# Song Lyrics Genre Detection Using RNN

A Project Report in partial fulfillment of the degree

**Bachelor of Technology**

in

**Computer Science & Engineering & Electronics & Communication Engineering**

**By**


| | |
|---|---|
| 19K41A04B6 | S. Amulya |
| 19K41A05A8 | P. Rithika Reddy |
| 19K41A05B1 | S. Shiva Keerthi |


**Under the Guidance of**
**Mr. D. Ramesh,**
**Assistant Professor,**
**Dept of CSE.**

**Submitted to**


**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**
**S.R. ENGINEERING COLLEGE(A), ANANTHASAGAR, WARANGAL**
**(Affiliated to JNTUH, Accredited by NBA) Dec-2021.**

## DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

## CERTIFICATE

This is to certify that the Project Report entitled "Song Lyrics Genre Detection Using RNN" is a record of bonafide work carried out by the student(s) S . Amulya, P. Rithika Reddy, S. Shiva Keerthi bearing Roll No(s) 19K41A04B6, 19K41A05A8, 19K41A05B1 during the academic year 2022-23 in partial fulfillment of the award of the degree of ***Bachelor of Technology*** in **Computer Science & Engineering/Electronics & Communication Engineering** by the Jawaharlal Nehru Technological University, Hyderabad.

.

**Supervisor** **Head of the Department**

**External Examiner**

# ABSTRACT

Digitalization of music is the new trend, and preferences of individuals are highly rated. Millions of songs are being streamed in the music applications. The companies providing these services need to sort and arrange a wide range of music tastes for all of its users. On top of that, fresh music from various artists in a wide spectrum of genres are popping up every day. To keep track of all this, a classification system can be handy. So, we propose an RNN based model based on Natural Language processing to classify the songs based on their lyrics into different genres. Additionally, this tool can be handy to the music lovers for quickly identifying which genre a particular song belongs to. In this paper, we apply Long Short Term Memory (LSTM) model with both Universal Serial Embedder (USE) and Bert embedders. A comparative study is performed to understand which combination of models works based to classify the genres based on lyrics. From our results, on the basis of accuracy of the model, we found that USE embedder with LSTM gives a slightly better performance than Bert embedder. The LSTM model with USE embedding gave the highest accuracy of 83.42% when trained over a range of five folds.

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1

# INTRODUCTION

Over the past decade, music industry has seen major improvements with digitization of songs. More and more artists are releasing their fresh music in digital marketing rather than the traditional physical means. This enables the artist's music to spread all over the world. Services such as Amazon, Apple music, Spotify, etc. are providing consumers services for direct download and continuous streaming of their favorite music through different payment schemes. To maintain this huge database has become a prominent issue for the digital music providers as well as third party entities. For instance, Spotify's web application programming interface has given the developers a technique to pick out a large set of data for a particular selected song. The data picked can include the name of the song, artist's name, acoustics, genre, bass measure, tempo, etc. Additionally, data can also contain information about the popularity of the particular song. The websites which display the music lyrics for a song are developing APIs to access massive amounts of lyrical data to the developers.

Genres in music can have a wide spectrum of diversity. They can be separated on the basis of sound, aesthetics, lyrics, topics, etc. Pop culture has segregated several genres depending on the topic it discusses. So mainly, the lyrical part of song can decide the genre. In this project, we implement RNN model with the concepts of Natural Language Processing to build a tool to predict the genre based on the song lyrics. This analysis can be helpful for the developers of music websites, but also to the producers and musicians, to determine the characteristics of the lyrics and to understand which genre their song will typically fall into. Additionally, music enthusiasts can use this model to predict the genres of their favorite music.

For classification, we can use CNN, Naïve Bayes or RNN. But natural language process usually gives a higher accuracy for word classification. So, to build this model, we use RNN Long Short Term Memory(LSTM) model for prediction of genre based on lyrics. On top of this, we are going to use two separate embedders, which are Universal Serial Embedder (USE) as well as Bidirectional Encoder (BERT) embedder. We perform a comparative study with these embedding techniques to find the best performing model. These are pre-trained models which can help us achieve higher accuracies so as to deploy the models in real life scenarios.

# CHAPTER 2
# LITERATURE REVIEW

## 2.1 Related Work

[1] In this paper, they tried to compare whether the language representation models are better than the traditional deep learning models. They examined using the language representation models like BERT, DistilBERT and RNN Model BiLSTM. They did a comparative study for both single label and multi label classification. They found that in each task BERT outperforms the other models. Even though the complexity is reduced in DistilBERT still the complexity of BERT enables it to learn more efficiently. Since BERT and D-BERT are the advanced models they have a capacity to handle the sequence data well compared to the RNN Model BiLSTM. Finally, after compiling all the results they observed that BERT performs well with an accuracy of 77.63% for one-label and 71.29% for multi-label genre classification.

[2] In this paper, they classified the songs into four genres namely Christian, Metal, Country and Rap. Here they have used both lyrics and album artwork for the classification purpose. For the album artwork they used CNN model to classify the image data. To process lyrical data, they used two natural processing techniques like Bag of Words, Term Frequency-Inverse Document Frequency (TFIDF). After this to classify they used Machine Learning Algorithms like SVM, Naïve Bayes, XGBoost. They observed that in the embedding part TFIDF outperforms Bag of Words as it not only considers frequency of a particular word but also considers its frequency in the entire sequence and assigns terms to each word.

[3] In this paper, they implemented a Transformer Classifier which analyzes the relationship between the various audio frames well and achieves good performance to classify the genres. For this task they have audio clips of 1000 songs in 10 different genres like Jazz, Country, Classical, Rock etc. Each audio clip is 30 seconds duration which is further split into 3 seconds smaller clips. With an overlap of over 50% for each audio clip they have 18 clips in total. Here they used two methods for analyzing the result. The first one is to choose the category of each chip based on the output of the network and select the one with highest incidence of the 18 categories. And the second method is to sum the respective probability vector of 18 chips and generate a general probability vector to classify the result. Finally, after running for 10 epochs they have developed a model with 76% accuracy by averaging the first clip and 75.8% by averaging the second clip.

[4] In this paper, they classified 138,368 Brazillian songs into 14 genres like Gospel, Sertanejo, Rock, Pop etc using only lyrics. To embed the lyrics into vector they have used word embedding techniques like Word2Vec, Wang2Vec, FastText and Glove. Then they applied SVM, Random Forest and BiLSTM Network to classify a song into its respective genre. For every model they tabulated the results by applying the above mentioned 4 embedding techniques. For SVM they used the default parameters, for Random Forest they have used 100 trees. The BLSTM model was trained with 4 epochs, with adam optimizer and 256 neurons in the hidden layer. In the result analysis, they observed that Wang2Vec provided best F1score in BLSTM Network it is not necessarily the same case with the other models. Finally, they observed that the BLSTM Network with Wang2vec word embedding provided the best F1-Score result.

[5] In this paper, they classified the songs into 4 genres namely Christian, metal, country, rap using lyrics. For embedding the sentences, they used techniques like Word2vec and Word2Vec with TFIDF. After converting it to a vector they used machine learning algorithms like Support Vector Machine, Random Forest, XGBoost, Deep Neural Networks etc to classify the songs into its respective genre. The initial data consist of around 19000 samples which are spread over 4 genres. The data is split in 7:3 ratio for training and testing part. They have pre-processed the data by performing tokenization, removal of delimiters, removal of non-alphabetic characters and stop words on the data. Also, they have removed the words that are not in the English dictionary. At last they observed that rather than using Word2Vec, Word2Vec with TFIDF performed better in all the models. Overall, Deep Neural Network with 3 layers and Word2Vec TFIDF model outperforms the other models with an accuracy of 74%.

[6] In this paper, they have examined various approaches to classify the music genre using lyrics. Here they aim to show that even simple combinations of these approaches can improve the accuracy. Here they have used 4 datasets and they tried to classify a song into 7 genres like fold, indie, country, metal etc. The pre-processing techniques used for this approach is setting all the words to lowercase, lemmatization on the generated tokens. Here they have used Bag of Words approach and deep learning approaches. In one of the model they used Bag of Words technique for embedding the sentences; to classify the vectors into genres they used Naïve Bayes, Support Vector Machine, XGBoost and Random Forest. In the other models they used a combined both LSTM and CNN models for classification. The final model they have used here is Hierarchical Attention. In comparison with the other models they observed that Bag of Words approach performed well in all the datasets with an accuracy of 64%.

[7] In this paper, they have conducted a set of experiments to classify the music genre using learned, hand crafted and the fusion of them. Handcrafted features are extracted from the audio signal, lyrics, chords and spectrogram images are obtained from the audio. Here they have used brazillian lyrics dataset to classify the songs into 6 genres such as Axe, Rap, Forro etc. To examine the handcrafted features, they used algorithms like K-Nearest Neighbor, Decision Tree, SVM, and Random Forest, CNN etc. While testing the performance of the model with handcrafted features they observed that Random Forest performed better than other models. After evaluating all the results, they observed that CNN with Spectograms which achieved an accuracy of 78%.

[8] Here they have proposed a new study in which they determine which part of the songs should be used to train the model so that the accuracy can be improved. A dataset of 200 songs is split into 160 and 40 songs which are used for training and testing respectively. Here they considered English songs in 20 different genres like hop, jazz, pop, rock etc. They first conducted an acoustic analysis to generate acoustic features of a song. In the second part, they used these features for a set of songs classify the genre. For both these parts, they have conducted a study to determine to use the full song or a part of song (30-60 second time interval). They also further analyzed the same one using Turkish songs. Finally, they observed that using the full song to train the model achieved the best accuracy of 55%.

[9] In this paper, they examined various methods to develop a multi-emotion classification model using song lyrics. Here they have classified the songs into 6 emotions such as Anger, Disgust, Fear, Joy etc. They have used CBET, TEC and DailyDialog lyrical datasets to train the model. For developing the model, they used algorithms like Naïve Bayes, Random Forest, and a Most Frequent Sense (MFS) model. The Naïve Bayes model was trained using Bag of Words embedding technique and Random Forest was trained based on transformed feature vectors. They also used complex embedding techniques like BERT. Separate BERT models were trained on CBET dataset individually for each emotion. They observed that BERT models did not generalize well on the training datasets. When compared the model performance with respect to each emotion they observed that BERT models only performed well on joy and sadness emotions and are outperformed by Naïve Bayes classifier on disgust and fear emotions.

[10] In this paper, they determined the music genre using lyrics. Here they have classified the genre into 11 classes such as Rock, R&B, Indie, Metal etc. They have removed the special characters in the lyrical data and the bracketed notes like [Chorus], [Verse 1] were alaso removed. For both the models they split the data into 0.66, 0.11, 0.22 ratio for training, validation and testing respectively. Here they have used two approaches to complete the task. The first one is LSTM

model. The validation accuracy for this model was 44.81% and the testing accuracy was 44.85%. In the second approach which is TF-IDF method they lemmatized the lyric data. So, this can be cause for the spike in accuracy. The validation accuracy was 55.15%, test accuracy was 55.36%. Also, using the TF-IDF data they also fit a Logistic Regression model and observed that the test accuracy was 53.98%. Finally, they analyzed the F1-scores for both the models and observed that for electronic, indie, jazz, indie and other the performance was poor.

***Table 2.1:*** *Brief Literature Survey*

| S.no | Authors | Title of the Paper | Description |
|---|---|---|---|
| [1] | Akalp, Hasan, et al. | Language representation models for music genre classification using lyrics | • Used models like BERT, DistilBERT, BiLSTM. Compared both Single and Multi-label Classification of genres.<br>• Observed that BERT outperforms the other models with an accuracy of 77.63% for single-label and 71.29% for multi-label classification.<br>• The complexity of BERT helps the model to learn better Even though DistilBERT is less complex BERT. |
| [2] | A. Kumar, A. Rajpal and D. Rathore, | Genre Classification using Feature Extraction and Deep Learning Techniques | • Used both lyrics and album artwork to classify the genre<br>• Classified the songs into four genres namely Christian, Metal, Country and Rap.<br>• Used Bag of Words, TF-IDF for embedding and to classify they used SVM, Naïve Bayes, and XGBoost.<br>• Observed TF-IDF embedding technique performs better than other models. |
| [3] | Zhuang, Y., Chen, Y., & Zheng, J. | Music genre classification with transformer classifier | • Used audio clips of 1000 songs in 10 genres like Jazz, Country, Classical etc.<br>• Each audio clip is of 30 seconds which is further split into 3 seconds clips hence each audio clip has 18 clips.<br>• Implemented a transformer classifier, based on category of each chip in the output network they chose the one with highest incidence.<br>• The other method is to sum the respective probability |

| | | | |
|---|---|---|---|
| | | | vector of 18 chips. Finally, the highest accuracy is 76%. |
| [4] | Araújo Lima, Raul de, et al. | Brazilian lyrics-based music genre classification using a BLSTM network | • Classified Brazillian songs into 14 genres like Gospel, Rock, Pop etc using song lyrics.<br>• Embedding techniques used are Word2vec, Wang2Vec, FastText, Glove.<br>• Classification models used are SVM Random Forest and BiLSTM.<br>• Observed that BiLSTM network with Wang2Vec embedding technique gave the best F1-Score result. |
| [5] | Kumar, Akshi, Arjun Rajpal, and Dushyant Rathore | Genre Classification using Word Embeddings and Deep Learning | • Classified the songs into 4 genres like Christian, Country, Metal and Rap using lyrics.<br>• Embedding technique sused are Word2Vec and Word2Vec with TFIDF.<br>• Used Machine Learning Algorithms like SVM, XGBoost, Deep Neural Networks etc to classify the genre.<br>• Finally Deep Neural Network with Word2Vec and TFIDF performed well than other models with an accuracy of 74%. |
| [6] | Ueno, C. L. R. S., & Silva, D. F. | On Combining Diverse Models for Lyrics-Based Music Genre Classification | • Examined various approaches to classify the music genre using lyrics.<br>• Used 4 datasets to classify the song genre into 10 classes such as Indie, Country, metal etc.<br>• For embedding they used Bag Of Words and for classification they used SVM, XGBoost Naïve Bayes etc.<br>• Observed that Bag Of Words approach performed well with accuracy of 64%. |
| [7] | Pereira, Rodolfo M., et al. | Representation Learning vs. Handcrafted Features for Music | • Conducted a set of experiments to classify the genre using learned, hand-crafted and both.<br>• Used Classified the songs into 6 genres such as Axe, Rap, Forro etc. |

| | | | |
|---|---|---|---|
| | | Genre Classification | • Used algorithms like SVM, KNN, Decision Tree, CNN etc.<br>• After evaluating all the results they observed that CNN model with Spectograms performed well with an accuracy of 78%. |
| [8] | Atsız, E., Albey, E., & Kayış, E. | Effective Training Methods for Automatic Musical Genre Classification | • Aim to determine which part of the songs should be used to train the model.<br>• Used a dataset of 200 songs consisting of 20 genres such as hop, jazz, pop, rock etc.<br>• Conducted an acoustic analysis to generate acoustic features for a song and then using these features they classified the genre.<br>• Comparative study whether to use the full song or a part of song to train the model. The best accuracy was 55% achieved by using full song. |
| [9] | Edmonds, D., & Sedoc, J. | Multi-Emotion Classification for Song Lyrics | • Classified the songs into 6 emotions such as Anger, Disgust, fear etc using Song lyrics.<br>• Used Naïve Bayes, Random Forest, Most Frequent Sense (MSF) model to classify the songs.<br>• Naïve Bayes was trained using Bag of Words and Random Forest was trained based on transformed feature vectors.<br>• Also used BERT model and observed that it did not generalize well on the all the genres. |
| [10] | Mubeen, M. | Using Deep Learning to Predict Music Genre from Song Lyrics | • Classified the songs into 11 genres such as Rock, R&B, Indie etc using lyrics.<br>• Removed special characters in lyrical data and bracketed notes.<br>• Used LSTM, TF-IDF, Logistic Regression models to determine the best model<br>• Observed that the test accuracy for TF-IDF with Logistic Regression model was 53.98%. |

# CHAPTER 3

# DATASET INSIGHTS

The dataset we are using is brought from the kaggle machine learning repository. This dataset contains two datasets of artists-data and lyrics-data. This dataset is originally obtained by scraping the Brazilian website called Vagalume [11]. In the lyrics-data csv file, there are different columns specifying the link of the song and artist, song name, its lyric and the language. In artists-data file, we see the name of the Artist, the genres they sing in, number of songs they have and their popularity rate along with link of the song. The entire dataset contains a total of 79 musical genres like hip hop, jazz, R&B, rock, Black music, Rap, etc. with different artists in multiple languages: English, Portuguese and other [11].

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | ALink | SName | SLink | Lyric | language |
| 2 | /ivete-sangalo/ | ArerÃª | /ivete-sangalo/arere.html | Tudo o que eu | pt |
| 3 | /ivete-sangalo/ | Se Eu NÃ£o Te Amasse Tantc | /ivete-sangalo/se-eu-nao-te-amasse-tanto-assin | Meu | pt |
| 4 | /ivete-sangalo/ | Quando A Chuva Passar | /ivete-sangalo/quando-a-chuva-passar.html | Quando a | pt |
| 5 | /ivete-sangalo/ | Sorte Grande | /ivete-sangalo/sorte-grande.html | A minha sorte | pt |
| 6 | /ivete-sangalo/ | A Lua Q Eu T Dei | /ivete-sangalo/a-lua-q-eu-t-dei.html | Posso te falar | pt |
| 7 | /ivete-sangalo/ | Mulheres NÃ£o TÃªm Que C | /ivete-sangalo/mulheres-nao-tem-que-chorar-cı | Hey, girl | pt |
| 8 | /ivete-sangalo/ | Eva / AlÃ´ PaixÃ£o / Beleza R | /ivete-sangalo/eva-alo-paixao-beleza-rara.html | "EVA" | pt |
| 9 | /ivete-sangalo/ | Flor do Reggae | /ivete-sangalo/flor-do-reggae.html | Um brilho de | pt |
| 10 | /ivete-sangalo/ | Carro Velho | /ivete-sangalo/carro-velho.html | Cheiro de | pt |
| 11 | /ivete-sangalo/ | NÃ£o Precisa Mudar | /ivete-sangalo/nao-precisa-mudar.html | NÃ£o precisa | pt |
| 12 | /ivete-sangalo/ | Nada Vai Nos Separar | /ivete-sangalo/nada-vai-nos-separar.html | Toda vez que | pt |
| 13 | /ivete-sangalo/ | Agora JÃ¡ Sei | /ivete-sangalo/agora-ja-sei.html | Duvidava, | pt |
| 14 | /ivete-sangalo/ | Deixo | /ivete-sangalo/deixo.html | Eu me lembro | pt |
| 15 | /ivete-sangalo/ | NÃ£o Me Conte Seus Proble | /ivete-sangalo/nao-me-conte-seus-problemas.h | Ivete | pt |
| 16 | /ivete-sangalo/ | PaÃ-s Tropical / ArerÃª / Taj | /ivete-sangalo/pais-tropical-arere-taj-mahal.htn | Moro... | pt |
| 17 | /ivete-sangalo/ | Na Bahia | /ivete-sangalo/na-bahia.html | Na Bahia ia ia | pt |

***Fig 3.1:*** *Samples from Lyrics-data file*

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | Artist | Genres | Songs | Popularity | Link |
| 2 | Ivete Sangalo | Pop; AxÃ©; RomÃ¢ntico | 313 | 4.4 | /ivete-sangalo/ |
| 3 | Chiclete com Banana | AxÃ© | 268 | 3.8 | /chiclete-com-banana/ |
| 4 | Banda Eva | AxÃ©; RomÃ¢ntico; Reggae | 215 | 2.3 | /banda-eva/ |
| 5 | Ã‰ O Tchan | AxÃ© | 129 | 1.6 | /e-o-tchan/ |
| 6 | Claudia Leitte | Pop; AxÃ©; RomÃ¢ntico | 167 | 1.5 | /claudia-leitte/ |
| 7 | Harmonia do Samba | AxÃ©; Samba; Pagode | 237 | 0.9 | /harmonia-do-samba/ |
| 8 | Ara Ketu | AxÃ©; Pop | 139 | 1.5 | /ara-ketu/ |
| 9 | Daniela Mercury | MPB; AxÃ© | 230 | 1.4 | /daniela-mercury/ |
| 10 | Olodum | AxÃ© | 74 | 1.3 | /olodum/ |
| 11 | Netinho | AxÃ© | 204 | 2 | /netinho/ |
| 12 | Asa de Ãguia | AxÃ©; ForrÃ³; RomÃ¢ntico | 267 | 1 | /asa-de-aguia/ |
| 13 | Cheiro de Amor | AxÃ©; RomÃ¢ntico; Chillout | 185 | 1 | /cheiro-de-amor/ |
| 14 | Timbalada | AxÃ© | 139 | 1 | /timbalada/ |
| 15 | Carlinhos Brown | AxÃ© | 111 | 1.4 | /carlinhos-brown/ |
| 16 | Tomate | RomÃ¢ntico; AxÃ©; Reggae | 78 | 0 | /tomate/ |
| 17 | Jammil e Uma Noites | AxÃ©; RomÃ¢ntico; Trilha Sonora | 141 | 1 | /jammil-e-uma-noites/ |

***Fig 3.2:*** *Samples from Artists -data file*

# CHAPTER 4

# DATA PRE-PROCESSING

## 4.1 Feature Selection

To select the required features for our prediction model, we gathered information from both of the csv files from kaggle dataset. For this, we created a new file of our own and combined both the lyrics-data and artist-data files. In this, first, we have eliminated all the data columns containing other languages and selected only English songs. Then we fixated on 3 genres which include Rock, Pop and Hip-hop. So, all the other genres songs are removed. For each artist, we mapped the respective genre and lyrics. As our project outline is to classify based on the lyrics of the song, we do not require the artist name, song name, artist link and song link. Hence, these four columns are removed from our dataset. So, finally we ended up with two columns in our dataset. First column is the genre of the song, and second column is the lyrics of the song. Figure 4.1.1 shows the final dataset with selected features.

| | A | B |
|---|---|---|
| 1 | Genre | Lyrics |
| 2 | Rock | I could feel at the time. There was no way of knowing. Fallen leaves in the night. Who can say where they're blowing. As free as the wind. Hopefully learning. Why the sea on the tide. Has no way of turnin |
| 3 | Rock | Take me now, baby, here as I am. Hold me close, and try and understand. Desire is hunger is the fire I breathe. Love is a banquet on which we feed. Come on now, try and understand. The way I feel under |
| 4 | Rock | These are. These are days you'll remember. Never before and never since, I promise. Will the whole world be warm as this. And as you feel it,. You'll know it's true. That you - you are blessed and lucky. It's |
| 5 | Rock | A lie to say, "O my mountain has coal veins and beds to dig.. 5RockRock men with axes and they all dig for me." A lie to ssay, "O my. river where mant fish do swim, half of the catch is mine when you haul. |
| 6 | Rock | Trudging slowly over wet sand. Back to the bench where your clothes were stolen. This is a coastal town. That they forgot to close down. Armagedon - come armagedon come armagedon come. Everyday is |
| 7 | Rock | Don't talk, I will listen. Don't talk, you keep your distance. For I'd rather hear some truth tonight. Than entertain your lies,. So take you poison silently. Let me be let me close my eyes. Don't talk, I'll believe |
| 8 | Rock | Well they left then in the morning, a hundred pairs of wings. In the light moved together in the colors of the morning. I looked to the clouds in the cirrus sky and they'd gone.. Across the marshes, across th |
| 9 | Rock | [ music: Dennis Drew/lyric: Natalie Merchant ]. . science. is truth for life. watch religion fall obsolete. science. will be truth for life. technology as nature. science. truth for life. in fortran tongue the. answe |
| 10 | Rock | On bended kneeI've looked through every window then.. Touched the bottom, the night a sleepless day instead. A day when love came,came easy like what's lost now found.. Beneath a blinding light that |
| 11 | Rock | For whom do the bells toll. When sentenced to die. The stuttering rifles. Will stifle the cry. The monstrous anger. The fear's rapid rattle. A desert inferno. Kids dying like cattle. Don't tell me. We're not pre |
| 12 | Rock | She walks alone on the brick lane,. the breeze is blowing.. A year had changed her forever,. just like her grey home.. He used to live so close here,. we'd look for places I can't remember.. The world was sa |
| 13 | Rock | Jenny. Jenny you don't know the nights I hide. below a second story room. to whistle you down. the man who's let to divvy up. time is a miser. he's got a silver coin. only lets it shine for hours. while you sl |
| 14 | Rock | You were looking away from me, western skies calling you.. Colors spilling, running dazzling you.. I was looking the other way, voices call from the east,. I saw my roots of the trees there planted at my feet |
| 15 | Rock | The color of the sky as far as I can see is coal grey.. Lift my head from the pillow and then fall again.. With a shiver in my bones just thinking about the weather.. A quiver in my lips as if I might cry.. Well by |
| 16 | Rock | Oh,. Baby blankets and. Baby shoes,. Baby slippers,. Baby spoons,. Walls of baby blue. Dream child in my head. Is a nightmare born. In a borrowed bed. Now I know. Lightning strikes again. It struck me once |
| 17 | Rock | the legs of Maddox kitchen tables. my whole life twisted on a lathe. in a foreman's torrent. my first English was. "faster boy if you want your pay". barking commands. loud and simple. we could all obey. th |
| 18 | Rock | Steep is the water tower. painted off blue to match the sky. can't ignore the train. night walks in the valley silent. you could swear the earth just moved. can't ignore the train. dust to be kicked up. in the c |
| 19 | Rock | [original version by Lulu]. Those schoolgirl days of telling tales and biting nails are gone. But in my mind I know they will still live on and on. But how do you thank someone who has taken you from crayon |
| 20 | Rock | That summer fields grew high with foxglove stalks and ivy.. Wild apple blossoms everywhere.. Emerald green like none I have seen apart from dreams that escape me.. There was no girl as warm as you.. H |

***Fig 4.1:*** *Finalized Dataset samples*

## 4.2 Stop Words Removal

In Natural Language Processing, stop words filtering is done to remove the excess information which does not add much value to the text. These are the fillers used in English language to make sentences grammatically accurate, but they don't provide any information while processing the text through the model. These stop words are most common in any language, they may include pronouns, conjunctions, prepositions, articles, etc. In our code, we included a few lines to remove the stop words of English from our lyrics dataset to help boost the learning process of the model.

## 4.3 Lower Case Conversion

After the stop words removal, we convert the whole lyrical data into lower case. The dataset contains different sentences and words with both upper and lower case. This is usually redundant for training the model. Hence, the technique of converting upper case into lower case letters is done to process the information and parsing. Additionally, when all the words of dataset are in lower case, it can help to understand a word's syntactic role, in later stages of the NLP model.

## 4.4 Padding the Sentences

In general, all the songs are usually not of same length. So, they number of words in each song also varies. The lyrics in our dataset are in the form of continuous sentences. But, for processing the data in the model, all sentences must be of equal number of words. Hence, using a piece of code, we find the lyrical column with maximum number of words in it, which is 231 words. So, all the remaining sentences are padded with zero to make them of equal length before further processing of the data.

## 4.5 Tokenization

In Natural Language Processing, tokenization is done to split the paragraphs and sentences into smaller units to effortlessly understand the meaning and process it. This process allows the model to understand the whole meaning behind the sentences by reading individual parts. Hence, it is an important step in the learning process of a model. There are different types of tokenization techniques, such as word tokenization, character tokenization, sub-word tokenization, etc. In our dataset, we are applying the word tokenization to split the lyrics into individual words.

# CHAPTER 5
# METHODOLOGY

## 5.1 USE

Universal Sentence Encoder (USE) model encodes the text data into high dimensional vectors which are known as embeddings. Embeddings are the numerical representations of the text data [12]. This is mostly used for NLP tasks such as text classification, semantic similarity etc. The pre-trained USE model is publicly available in tensorflow-hub. It is trained on various data sources such as Wikipedia, web news, discussion forms etc. The input for this model is variable-length English text and the output is a 512 dimension vector. It has shown a good performance on semantic textual similarity. It comes with two variations one is a transformer encoder model which has a higher accuracy and is computationally more expensive. While the other model is trained with a Deep Averaging Network which is computationally less expensive and has a little lesser accuracy. Usually training the embeddings require a large amount of data. In this scenario, Pre-trained sentence embeddings have been proven much useful for various NLP tasks.

## 5.2 BERT

BERT (Bidirectional Encoder Representation from Transformers) is a language representation model developed by Google AI Language. This model is developed in contrast to the previous efforts of embedding which look into the sequence from either left to right or combined left-to-right and right-to-left training [13]. It is observed that a language model which is bidirectionally trained has a very good knowledge of the context and flow than the single direction language models. BERT makes use of transformer model which learns the contextual relationship between words in a text. In its vanilla form, Transformer includes both encoder and decoder mechanisms. Since BERT's goal is to generate a language model only the encoder mechanism is used. Here the Transformer encoder reads the entire sequence of words at once. This is what makes the model learn the context of word based on the words in its left and right side.

In this model the researchers used a technique MaskedLM (MLM) which allows Bidirectional training in models in which it was previously not possible. Before feeding the sequence of words to BERT, 15% of the words in every sequence are replaced with a MASK token. The model aims to predict the initial value of the masked words based on the information provided by the non-masked words. BERT's loss function only considers the prediction of masked values but not the non-masked words. So, the model converges slower when compared with the directional models but still bidirectional training outperforms the traditional single directional language models.

## 5.3 LSTM

The standard RNN Model has a input layer, hidden layer and output layer. RNN Models make use of recurrent units in its hidden layer to process the sequence data. This is done by recurrently so these models suffer from Vanishing Gradient Problem. Recurrent Units are also used in LSTMs but what happens in the unit is different in the two models. Unlike the standard feed forward neural networks LSTM has feedback connection. So this can not only process single data points but also the entire sequence of data. LSTM architecture aims to provide short term memory for RNN which can last for thousands of steps. The recurrent unit in LSTM is much more complex than the one in RNN which improves learning but more computational resources [14]

LSTM module has a cell state and three gates which provides the cell states to learn, unlearn and retain information from every unit. This model helps the information to flow through units without being altered and allows only a few linear interactions. Each unit has an input output and a forget gate which has a capability to add or remove the information to the cell state. The input gate controls the flow of information to a cell state by using a point-wise multiplication operation of sigmoid and tanh. The forget gate the information from the previous cell state to be forgotten for which it uses a sigmoid function. The output gate decides which information should be sent to the next hidden stage.

# CHAPTER 6
# RESULT ANALYSIS

To develop a model with best accuracy, we have used two different language representation models Universal Sentence Encoder (USE) and Bidirectional Encoder Representation from Transformers (BERT). We chose Long Short Term Memory (LSTM) model for classification purpose. We trained the Language representation models USE and BERT over five folds to determine the best accuracy.

- As shown in Table 2 we have tabulated the best accuracy in each fold for both USE and BERT models.
- We can observe that the best accuracy with USE-LSTM model is 83.42% in Fold 2 and for BERT-LSTM model the best accuracy id 76.55% in second Fold.

***Table 6.1:*** *Table of Accuracies*

| | **USE** | **BERT** |
|---|---|---|
| **Fold 1** | 79.23% | 64.49% |
| **Fold 2** | 83.42% | 76.55% |
| **Fold 3** | 79.4% | 74.54% |
| **Fold 4** | 80.23% | 75.38% |
| **Fold 5** | 79.23% | 64.15% |

- As displayed in Table 3 we have also analyzed the total parameters for both models. We observed that USE has more parameters than BERT.
- Here we can conclude that even though BERT has less parameters than USE the complexity of USE enables it to learn efficiently.

***Table 6.2:*** *Table of Parameters*

| | **USE** | **BERT** |
|---|---|---|
| **Total Parameters** | 142,851,170 | 36,406,370 |
| **Trainable Parameters** | 142,851,170 | 36,406,370 |

- In Fig 6.1 and Fig 6.2 we can view the training and testing loss of both USE and BERT models.

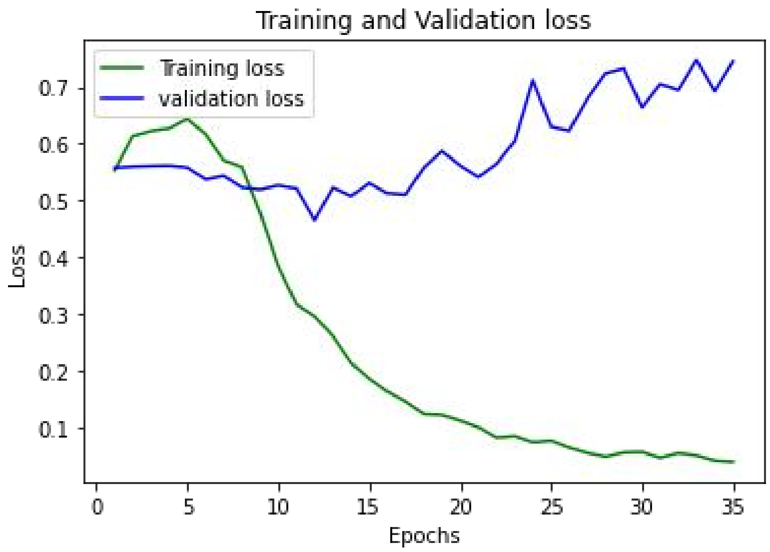


***Fig 6.1:*** *Training and Validation loss of USE*

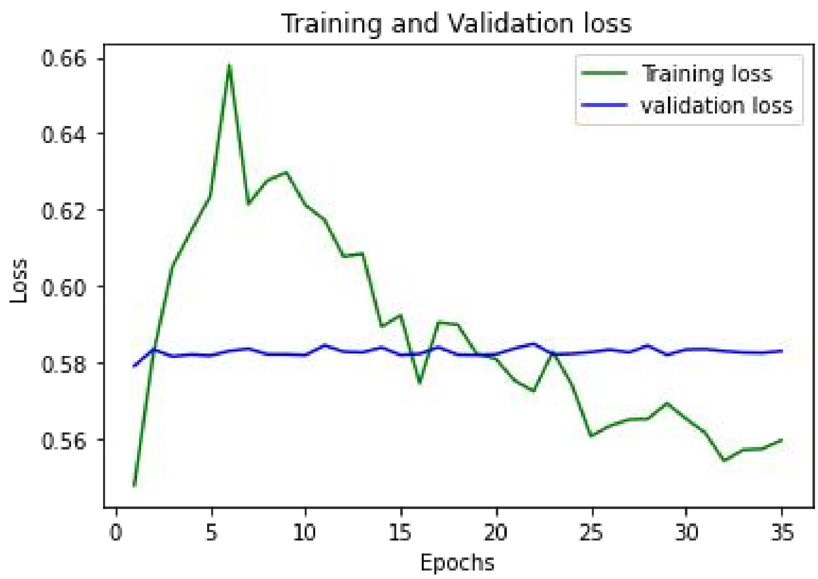


***Fig 6.2:*** *Training and Validation loss of BERT*

# CHAPTER 7

# CONCLUSION

In Conclusion, in this paper compared various language models to determine the best model. Here we have developed a song genre classification model suing lyrics with Language Representation models and Recurrent Neural Networks with a highest accuracy of 83.42%. This model will provide an automatic genre classification to manage the songs according to its genre in music platforms.

# REFERENCES


**Language Representation Models for Music Genre Classification Using Lyrics**

[1] Akalp, Hasan, et al. "Language representation models for music genre classification using lyrics." 2021 International Symposium on Electrical, Electronics and Information Engineering. 2021.

**Genre Classification using Feature Extraction and Deep Learning**

[2] A. Kumar, A. Rajpal and D. Rathore, "Genre Classification using Feature Extraction and Deep Learning Techniques," 2018 10th International Conference on Knowledge and Systems Engineering (KSE), 2018, pp. 175-180, doi: 10.1109/KSE.2018.8573325.

**Music Genre Classification with Transformer Classifier**

[3] Zhuang, Y., Chen, Y., & Zheng, J. (2020, June). Music genre classification with transformer classifier. In Proceedings of the 2020 4th International Conference on Digital Signal Processing (pp. 155-159).

**Brazilian lyrics-based music genre classification using a blstm network**

[4] Araújo Lima, Raul de, et al. "Brazilian lyrics-based music genre classification using a BLSTM network." International Conference on Artificial Intelligence and Soft Computing. Springer, Cham, 2020.

**Genre Classification using Word Embeddings and Deep Learning**

[5] Kumar, Akshi, Arjun Rajpal, and Dushyant Rathore. "Genre classification using word embeddings and deep learning." In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2142-2146. IEEE, 2018.

**On Combining Diverse Models for Lyrics-Based Music Genre Classification**

[6] Ueno, C. L. R. S., & Silva, D. F. (2019, October). On Combining Diverse Models for Lyrics-Based Music Genre Classification. In *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)* (pp. 138-143). IEEE.

**Representation Learning vs. Handcrafted Features for Music Genre Classification**

[7] Pereira, Rodolfo M., et al. "Representation learning vs. Handcrafted features for music genre classification." *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019.

**Effective Training Methods for Automatic Musical Genre Classification**

[8] Atsız, E., Albey, E., & Kayış, E. (2019). Effective training methods for automatic musical genre classification. In *DATA 2019-Proceedings of the 8th International Conference on Data Science, Technology and Applications*. SciTePress.

**Multi-Emotion Classification for Song Lyrics**

[9] Edmonds, D., & Sedoc, J. (2021, April). Multi-emotion classification for song lyrics. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 221-235).

**Using Deep Learning to Predict Music Genre from Song Lyrics**

[10] Mubeen, M. Using Deep Learning to Predict Music Genre from Song Lyrics.

**Dataset**

[11] https://www.kaggle.com/datasets/neisse/scrapped-lyrics-from-6-genres?select=lyrics-data.csv

**USE**

[12] https://www.analyticsinsight.net/googles-universal-sentence-encoder-would-revolutionize-the-application-of-neural network/#:~:text=The%20universal%20sentence%20encoder%20model,%2C%20semantic%20similarity%2C%20and%20clustering

**BERT Model**

[13]https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270

**LSTM Model**

[14]https://towardsdatascience.com/lstm-recurrent-neural-networks-how-to-teach-a-network-to-remember-the-past-55e54c2ff22e