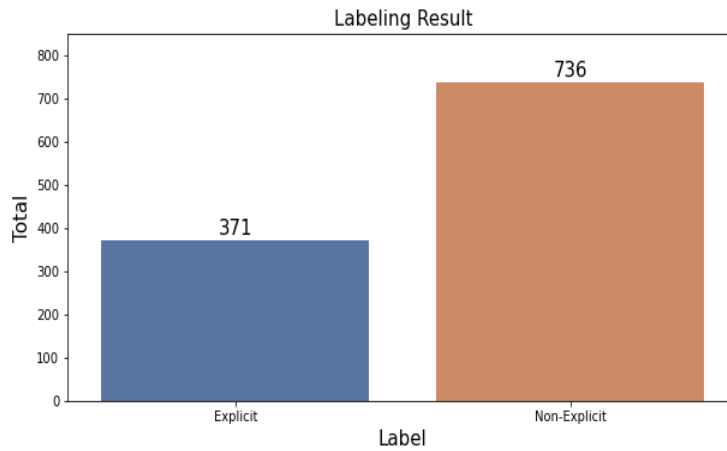


Figure 10. Labeling Result

3.3 Split Lyric

The split lyric process involves dividing the lyrics into individual sentences, followed by the selection of representative excerpts that correspond to the song's explicit or non-explicit label. This selection is performed manually, taking into consideration the presence of offensive language, depictions of violence, sexual behavior, substance abuse, offensive and discriminatory words, as well as physical or mental abuse within the lyrics. The split lyric stage resulted in a collection of 3,000 lyrics, evenly distributed between explicit and non-explicit classes, with 1,500 data points in each category. Table 3 provides an example of the split lyric results.

Table 3. Split lyric results

Lyric	Label
See nigga ain't thugged out of a reason.	Explicit
Niggaz ain't thuggin, because, they like the look nigga.	Explicit
If it ain't one of these bitches the these niggaz won't test.	Explicit
It's been a long time coming.	Non-Explicit
Not a single word comes to mind.	Non-Explicit
I'll think of you.	Non-Explicit

3.4 Preprocessing

The preprocessing steps applied to the lyrics include lowercasing, cleaning, stopwords removal, and lemmatization. Lowercasing involves converting all characters/letters to lowercase. Cleaning entails removing punctuation, numbers, and non-ASCII characters from the string. Stopwords removal is performed by eliminating highly frequent words in the language that lack substantial meaning on their own. The stopwords used are a combination of NLTK library, sklearn, and

additional custom stopwords. Lemmatization is conducted to transform words into their base form based on the corpus. The lemmatization process utilizes the NLTK library. Table 4 provides an example of the preprocessing results.

Table 4. Preprocessing results

Preprocessing	Lyric
Before preprocessing	If it ain't one of these bitches the these niggaz won't test
After Lowercasing	if it ain't one of these bitches the these niggaz won't test.
After Cleaning	if it aint one of these bitches the these niggaz wont test.
After Stopwords Removal	bitches niggaz test
After Lemmatization	bitch niggaz test

3.5 Comparison of Classification Model

The classification model was developed using four different data distributions: 90:10, 80:20, 70:30, and 60:40. The random forest algorithm and the TF-IDF vectorization method were employed in constructing the classification model. Performance parameters, including accuracy, precision, recall, and F1-score, were used to evaluate and compare these models. Figure 10 presents a comparison of the classification models.

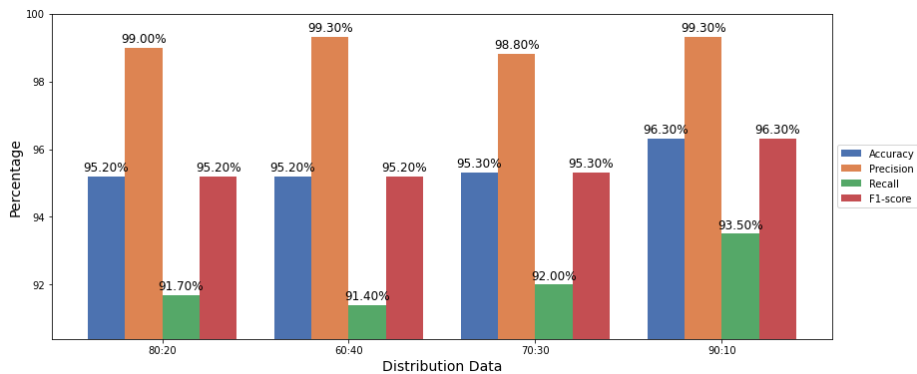


Figure 11. Comparison of Classification Model

Figure 11 illustrates the comparison of model classification performance. The model with a training-testing data distribution of 90:10 achieved the highest accuracy, precision, recall, and F1-score. The accuracy reached 96.3%, precision reached 99.3%, recall reached 93.5%, and the F1-score reached 96.3%. These elevated values of accuracy, precision, recall, and F1-score indicate excellent classification performance. Notably, this model surpassed previous studies that utilized the random forest algorithm. Compared to [4], our model exhibited a 12.3% higher accuracy, 7.6% higher precision, 18.4% higher recall, and 13.7% higher F1-score. Furthermore, in comparison to [5], our model demonstrated a

2.8% higher accuracy, 7.3% higher precision, and 2.3% higher F1-score. Given the remarkable values of accuracy, precision, and recall, the 90:10 training-testing data distribution was selected for the classification process.

3.6 Visualization of Classification Result

Upon successful completion of the classification process, various aspects of the results are visualized to provide a comprehensive overview. The visualizations include a comparison of the number of songs classified into explicit and non-explicit categories, the percentage of explicit songs within each genre, the annual growth of explicit songs, the frequency of words in explicit songs, and the singer with the highest number of explicit songs.

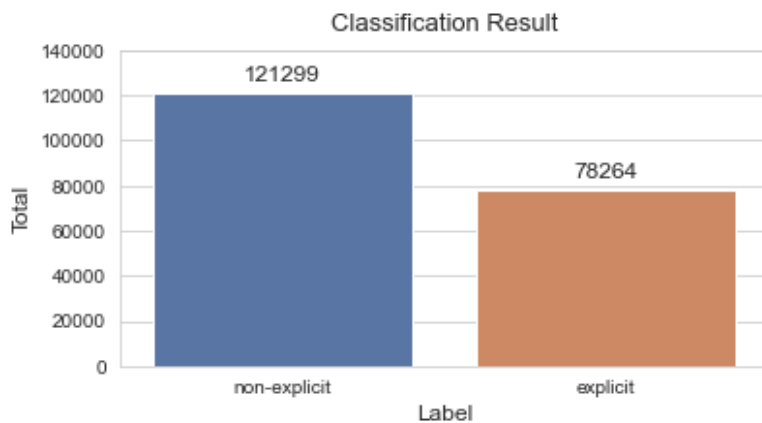


Figure 12. Classification results

Figure 12 depicts a visualization of the classification outcomes, revealing that the number of songs classified as non-explicit exceeds that of explicit songs. Out of the total dataset comprising 199,563 songs, 121,299 were classified as non-explicit, while 78,264 were classified as explicit. Despite the significant difference in the number of songs between the two categories, both hold substantial representation within the dataset, each consisting of over 78,000 songs. This demonstrates the significance of explicit and non-explicit classifications, as they provide valuable insights for further analysis, such as identifying music genres that tend to contain explicit content and observing the trends in explicit songs over the years.

The percentage of explicit songs is calculated by dividing the total number of explicit songs in each genre by the overall number of songs within that genre, encompassing both explicit and non-explicit categories. Notably, the Hip-Hop/Rap genre exhibits the highest percentage of explicit songs, comprising 91.97% of all songs within that genre. This elevated percentage can be attributed

to the frequent use of profanity in hip-hop songs. According to a study [13], hip-hop songs contain the highest number of profanity words, with one profanity word occurring approximately every 47 words. Conversely, the jazz genre demonstrates the lowest percentage of explicit songs, accounting for merely 15.72% of the total songs in that genre. Figure 13 provides a visualization depicting the percentage of explicit songs for each genre.

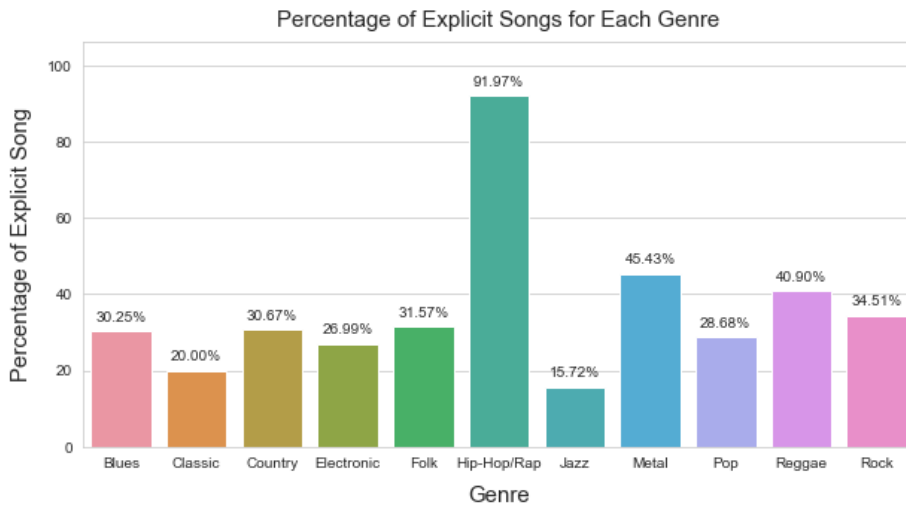


Figure 13. Percentage of explicit songs for each genre

The growth of explicit songs has been visualized starting from 1985, which marks the introduction of the "Parental Advisory, Explicit Content/Lyrics" label. The visualization reveals that the percentage of explicit songs exhibits a fluctuating pattern over the years. In the early years of 1985, the percentage of explicit songs stood at approximately 29.97%. It then rose to 37.55% in 1988 but decreased again to 36.94% in 1989. Subsequently, the percentage experienced fluctuations, reaching its peak in 2017 and 2018, with percentages of approximately 48.50% and 48.65%, respectively. However, despite the high number of explicit songs in 2018, there was a sharp decline in 2020, where the percentage dropped to 37.26%. It then rose again to 39.01% in 2021 and further increased to 42.67% in 2022. From this data, it can be concluded that the percentage of explicit songs tends to rise and fall from year to year, lacking a clear overall trend. This fluctuation could be influenced by various factors, such as shifts in musical preferences or changes in the popularity of specific genres during a given year. Figure 14 provides a visualization depicting the growth of explicit songs each year.

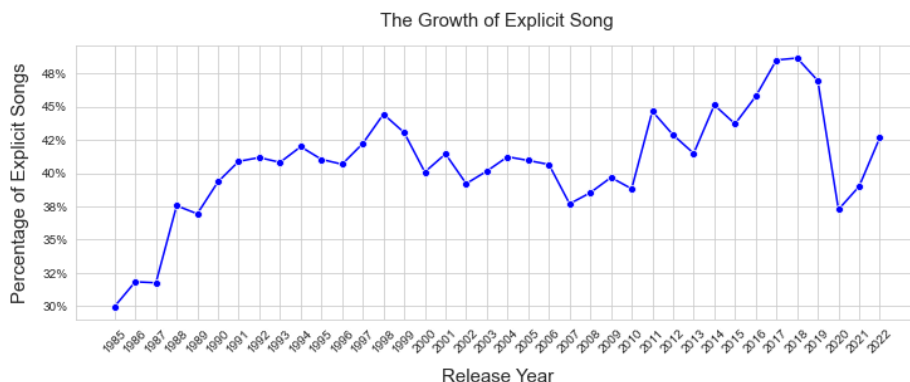


Figure 14. The growth of explicit song

The size of each word corresponds to its frequency of occurrence. Upon analyzing the visualization, it becomes evident that certain words such as 'hell', 'kill', 'drink', 'fck', and 'sh't prominently feature in explicit songs. This observation indicates that explicit songs frequently incorporate violent, vulgar, and aggressive content, which has the potential to influence the perceptions and behaviors of listeners, particularly children and adolescents. Figure 15 presents the visualization illustrating the word frequency in explicit songs.



Figure 15. Word frequency in explicit song

Among the singers analyzed, Master P emerges as the artist with the highest number of explicit songs, totaling 148. Master P, an American rapper primarily associated with the hip-hop/rap genre, is joined by nine other artists on the list, all of whom are also rappers or hip-hop groups/duos. These include Jay-Z, Lil Wayne, Nas, LL Cool J, B.o.B, as well as Insane Clown Posse, Twiztid, Cypress Hill, and Three 6 Mafia. This visualization highlights the tendency of singers in the hip-hop/rap genre to produce a substantial number of explicit songs, likely due to their utilization of bold and provocative lyrics, often containing offensive

language. Figure 16 showcases the visualization of the top 10 singers with the highest number of explicit songs, based on the web scraping dataset employed in this study.

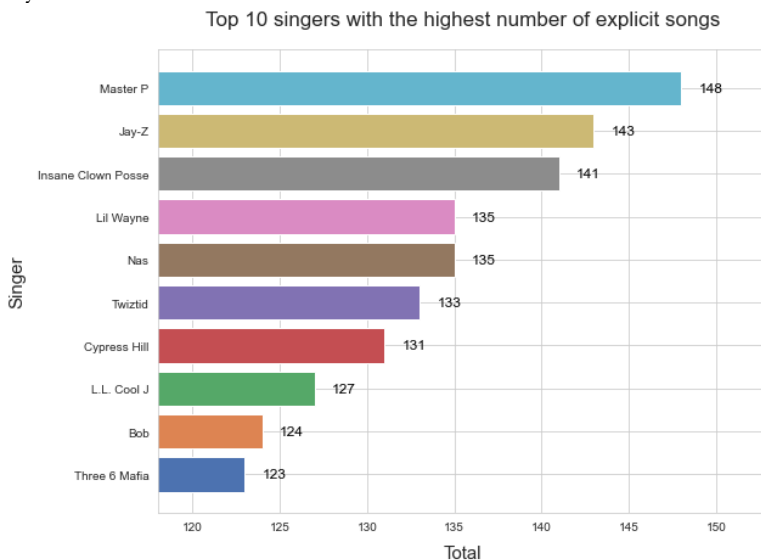


Figure 16. Top 10 singers with the highest number of explicit songs

4. CONCLUSION

The classification model, utilizing the TF-IDF vectorization method and the random forest algorithm with song lyrics as input, proves to be effective in accurately classifying explicit songs. Performance evaluation metrics, including accuracy, precision, recall, and f1-score, consistently exceed 90%. The optimal model, with a training-testing data distribution of 90:10, achieves exceptional performance with 96.3% accuracy, 99.3% precision, 93.5% recall, and 96.3% f1-score. Notably, the accuracy of this model surpasses previous research utilizing similar algorithms and tasks. The utilization of lyric snippets and efficient data processing contribute to the model's superior performance. Visualizing the classification results reveals that explicit songs exhibit fluctuations from year to year, devoid of a discernible trend. This variability can be attributed to various factors, such as shifting musical preferences and changes in the popularity of different genres within a given year. Conversely, the classification results highlight the dominance of the Hip-Hop/Rap genre in the explicit song category. This observation aligns with the visualization findings, which illustrate that singers in the Hip-Hop/Rap genre have the highest number of explicit songs. The explicit nature of this genre is likely influenced by its frequent use of profanity. Thus, the visualization supports the conclusion that the Hip-Hop/Rap genre tends to produce explicit content.

REFERENCES

- [1] S. L. Keenan-Kroff *et al.*, “Associations between sexual music lyrics and sexting across adolescence,” *Comput. Hum. Behav.*, vol. 140, p. 107562, Mar. 2023, doi: 10.1016/j.chb.2022.107562.
- [2] M. Fell, E. Cabrio, M. Corazza, and F. Gandon, “Comparing Automated Methods to Detect Explicit Content in Song Lyrics,” in *Proceedings - Natural Language Processing in a Deep Learning World*, Incoma Ltd., Shoumen, Bulgaria, Oct. 2019, pp. 338–344. doi: 10.26615/978-954-452-056-4_039.
- [3] Moch. F. Shadiqin Thirafi and F. Rahutomo, “Implementation of Naïve Bayes Classifier Algorithm to Categorize Indonesian Song Lyrics Based on Age,” in *2018 International Conference on Sustainable Information Engineering and Technology (SIET)*, Malang, Indonesia: IEEE, Nov. 2018, pp. 106–109. doi: 10.1109/SIET.2018.8693201.
- [4] L. Bergelid, “Classification of explicit music content using lyrics and music metadata,” Master’s Thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2018.
- [5] M. A. S. Siddique, M. I. Sarker, R. Ghosh, and K. Gosh, “Toxicity Classification on Music Lyrics Using Machine Learning Algorithms,” in *2021 24th International Conference on Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh: IEEE, Dec. 2021, pp. 1–5. doi: 10.1109/ICCIT54785.2021.9689865.
- [6] H. Chin, J. Kim, Y. Kim, J. Shin, and Mun. Y. Yi, “Explicit Content Detection in Music Lyrics Using Machine Learning,” in *2018 IEEE International Conference on Big Data and Smart Computing (BigComp)*, Shanghai: IEEE, Jan. 2018, pp. 517–521. doi: 10.1109/BigComp.2018.00085.
- [7] M. Khder, “Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application,” *Int. J. Adv. Soft Comput. Its Appl.*, vol. 13, no. 3, pp. 145–168, Dec. 2021, doi: 10.15849/IJASCA.211128.11.
- [8] Y. HaCohen-Kerner, D. Miller, and Y. Yigal, “The influence of preprocessing on text classification using a bag-of-words representation,” *PLOS ONE*, vol. 15, no. 5, p. e0232525, May 2020, doi: 10.1371/journal.pone.0232525.
- [9] H. S. Sint and K. K. Oo, “Comparison of two methods on vector space model for trust in social commerce,” *TELKOMNIKA Telecommun. Comput. Electron. Control*, vol. 19, no. 3, p. 809, Jun. 2021, doi: 10.12928/telkomnika.v19i3.18150.
- [10] R. R. Waliyansyah and N. D. Saputro, “Forecasting New Student Candidates Using the Random Forest Method,” *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 11, no. 1, p. 44, Apr. 2020, doi: 10.24843/LKJITI.2020.v11.i01.p05.
- [11] V. Maslej-Krešňáková, M. Sarnovský, P. Butka, and K. Machová, “Comparison of Deep Learning Models and Various Text Pre-Processing

- Techniques for the Toxic Comments Classification,” *Appl. Sci.*, vol. 10, no. 23, p. 8631, Dec. 2020, doi: 10.3390/app10238631.
- [12] Abdulkareem, Nasiba Mahdi and Abdulazeez, Adnan Mohsin, “Machine Learning Classification Based on Radom Forest Algorithm: A Review,” Jan. 2021, doi: 10.5281/ZENODO.4471118.
- [13] C. Thompson, “Lyric-Based Classification of Music Genres Using Hand-Crafted Features,” *Reinvention Int. J. Undergrad. Res.*, vol. 14, no. 2, Oct. 2021, doi: 10.31273/reinvention.v14i2.705.