Emotion Analysis of Pages in a Book to Play Background Music

MSc Research Project
Data Analytics

Joshma Joseph
Student ID: X20206968

School of Computing
National College of Ireland

Supervisor: Hicham Rifai

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Joshma Joseph
Student ID:	X20206968
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Hicham Rifai
Submission Due Date:	15/08/2022
Project Title:	Emotion Analysis of Pages in a Book to Play Background Music
Word Count:	6124
Page Count:	18

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	10th August 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Emotion Analysis of Pages in a Book to Play Background Music

Joshma Joseph
X20206968

Abstract

The approach of natural language processing, which has been extensively studied, includes recognizing emotions within texts. To make pertinent product and response recommendations to users, it may be necessary to extract human emotions or thoughts from these text. In this paper we will research on how we can analyse the emotions of pages in a book using social media data-set and play background music related to that emotion. This approach can build the mood for the chapters in the book and depict the authors emotion. With this strategy, the reader will be able to fully immerse themselves in the novel-reading experience and gain a deeper understanding of the author's feelings through music. With the help of various Machine Learning and Deep Learning models, such as Multinomial Nave Bayes and Recurrent Neural Network-based approach, which will be applied to the word vectors produced in the feature extraction process and the padded phrases, the classification and identification of emotions are carried out in this study. In recognizing the emotions in the book's pages, the Multinomial Naive Bayes approach had an accuracy of 62 percent and the Recurrent Neural Network methodology had an accuracy of 82 percent. The precise identification of the emotions enabled the playing of appropriate background music from a library that was specially curated with songs for various emotions.

1 Introduction

Aside from words, gestures, and facial and vocal emotions, there are many more methods to show emotion. Everyday existence is significantly influenced by human emotions. Intuition that is separate from intelligence or comprehension is what is referred to as emotion. A person's ability to consider many scenarios and adapt their behavior to rewards is impacted by emotion. Emotional acceptance is used in many industries, including medicine, law, advertising, and e-learning. Another crucial component of sophisticated human communication is the ability to describe emotions. A person's voice, gestures, and writing may all be used to psychologically portray speech, appearance, and text emotion. Although substantial work has gone into speech and face expression recognition, a framework for text-based emotion identification is still lacking. From a data analysis perspective, language modelling makes recognizing human emotions in a text quite helpful. Emotions like happiness, rage, contempt, fear and other things are shown. Xu et al. (2020). With the ever-evolving generations, the conversations amongst people have become vastly different than they used to be earlier. The perspective of people has been different on different topics now. The up-to-date platform that depicts what people in the current generation are biased towards is social media platforms like Twitter. Hence use of this Twitter data to identify and classify the emotions in different texts can be more beneficial to the current trends of people's thinking.

1.1 Background and Motivation

Books have always been a source of knowledge since they included the experiences and tales of people from various backgrounds. It served as a forum for the sharing of life experiences by people from all walks of life. However, fewer people currently read books due to people's hectic lifestyles. Audio-books have been offered as alternatives to reading books, but the artificial voice of an audio-book does not capture the true essence of the work. Providing background music in flow with the emotions of the book will help readers better grasp each chapter of the book. In this way, it would be possible to enhance a reader's reading experience. This strategy focuses mostly on maintaining the reader's interest and motivating them to finish the book. Computing devices nowadays are clever in a number of ways, including the capacity to recognize speech, gauge the strength of a keyboard, predict the next word as you type, and detect body position. However, the human emotion associated with or buried inside each piece of information that modern computing devices get is still not detectable. It is possible to minimize the semantics of the interaction that occurs between humans and computing technologies by giving them the capacity to sense human emotion. This will enable modern data processing equipment to react to orders adequately, if not more sympathetically. Therefore, effective computing is the use of pattern recognition to enable emotionally intelligent machines. There are several ways to identify emotions, including facial expressions, hand gestures, body language, blood pressure, heart rate, body temperature, and voice tone, to name a few.

Machine learning (ML) techniques can assist in identifying concealed emotions in writing. The users can receive suitable recommendations or suggestions based on the observed emotions. Machine learning and data analytics are frequently used to make machines think more like humans. Machines can now analyse human emotions and predict human responses to a circumstance thanks to advances in data analytics and machine learning algorithms. Human behaviour has been studied in a variety of ways in the past, but sentiments discovered in literature or novels have not yet been thoroughly investigated. In order to accomplish this, we use Natural Language Processing to decipher the sentiments included inside a specific page of a novel. Once the feelings are identified, an emotion-based music recommendation system will make suggestions or proposals for music that will enhance the reading experience and compliment the emotions of the text.

1.2 Research question

Can machine learning and deep learning models trained from social media data-sets be used to classify the emotion of pages in a book and consequently play relevant background music?

For this research the data-sets used consists of twitter data and dialogues from the famous sitcom FRIENDS which have been classified into 8 different emotions. A free e-book from Gutenberg library [1] will be used for testing the emotion analysis model and free sample music will be used to be played as background music.

[1]Project Gutenberg: `https://www.gutenberg.org/`

The following is a description of the report's structure. In part 2, a survey of related literature is conducted in this field of study. Section 3 then elaborates on the the methodology of the research. Section 3 is further divided into five parts, each of which discusses a different aspect of business understanding, data understanding, or data preparation techniques. The design specification portion is covered in Section 4, which is followed by Section 5 on model implementation and Section 6 on model assessment. Finally, the report's conclusion and future work are discussed in section 7, and references are provided in the next section.

2 Related Work

This part will go over earlier studies on how to analyze a text's emotions. The methods mentioned below were mostly used to analyze the feelings and sentiments expressed in social media texts in order to comprehend the general feelings of the population about a certain subject. The past several years have seen a significant amount of study in this area. In order to recognize people's emotions from their social media interactions, several academics have employed a variety of methods and algorithms. We will talk about a few research that has been done recently in the field of emotion analysis.

2.1 Natural Language Processing for Emotion Analysis

(Gosai et al.; 2018) advocated the use of natural language processing in this study to identify emotions in texts. The input text is categorized into emotions using this method by looking for emotional components such as adjectives, verbs, phrases, adverbs, or combinations. After the text has undergone pre-processing, which entails eliminating punctuation, repetitive characters, stop words, as well as stemming and lemmatizing the texts, the data is specified in a data dictionary that comprises six kinds of emotions. The words are tokenized and tagged using the NLP tool kit (NLTK), which is then utilized to calculate the sentiment metrics.

For the purpose of detecting emotions in an English text,(Gaind et al.; 2019) recommended combining two alternative techniques. This essay looked at how emotions may be divided into six groups: happiness, surprise, anger, disgust, fear, and sadness. The study's first proposal used elements including degree words, emoticons, and negations and was based on a model of natural language processing. The second strategy was using machine learning algorithms to categorize emotions. The researchers were able to create an emotion word bag automatically for each block of text (EWS). They were also able to create a labeled training set using a keyword matching strategy to train the different classifiers.

A complete framework for emotion modeling and analysis was proposed in the research article by (Adikari et al.; 2021), which was founded on the ideas of self-structuring artificial intelligence. This method used modeling approaches to social media's unstructured and unlabeled data in a methodical way. The methodology included emotion extraction based on Pluchik's psychological model, the development of a curated emotion vocabulary, the use of NLP techniques to mine emotional expressions from textual content, the calculation of emotion intensity and emotion state for each created profile, and emotion classification followed by topic modeling-based granular exploration. Deep emotions were particularly examined and simulated by the researchers employing cutting-edge word embedding, Markov chains, natural language processing, and developing self-organizing maps.

2.2 Use of different datasets for Emotion Analysis

The multi-label classification approach, which falls under sentiment analysis, is examined in (Almeida et al.; 2018)'s study for diverse cognitive stances. To find a wide range of multi-label solutions, they used several strategies such as algorithm adaptation, issue transformation, and ensemble techniques. For this investigation, two datasets were employed, one consisted of entertainment news articles written in Brazilian Portuguese and acquired from Buzzfeed and other online news sources in Brazil. Different multi-label algorithms were evaluated for the sentiment and emotion analysis of the datasets, and the ensemble technique outperformed the neural network-based approaches for ranking-based metrics.

For this study, (Sailunaz and Alhajj; 2019) gathered information from tweets and responses on specific subjects on Twitter in order to identify and analyze the attitudes and emotions individuals exhibited. The influential scores of the discovered attitudes and emotions were calculated based on a variety of user- and tweet-based characteristics. The users were then given a customised suggestion featuring a list of people who had exhibited similar interests and agreed on the same subject using these influential ratings. The Nave Bayes algorithm was employed to categorize the users' various emotions and attitudes.Through the investigation, it was shown that the accuracy level declined as the number of classes rose; as a result, sentiment classifiers with just three classes were more accurate than emotion classifiers with seven classes.

In-text characteristics were used by (Halim et al.; 2020) to recognize emotions in short texts and create datasets for the purpose. A benchmark dataset and a local dataset were included in the three datasets used for the investigation. Several machine learning and deep learning techniques, including the Artificial Neural Network (ANN), Support Vector Machine (SVM), and Random Forest, were used to classify emotions (RF). Evaluation of these approaches' precision revealed that the recommended framework produced superior outcomes with an average accuracy of 83 percent.

In this study, (Asif et al.; 2020) provided an overview of the various tools and techniques used to categorize the feelings and sentiments in a text. Future research in the area of sentiment and emotion analysis of texts can benefit from this study. The researchers have provided a description of all the tools, methods, and datasets utilized, particularly for the Urdu language, to analyze the emotions in a text. The six fundamental emotions—surprise, fear, happiness, anger, disgust, and sadness—were the focus of the study. This study explored a wide range of strategies for the analysis of emotions, including approaches based on keywords, lexicons, and machine learning algorithms.

Ameer et al. (2022) recognized all potential emotions in a given text in their study work. This study provided a sizable benchmark corpus for a multi-label emotion classification task for code-mixed SMS messages, manually annotated using twelve sets of emotions like fear, love, anticipation, and optimism. These emotions included disgust, anger, sadness, trust, joy, pessimism, and surprise. After creating a sizable standard benchmark code-mixed corpus, the researchers performed a variety of deep learning, traditional machine learning (content-based approaches), and cutting-edge transfer learning techniques on the suggested corpus. They noticed that using traditional machine learning techniques along with OVR multi-label and SVC single label produced the greatest results.

2.3 Deep Learning approach for Emotion Analysis

For the purpose of classifying emotions through text,(Su et al.; 2018) created a database of seven different emotions, including disgust, melancholy, anxiety, happiness, boredom, surprise, and rage. Using the input text, a semantic word vector was produced using the word2vec model. In order to create a vector of emotion words, each lexical term is also connected to every emotion word specified in an affective lexicon. An autoencoder is used to obtain bottleneck features for dimensionality reduction. In order to categorize the text's emotions, the Long-Short Term Memory (LSTM) methodology, an expansion of the RNN method, is utilized. When compared to the CNN technique, this method's accuracy in identifying emotions was 5.33 percent higher, coming up at 70.66 percent. Additionally, this model included semantic and emotional word vectors, which was a superior method than using separate feature vectors.

In this research by (Singh et al.; 2022), a deep learning method for sentiment analysis of evaluations of the COVID-19 on Twitter is presented. The suggested approach uses increased featured weighting by attention layers and an LSTM-RNN-based network as its foundation. Through the attention mechanism, this algorithm makes advantage of an improved feature transformation framework. In this investigation, four class labels from publicly accessible Twitter data stored in the Kaggle database—sad, pleasure, fear, and anger—were employed. The suggested deep learning strategy considerably improved the performance metrics based on the usage of attention layers with the existing LSTM-RNN approach, with a gain of 20% in accuracy and 10% to 12% in precision but only 12-13% in recall when compared with the present approaches. Positive, neutral, and negative tweets made up 45 percent, 30 percent, and 25 percent, respectively, of the 179,108 tweets that were connected to COVID-19. This demonstrates how effective and practical the suggested deep learning strategy is and how simple it is to apply for sentiment categorization of COVID-19 reviews.

For Natural Language Processing, several Deep Learning methods have been employed (NLP) in this paper by (Raza et al.; 2022). Finding the ideal algorithm is essential for effective sentiment management in the Airbnb business. In order to distinguish between various elements of guest reviews and assess their accuracy, the article employs a number of Deep Learning algorithms. Results are analyzed using four accuracy measurement benchmarks: Precision, Recall, F1-score, and Support. Comparing the GRU approach to RNN and LSTM, the study reveals that it produces the best results with the greatest classification metrics values.

This study by (Mohanty and Mohanty; 2022) uses a deep recurrent neural network to interpret spoken input to determine how people feel (RNN). The work's current status of development is mentioned first. There are two major feelings and two submajor sentiments that are determined. Some of the variations have recordings of people in various emotional states. These are thoroughly examined as opposed to only examining the good and negative possibilities. Different spectrum approaches are utilized to generate features because spectral features are effective at doing so. Along with Mel-frequency cepstral coefficients, other factors are taken into account, such as segmental characteristics as speech components. For the classification job, neural network-based models are put to the test. Last but not least, a deep RNN is employed, assessed, and determined to be superior to other techniques that can result in voice data mining. The stochastic gradient descent approach is used to train RNN using these input characteristics. The outcomes demonstrate that the efficacy is more than that which was previously reported in works.

2.4 Summary

To summarize the previous researches performed on this topic, we can see that various data-sets have been used and analysed for the identification of emotions in day-to-day conversations. We see that data-sets like news headlines, dialogues, tweets and emotion intensity data have been used for emotion recognition. Hence to better understand human emotions for this research we have used a combination of data-sets consisting both social media messages i.e twitter data and daily dialogues liked dialogues from the sitcom FRIENDS. The Naive Bayes and RNN model have been widely used throughout the researches and have proven to be successful in accurately identifying the emotions within short-texts. Identifying emotions within texts in a book using social media data-sets and playing background music based on the identified emotion is the novel approach of this research.

3 Methodology

The method of deciphering various human emotions in each text is known as emotion analysis. In this research, we'll concentrate on analyzing eight fundamental human emotions: joy, sadness, fear, surprise, anger, shame, disgust, and neutral (no emotion). Since there are no existing data sets that classify the pages according to emotions, we will combine publicly accessible social media data sets in order to get better recognition rate.The database's classified collection of music will be played based on the emotions identified. For emotion classification we will use the natural language processing methodology. In the given research of emotion analysis of pages in a book for playing background music, the entire data mining process is organized into a series of phases based on CRISP-DM technique. Figure 1 presents each of these phases in detail.

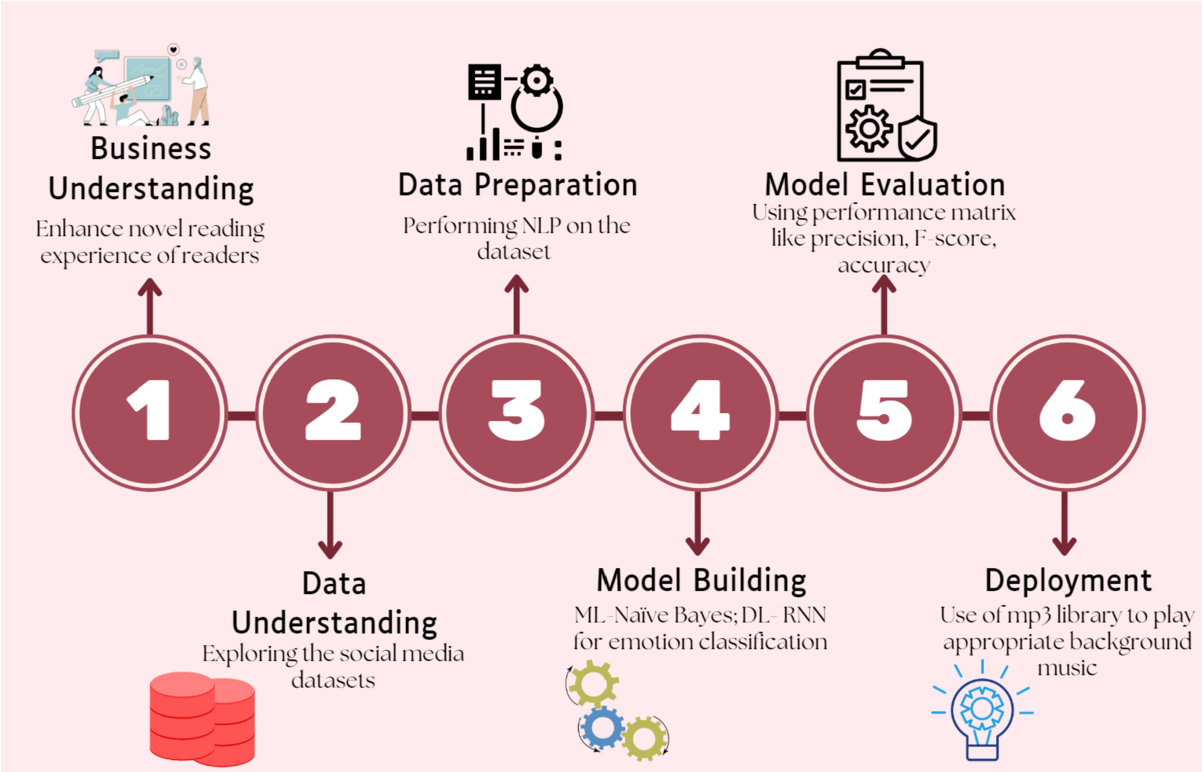

Figure 1: CRISP-DM phases

3.1 Business Understanding

As stated in the preceding sections, the main objective of this project is to increase interest in reading books and in understanding and sharing the author's feelings through music.We will be employing background music to encourage individuals to read more and make the reading experience delightful because music has a significant influence in evoking emotions. It is necessary to analyze the chapter's emotions before recommending any music. In order to keep readers interested in reading the novel, we will use this study to identify emotions like joy, fear, etc. in a chapter of the book and offer background music for each chapter. The author's portrayal of real emotions in the book may be experienced by the reader through music. Additionally, this enables people to read a chapter from an old book while reestablishing their musical connection to the narrative. This research will make use of social media data-sets, and hence the machine learning and deep learning models will be trained with the current trends and sensitive topics to people predicting emotions in a chapter of a book.

3.2 Data Understanding

A social media data repository along with a data-set including the dialogues from the series FRIENDS is examined in the research that is being presented. Textual information is accessible as a comma-separated file (.csv) with various attributes pertaining to each distinct scenario. The data needed for the study is obtained from a number of sources that are detailed below.

- Social media data-set [2]

  This data collected from the Kaggle website consists of social media interactions or tweets. It consists of data labelled into 7 emotions namely joy, surprise, disgust, anger, fear, sadness and neutral. There are more than 30,000 tweets which have been labelled into different emotions in this data-set.

- Emotion Line data-set [3]

  (Ekman; 1999)'s fundamental emotions and a neutral class are used to categorize the data in the Emotion Lines data set. This information was gathered from Facebook Messenger conversations and lines from the Friends comedy series. There are more than 10,000 dialogues and 29,245 utterances in the data. This data has been taken from the years 2018 and 2019. The data-set was available upon online registration on the website.

Both the data-sets are combined together to form a heterogeneous mixture of data labelled with 7 different emotions. This kind of data-set will allow us to better train the model since it has a wide range of data sources. By using the social media interactions the trending conversations and sensitive topics which can peak people's emotions can be easily identified. The use of Emotion line data-set allows the machine learning models to be trained to understand emotions within daily dialogues or interactions amongst people. The below Figure 2 depicts a bar graph of dialogues and tweets labelled into different emotions. As seen in the graph, the target variable's class distribution is imbalanced, with more records for neutral emotion accounting for around half of the overall count while disgust and shame emotions are represented by fewer records.

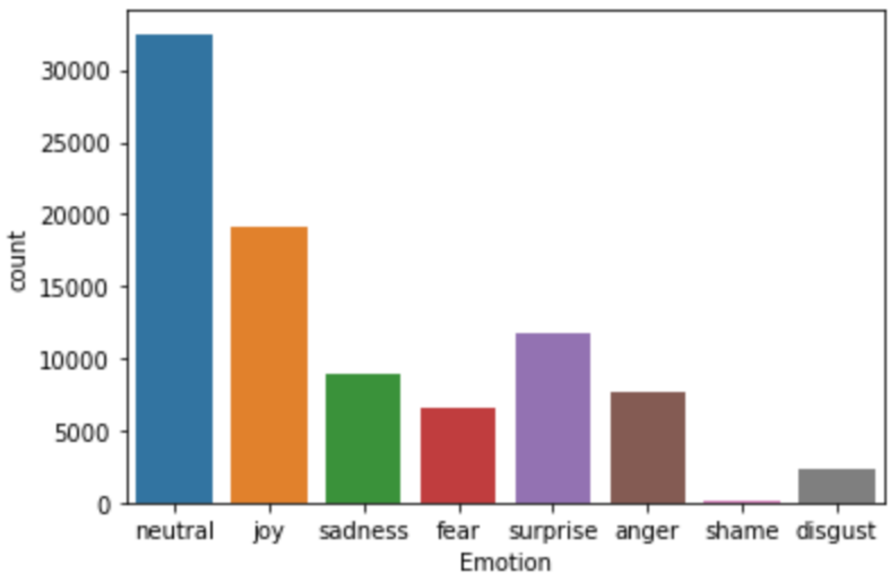

Figure 2: Class distribution of emotions

[2] Social media data-set: https://github.com/Jcharis/end2end-nlp-project/blob/main/data/emotion_dataset_2.csve

[3] Emotion Line data-set: https://sites.google.com/view/emotionx2019/datasets

3.3 Data Preparation

We must prepare the data before we can begin the model-training procedure. Since several data-sets were utilized for the research, the data must be cleansed and synchronized so that it only contains the eight emotions that will be examined: joy, sadness, fear, surprise, anger, shame, disgust, and neutrality. Final data-set will include texts and the relevant labelled emotion. These pre-processing steps need to be carried out for both the training data as well as the text from the pages of the book which needs to be analysed. By eliminating the unnecessary clauses from the phrases during the pre-processing stage, it will be simpler to determine the appropriate emotion in the book. As a result, the models' ability to accurately identify emotions will improve.

3.3.1 Natural Language Processing

Tokenization, punctuation removal, stop word removal, and lemmatization are the few pre-processing procedures that are described below as being a component of the natural language processing domain.

- Tokenization:

  Phrases, conversations, or tweets that are two or sometimes three sentences in length make up the training data. As a result, these sentences must be divided into tokens, which are shorter portions. In a similar way, a book's pages may include a number of sentences that are difficult to comprehend all at once. As a result, the tokenization procedure can facilitate a computer's interpretation of these sentences.

- Removing Punctuation:

  To further process the smaller parts or tokens of the phrases, the punctuation is removed after the tokenization process. In order to identify the emotions in the text, this step entails eliminating punctuation marks like full-stop, commas, question mark, etc. from the tokenized data.

- Removing stop words:

  Many words in the English language are there only to give a statement additional significance. This contains stop words, which are words like "and," "the," and so forth. Although these words give phrases more meaning, they provide very little information and are useless for identifying the emotions expressed in sentences. Therefore, after eliminating the punctuation from the tokenized data, we will eliminate the stop words from the tokenized data. The quality of the data will significantly improve as a result.

- Lemmatization:

  Following the stop words removal process, we'll try to identify the tokens' basic structure. We shall discover the origin of words throughout the lemmatization process. A lemma is another name for a word's fundamental unit. For instance, the terms try, trying, and tried all have the basic word try as their common origin. We locate every word and change it into its base form.

3.3.2 Feature Extraction/ Feature Engineering

Apart from the basic Natural Language processing steps for pre-processing the important parts of the sentences which will be beneficial for identifying and classifying emotion within them need to be extracted. This process is called called feature extraction. Based on the model used for the analysis the feature extraction process differs. We will be training and using two models for our emotion classification namely the Naive Bayes and Recurrent Neural Networks. The feature extraction/ feature engineering process for both these models will be explained further:

- Naive Model:

  The feature extraction method that will be used for this model is the keyword extraction technique. In this process the most common words per class of emotion is extracted. For example, if we deep dive into the emotion 'joy' in the data-set of our research, we can observe that the words 'Oh', 'like' and 'love' are most common. Similarly for all the 8 emotions in our analysis we will identify the most common words for these emotions to train our machine learning model. The below Figure 3 shows the plot generated using WordCloud package in python depicting the most common words in the emotion 'joy' with the help of different text size corresponding to the occurrence of the words. This process is followed by converting these words into vectors or numerical format with the countVectorizer which converts the data into sparse matrix.

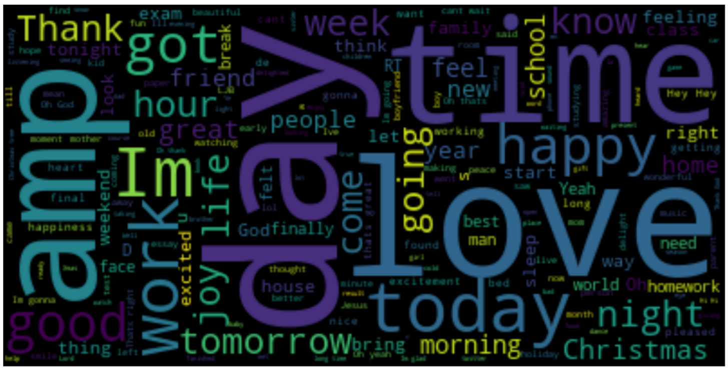

Figure 3: Keyword extracted in joy emotion

- Recurrent Neural Network (RNN):

  For the RNN model, the cleaned text need to be converted into numerical format for the model to understand and process the information. Hence for this purpose the cleaned text need to be initially converted to separate words and coded into numerical. One-Hot encoding methodology is used to encode the words into relevant numbers. The words are also padded further to keep them consistent with the required input for the model. The below Figure 4 depicts the process of one-hot coding and padding process in TensorFlow.

Figure 4: One Hot encoding technique

4 Design Specification

The design architecture of our system for analyzing emotions is demonstrated in this section. The processes shown in the graphic are followed to execute the text emotion categorization of a book page. The Figure 5 below depicts the research work's flow diagram. Initially the raw data is pre-processed with the steps as explained in the previous sections followed by the process of feature extraction—keyword extraction for the Naive Bayes model and word encoding for the RNN model—is performed. Naive Bayes and RNN models are consequently trained and built for decision-making. Then the pages of a book are provided as separate input back to back and the text within them are pre-processed to meet the requirements of the input data for the models. The process of emotion analysis is performed on these texts.The identified emotion is then used to play appropriate music from a emotion labelled music library.

This architecture can be converted into layers during deployment like database layer, application layer and presentation layer. The database layer will consist of the data pre-processing and the data transformation part. The application layer will consist of the model and its evaluation to identify the best suited for the analysis. Finally the application layer will consist of the actual output which includes the identified emotion along with the background music.

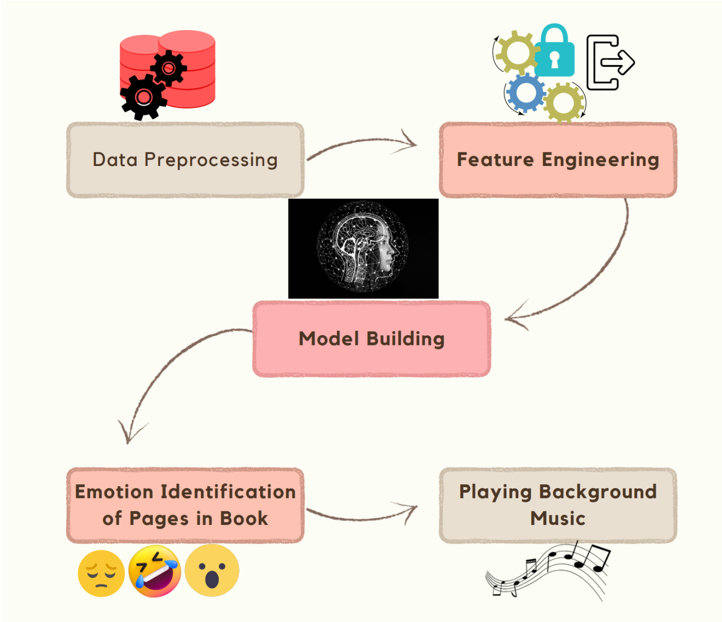

Figure 5: Emotion identification and music playing flow

5 Implementation

In this section we will explain the model building and implementation of our emotion analysis system.A machine learning model Naive Bayes and a deep learning model will be trained and built for the emotion analysis. Further we will also elaborate the playing of background music for the identified emotion.

5.1 Naïve Bayes Model Building

The Naïve Bayes classifier is appropriate for large amounts of data and is based on the Bayes theorem. Less training data is needed for this classifier. A probabilistic model called Naïve Bayes uses naive assumptions to assign data to a certain category or class.This Bayes Theorem is shown in the below formula.

$$P(\theta|\mathbf{D}) = P(\theta)\frac{P(\mathbf{D}|\theta)}{P(\mathbf{D})}$$

where,

$P(\theta)$ is probability of the hypothesis,
$P(\mathbf{D}|\theta)$ is the probability that given the hypothesis the data is correct
$P(\theta|\mathbf{D})$ is the probability that based on the data, the hypothesis is correct
$P(\mathbf{D})$ is the probability of data.

As we have to deal with multiple classes we have used Multinomial Naive Bayes solution. A document in the multinomial model is an organized list of word events taken from the same vocabulary. We presume that document durations are unrelated to a document's class. The probability of each word event in a document is independent of the word's context and position in the document, which is a similar Naive Bayes assumption that we make once again. Each document is therefore selected using a multinomial distribution of words and as many independent trials as there are words in the text. This results in the well-known portrayal of papers as a ”bag of words.” [4]. To train the model and evaluate its effectiveness, the cleaned data is separated into test and train data. The Naïve model is given the train data as input. The cleaned vectorized text data and it’s labelled emotion is used to train the model. For test purpose the provided text data can be classified based on the words that are mostly occurring in a particular emotion class. Once the model is trained the trained Naïve Bayes model is used for the analysis of the texts present in d pages of the book.

5.2 RNN Model Building

RNN is a common neural network that examines input data in a solitary setting. It is sometimes referred to as a ”forward-pass” neural network. In other words, it forgets previous bits of data when training on a new piece of data. Having its own internal memory, it is a model. Iteratively processing the elements is how RNNs process sequence input. The outputs from one timestep are sent to RNNs’ inputs for the following timestep. Because they can examine each word in the context of the words that came before it, RNNs are excellent for NLP applications because they can extract more meaning from a phrase. The below figure shows the network of the RNN model of the research. The model consists of an embedding layer, a RNN layer along with dense layers as seen in the figure. A 64-neuron RNN layer is used to construct a sequential model. With the exception of the output layer’s seven output neurons, all of these layers are activated using the sigmoid activation function. It uses a recurrent dropout rate of 0.2. The Figure 6 below gives a summary of this built model.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 1160, 150)         1650000

 dropout (Dropout)           (None, 1160, 150)         0

 simple_rnn (SimpleRNN)      (None, 128)               35712

 dropout_1 (Dropout)         (None, 128)               0

 dense (Dense)               (None, 64)                8256

 dropout_2 (Dropout)         (None, 64)                0

 dense_1 (Dense)             (None, 8)                 520

=================================================================
Total params: 1,694,488
Trainable params: 1,694,488
Non-trainable params: 0
_________________________________________________________________
```

Figure 6: RNN Model network

5.3 Playing Background Music

Once the models are trained, the most accurate models are used for the identification of emotions in the pages of the book. The identified emotion is helpful is selecting the appropriate song for the music database of 8 songs corresponding to the 8 emotions. This

[4]MNB: `https://www.slideserve.com/lindaray/naive-bayes-for-document-classification-powerpoint-ppt-`

is done with the help of the playsound library of python which allows the use of mp3 player to play the relevant music.

6 Evaluation

In this section we will critically evaluate the models we used for our analysis and the experiments we performed. For the analysis we performed two experiments the first included emotion recognition of sample e-books downloaded from online websites and the second experiment included the use of a sample book for the analysing and the training the models. Based on the results or the identified emotion the appropriate music was played from the manually created library of sample music.

6.1 Experiment 1: Sample e-book 'Paying the price'

For the first experiment we used the e-book "Paying the price" available on the Gutenberg website as a sample book for the analysis. As a method of cross verifying the results. The book was read and each page was analysed, it was observed that even though there various emotions in the pages the most prominent of them was 'Neutral'. Hence to test if the models can accurately predict theses emotions, the models were tested on the book. As identified by reading the book, the models too recognized a neutral emotions throughout each page of the book.

6.2 Experiment 2: Manually created book for analysis

Since the previous experiment only had one emotion that was predicted, various other famous books available online were tested like "When Siggy met Phyllis" [5]. This time the models were able to predict more emotions like "Joy". So to produce better results a book that depicted different emotions was manually created. These book consisted various keywords that are associated with different emotions for people. This book when used as an test to validate our models depicted more varying emotion outputs.

6.3 Model evaluation using different metrics

Now that the emotions have been identified by the model, the models used for the research need to be evaluated further. We refer to different evaluation metrics like precision, recall, f1-score, accuracy and confusion matrix which will be explained below:

- Precision: Starting with precision, which provides the following information- we can determine what percentage of anticipated Positives are in fact Positive. When we wish to be absolutely certain about our forecast, precision is an acceptable option of assessment metric. Equation 2 below depicts the formula for precision.

$$Precision = \frac{TP}{TP + FP}$$

  where TP is true positive, TN is true negative and FP is false positive.

[5] `https://freekidsbooks.org/when-siggy-met-phyllis/`

- Recall: Recall is a separate measurement that provides an answer to the issue of how many real Positives are accurately categorized. When we aim to catch as many positives as possible, recall is a good option of assessment statistic.

$$Recall = \frac{TP}{TP + FN}$$

where FN is the false negative.

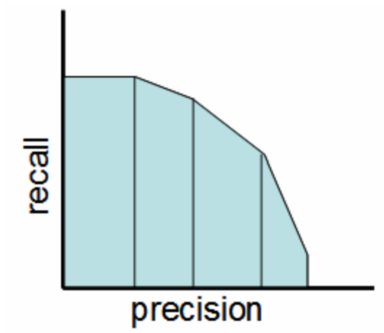

Figure 7: Recall and precision trade-off
6

- F1-score: The Figure 7 represents the recall and precision trade-off. The harmonic mean of accuracy and recall is represented by the F1 score, which ranges from 0 to 1. A model with strong recall and precision would be ideal. Simply put, the F1 score sort of keeps the precision and recall of your classifier in balance. If your memory is poor, your F1 score will also be poor if your accuracy is poor.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN}$$

where FP is false positive.

- Accuracy: The fundamental classification metric is accuracy. It is rather simple to comprehend. Additionally, it is ideally suited to binary and multi-class classification problems. The percentage of real outcomes in relation to the total cases investigated is the accuracy.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- Confusion Matrix: One of the most used measures for assessing the effectiveness of multi-class classifiers is "confusion matrices". Each row in the confusion matrix, which is occasionally shown as a heat map, represents an actual or goal class. The model's forecast for each class is displayed in each column. As a result, each cell's value is a composite of its predicted and actual value. The matrix's diagonal values display the proportion of accurate model predictions for each class.Belyadi and Haghighat (2021)

6.3.1 Multinomial Naive Bayes

For the evaluation of the Naive Bayes model we will refer the confusion matrix , precision , recall and f1-score.As seen in the Figure 8 below we can see the recall, precision and f1-score of the different 8 emotions of anger, disgust, fear, joy, neutral, sadness, shame and surprise. We can also see that the accuracy of the model is only 63 percent.

	precision	recall	f1-score	support
anger	0.67	0.35	0.46	2304
disgust	0.76	0.10	0.18	708
fear	0.77	0.49	0.60	2003
joy	0.58	0.62	0.60	5721
neutral	0.62	0.94	0.75	9819
sadness	0.63	0.39	0.48	2741
shame	0.00	0.00	0.00	43
surprise	0.71	0.33	0.45	3413
accuracy			0.63	26752
macro avg	0.59	0.40	0.44	26752
weighted avg	0.64	0.63	0.60	26752

Figure 8: Naive model evaluations

As seen in the figure the identification of the emotion class is done with the highest accurate emotion identified amongst all the emotions present in the text of a particular page in a book.

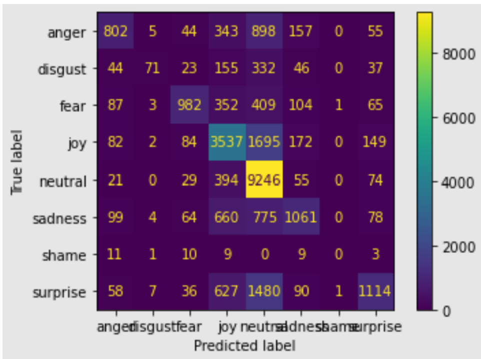

Figure 9: Confusion Matrix for Naive Bayes model

6.3.2 Recurrent Neural Network

For the RNN model the last layer consists of a Softmax layer.This Softmax layer outputs a tensor of dimension 8 , since we have 8 emotion classes, using a softmax activation function as softmax is an extension of sigmoid[7] which handles multiple output classes. Softmax assigns decimal probabilities to each class in a multi-class problem. Those decimal probabilities must add up to 1.0. This additional constraint helps training converge more quickly than it otherwise would. To avoid overfitting Dropout layers have been added with 0.2 values. For the RNN model the below Figure 10 depicts the precision , recall and accuracy of the model.

[7]Softmax:`https://developers.google.com/machine-learning/crash-course/multi-class-neural-networks/softmax`

```
Epoch 1/10
836/836 [==============================] - 790s 944ms/step - loss: 1.4976 - accuracy: 0.4655 - precision_1: 0.6525 - recall_1: 0.2409
Epoch 2/10
836/836 [==============================] - 796s 952ms/step - loss: 1.1890 - accuracy: 0.5906 - precision_1: 0.7264 - recall_1: 0.4355
Epoch 3/10
836/836 [==============================] - 788s 943ms/step - loss: 0.9973 - accuracy: 0.6670 - precision_1: 0.7657 - recall_1: 0.5547
Epoch 4/10
836/836 [==============================] - 793s 949ms/step - loss: 0.8683 - accuracy: 0.7125 - precision_1: 0.7887 - recall_1: 0.6290
Epoch 5/10
836/836 [==============================] - 792s 948ms/step - loss: 0.7701 - accuracy: 0.7464 - precision_1: 0.8122 - recall_1: 0.6752
Epoch 6/10
836/836 [==============================] - 798s 954ms/step - loss: 0.6927 - accuracy: 0.7758 - precision_1: 0.8306 - recall_1: 0.7140
Epoch 7/10
836/836 [==============================] - 797s 953ms/step - loss: 0.6313 - accuracy: 0.7942 - precision_1: 0.8448 - recall_1: 0.7412
Epoch 8/10
836/836 [==============================] - 796s 952ms/step - loss: 0.6025 - accuracy: 0.8020 - precision_1: 0.8466 - recall_1: 0.7562
Epoch 9/10
836/836 [==============================] - 795s 951ms/step - loss: 0.5646 - accuracy: 0.8158 - precision_1: 0.8564 - recall_1: 0.7707
Epoch 10/10
836/836 [==============================] - 790s 945ms/step - loss: 0.5201 - accuracy: 0.8294 - precision_1: 0.8661 - recall_1: 0.7919
```

Figure 10: Evaluation of RNN model

6.4 Discussion

The key findings are critically reviewed in this section. By raising the epoch values, we may improve the training and accuracy of the RNN model. However, when the number of epochs rises, the weight in the neural network is altered more often, and the curve shifts from being under-fitting to being ideal to being over-fitting. Because of this, we selected an epoch value of 10 for our investigation, which showed an accuracy of 82 percent.One of the issues observed in the research is that most of the pages in the book depict a neutral emotion. The logic behind the models is that out of all the emotions observed in the text of a page the most relevant or the emotion with more weight is selected. Hence in most scenarios, the identified emotion points towards a neutral emotion. By using the model on a children's book, we can observe more variations in identified emotions.

7 Conclusion and Future Work

In this study, we used a social media data collection and two separate algorithms to determine the emotions present in the text of a book page. The algorithms were able to correctly categorize the emotions despite the fact that the data sets were very different from the slang and expressions used in a book.Here, the textual input is pre-processed using Nltk, and the words are transformed into a format that computers can comprehend using a counter vectorizer for Naive Bayes and one-hot encoding for RNN model. As observed by the evaluation results the RNN model has proven to be a better model for the emotion identification of the pages in the sample book with an accuracy of 82 percent as opposed to the accuracy of 63 percent of the Naïve Bayes model.

For this research, a simple RNN model was used for the analysis using the social media data-sets, for future work other models such as CNN, BERT and LSTM models can be used for emotion identification. The data-sets used for this research consisted of daily dialogues and tweets, hence more variations in the data-sets can be made like using the data-sets of emotion labelled books. All the music used in this research is around 15 seconds duration, in further studies the time taken for a person can be identified and the duration can be set accordingly so that the pages and the music played in the background will be in the same flow. This idea can be made into a web application and combined as feature in different e-books or as an add-on to PDF files.

References

Adikari, A., Gamage, G., De Silva, D., Mills, N., Wong, S.-M. J. and Alahakoon, D. (2021). A self structuring artificial intelligence framework for deep emotions modeling and analysis on the social web, *Future Generation Computer Systems* **116**: 302–315.

Almeida, A. M., Cerri, R., Paraiso, E. C., Mantovani, R. G. and Junior, S. B. (2018). Applying multi-label techniques in emotion identification of short texts, *Neurocomputing* **320**: 35–46.

Ameer, I., Sidorov, G., Gomez-Adorno, H. and Nawab, R. M. A. (2022). Multi-label emotion classification on code-mixed text: Data and methods, *IEEE Access* **10**: 8779–8789.

Asif, M., Ishtiaq, A., Ahmad, H., Aljuaid, H. and Shah, J. (2020). Sentiment analysis of extremism in social media from textual information, *Telematics and Informatics* **48**: 101345.

Belyadi, H. and Haghighat, A. (2021). *Machine learning guide for oil and gas using Python*, Vol. 10, Elsevier.

Ekman, P. (1999). Basic emotions, *Handbook of cognition and emotion* **98**(45-60): 16.

Gaind, B., Syal, V. and Padgalwar, S. (2019). Emotion detection and analysis on social media, *arXiv preprint arXiv:1901.08458* .

Gosai, D. D., Gohil, H. J. and Jayswal, H. S. (2018). A review on a emotion detection and recognization from text using natural language processing, *International Journal of Applied Engineering Research* **13**(9): 6745–6750.

Halim, Z., Waqar, M. and Tahir, M. (2020). A machine learning-based investigation utilizing the in-text features for the identification of dominant emotion in an email, *Knowledge-based systems* **208**: 106443.

Mohanty, M. D. and Mohanty, M. N. (2022). Verbal sentiment analysis and detection using recurrent neural network, *Advanced Data Mining Tools and Methods for Social Computing*, Elsevier, pp. 85–106.

Raza, M. R., Hussain, W. and Varol, A. (2022). Performance analysis of deep approaches on airbnb sentiment reviews, *2022 10th International Symposium on Digital Forensics and Security (ISDFS)*, IEEE, pp. 1–5.

Sailunaz, K. and Alhajj, R. (2019). Emotion and sentiment analysis from twitter text, *Journal of Computational Science* **36**: 101003.

Singh, C., Imam, T., Wibowo, S. and Grandhi, S. (2022). A deep learning approach for sentiment analysis of covid-19 reviews, *Applied Sciences* **12**(8): 3709.

Su, M.-H., Wu, C.-H., Huang, K.-Y. and Hong, Q.-B. (2018). Lstm-based text emotion recognition using semantic and emotional word vectors, *2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia)*, IEEE, pp. 1–6.

Xu, D., Tian, Z., Lai, R., Kong, X., Tan, Z. and Shi, W. (2020). Deep learning based emotion analysis of microblog texts, *Information Fusion* **64**: 1–11.