UNIVERSITÀ
DEGLI STUDI
DI MILANO

Università degli Studi di Milano

Text Mining and Sentiment Analysis course

Emotion detection in song lyrics Application

Author: Andrea Ierardi
Student ID: 960188

May, 2021

Table of Contents

List of Figures

1 Introduction

The aim of the project is the the developing of different models for predicting the emotions in text lyrics. Moreover, the model have been used to exploit a correlation between different genres and emotions. In particular, a random algorithm playlist suggestion based on genre and emotion has been created. Moreover, a simple web application has been developed for the deploying of the work. The considered data is the WASSA-2017 Shared Task on Emotion Intensity (EmoInt) dataset. It comprehends a series of Tweet from Twitter social network where are provided for four emotions: joy, sadness, fear, and anger. A comparison of the predictions of different models have been made: Logistic Regression, Random Forest, Multi-layer Perceptron classifier and different Neural Networks architectures. An inspection of the different predictions from the models have been done and different results have been obtained.
The web application backend has been developed using Django Python library and the fronted using HTML, Javascript and AJAX languages.

2 Research question and methodology

The aim is the developing of a web application that is able to predict emotion from lyrics song. Different models have been trained for the classification task.

2.1 Data

2.1.1 Datasets

The data considered is the *WASSA-2017 SharedTask on Emotion Intensity (EmoInt)* dataset. It is composed by a series of Tweets and each of them is labeled with one of the following four emotions: joy, sadness, fear, and anger. Another dataset has been considered for the songs lyrics and it is called *"Music Dataset: Lyrics and Metadata from 1950 to 2019"*. The dataset provides a list of lyrics from 1950 to 2019 describing music metadata as sadness, danceability, loudness, acousticness, etc. The audio data was scraped using Echo Nest API integrated engine with spotipy Python's package. The spotipy API permits the user to search for specific genres, artists,songs, release date, etc. To obtain the lyrics they used the Lyrics Genius API as baseURL for requesting data based on the song title and artist name. For the purpose of the project only title, lyrics text and author of the songs have been considered.

2.1.2 Preprocessing

The EmoInt raw dataset was already divided in train, validation and test sets for each emotion. The first approach was to merge all the different emotion train, validation and tests to then shuffled to avoid having emotions in consecutive order.
For the pre-processing different approaches have been used:

- **Pipeline**: stemming, stop words removal and tokenization. Then vectorization (unigrams) of the document to apply features selection and standard scaler.
- **Pipeline with 2-grams**: same procedure of Pipeline but also including bigrams.
- **Sequences**: Tokenization and sequentialization.

2.2 Models

2.2.1 Logistic Regression Classifier

Logistic Regression is a Statistical Learning technique categorized in Supervised Machine Learning (ML) methods dedicated to Classification tasks.

2.2.2 Random Forest Classifier

Random forests is a supervised learning algorithm that can be used both for classification and regression. Random forests creates decision trees on randomly selected data samples, gets prediction from each tree and selects the best solution by means of voting.

2.2.3 Multi-layer Perceptron Classifier

A multilayer perceptron (MLP) is a class of feedforward artificial neural network. It consists of three or more layers (an input and an output layer with one or more hidden layers) of nonlinearly-activating nodes.

2.2.4 Neural Networks

A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature.
In particular different architecture types have been explored:

- Simple FeedForward Neural Network,
- Deep Neural Network,
- Neural Networks with Embeddings,
- Neural Networks trained with GridSearch CV hyperparameters tuning.

2.3 Application

2.3.1 FrontEnd

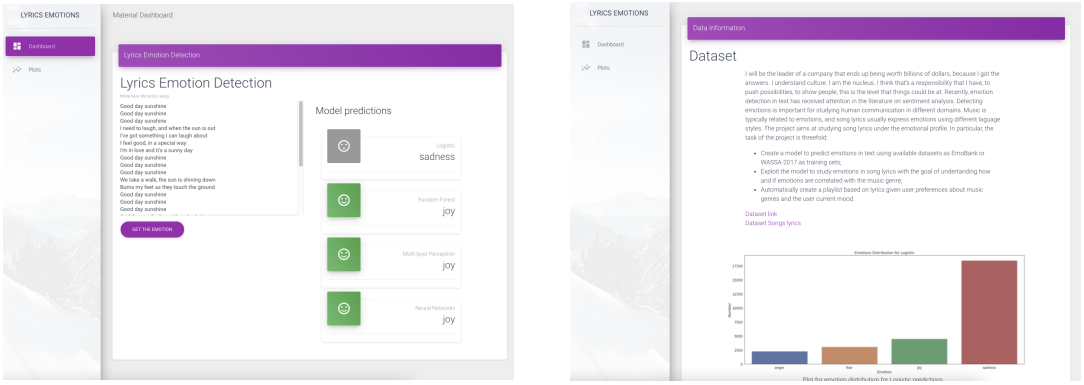
The application has a user interface developed in HTML, CSS, Javascript and AJAX languages. The interface is based on a template in AppSeed.us called Django Dashboard Material.

2.3.2 BackEnd

The application has a web server developed using Django Python library. The principal scope of the server is to read the model, the data and serving prediction request from the user interface.

2.3.3 Main section: dashboard

The main section (Figure 1a) of the web app include the dashboard page in which there is a text are in which the user can write or paste a lyrics text and obtain the prediction.
The predictions are listed in different cards for each model: Logistic, Random Forest, MLP e Neural Networks.
The plot section (Figure 21) dispays the dataset source and the different plots representing the distribution of genre and emotions (Described later in the paper).



(a) Main section **(b)** Plot section

Figure 1: Page structure of the App

3 Experimental results

For the simplicity of the experiment to measure the accuracy prediction, the *f1 score* has been used. In particular the weighted average criteria has been used to obtain the f1 score since it is a multi-class classification problem.

3.1 Data

Each data pre-processing considered output different results. For the pre-processing using basic pipeline performed well in general but the pipeline using 2-grams was the best performing in terms of accuracy. The worst performing was the pre-processing using sequences but it was needed for training Neural Networks with Embeddings. Labels has been pre processed using Label Encoder and also for Neural Networks the vectors were converted to binary class matrix.

3.2 Model results

3.2.1 Logistic Regression results

The logistic regression models gave a acceptable results in terms of f1 score. In particular, the one tuned with Grid Cross Validation output the best results for this class of predictor. It had an accuracy f1 score of 0.826. Sadness is the more predicted emotion, moreover it is predicted more than 60% of the time. From Figure 2 is possible to see the emotion distribution of the prediction outputs.

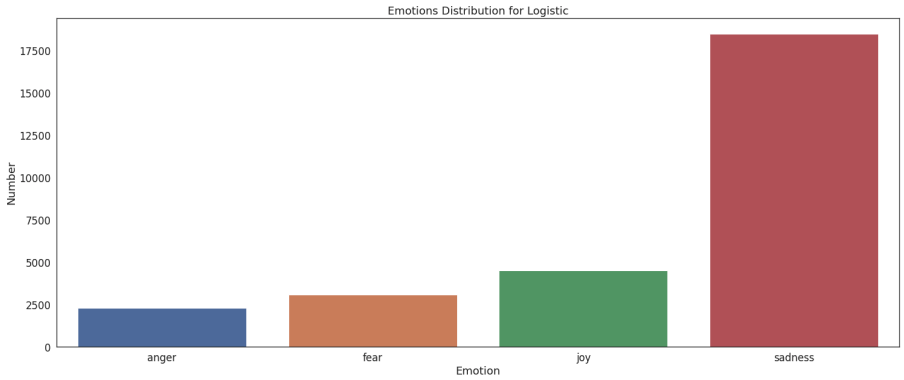


Figure 2: Emotion distribution predicted by Logistic

From Figure 3 is possible to see the emotion distribution of the top-ten different genre and different predicted emotions. Pop sadness songs are the most common, based on the prediction o f the Logistic model.

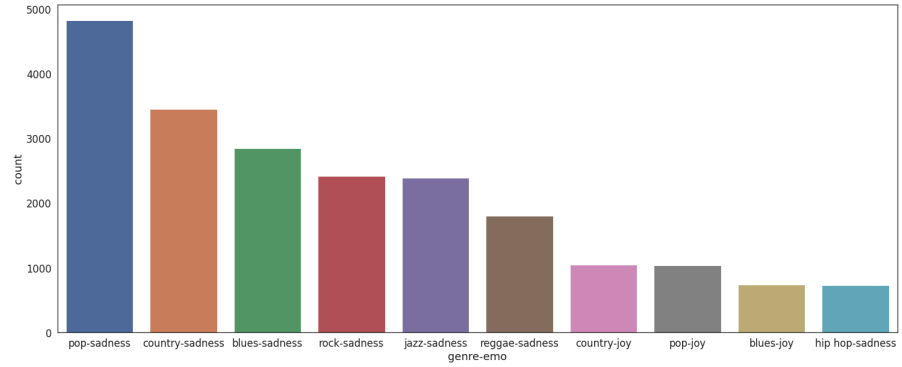


Figure 3: Top-ten genre-emotion distribution Logistic

3.2.2 Random Forest results

The Random Forest f1 score results are acceptable. In particular, the one tuned with Grid Cross Validation outputs the best results with an accuracy f1 score of 0.802. From Figure 4 is possible to see the emotion distribution of the prediction outputs. Sadness is predicted 60% of the time.

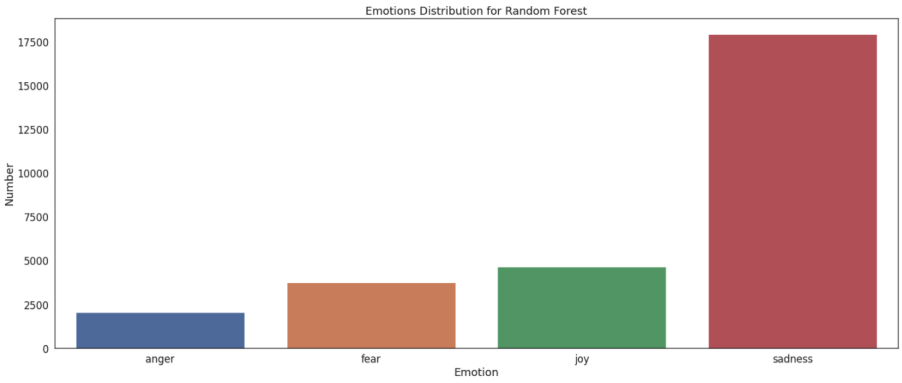


Figure 4: Emotion distribution predicted by Random Forest

From Figure 5 is possible to see the emotion distribution of the top-ten different genre and different predicted emotions.

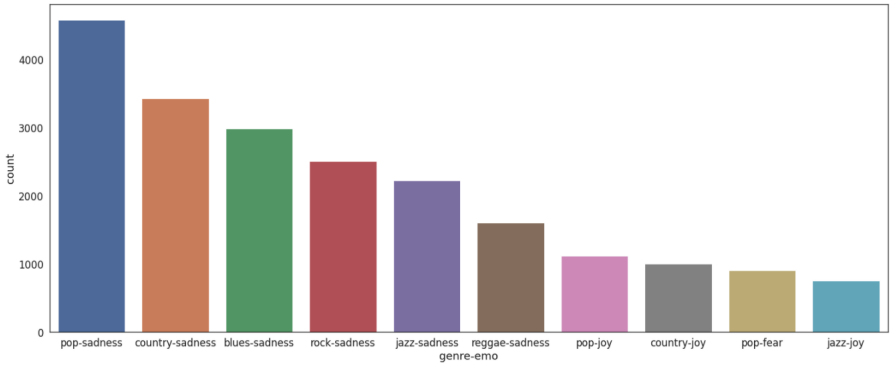


Figure 5: Top-ten genre-emotion distribution Random Forest

3.2.3 Multi-layer Perceptron results

The Multi-layer Perceptron model tuned with Grid Cross Validation output an accuracy f1 score of 0.803. From Figure 6 is possible to see the emotion distribution of the prediction outputs. Fear is the more predicted emotion by MLP.

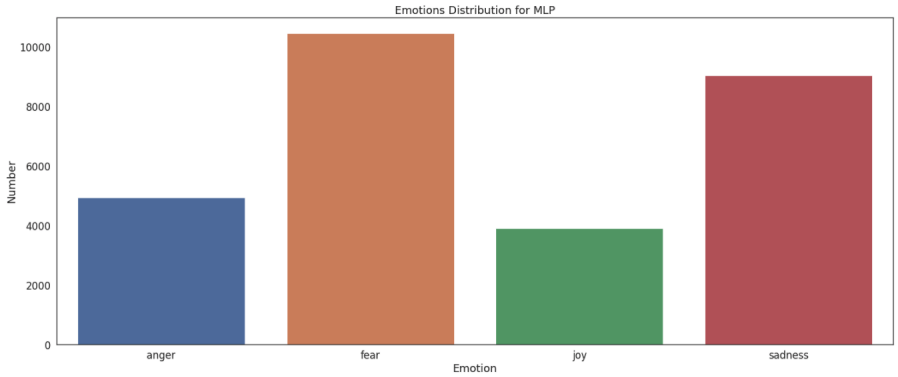


Figure 6: Emotion distribution predicted by MLP

From Figure 7 is possible to see the emotion distribution of the top-ten different genre and different predicted emotions. Pop and fear are the most common songs based on the prediction of MLP.

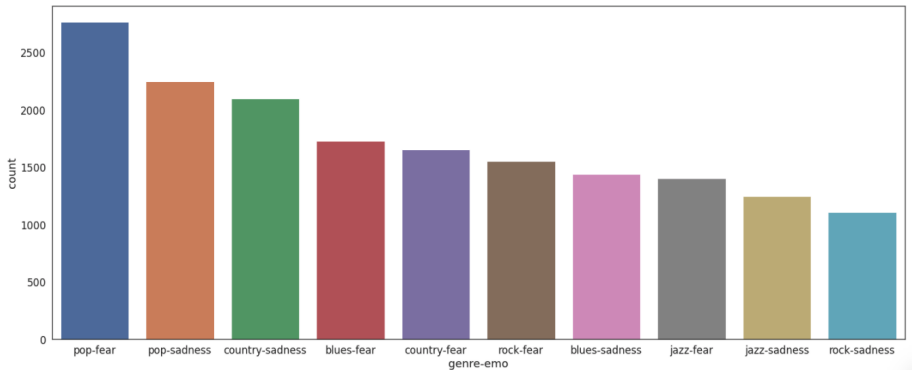


Figure 7: Top-ten genre-emotion distribution MLP

3.2.4 Neural Networks results

Different Neural Networks Architectures have been test:

- Dense Neural Networks: 3 dense layers with different number of nodes.
- Neural Networks with Embeddings: Embedding layer with two Dense layers
- Simple Feed Forward Neural Networks
- Tuned Dense Neural Networks with Grid Cross Validation

The Neural Networks models gave similar results in terms of f1 score.

Dense Neural Networks 3 node configurations have been tested: two dense layers with 12-8 nodes, 16-8, 32-16. The one with the major accuracy score is the first one with 0.803. (Figure 8)

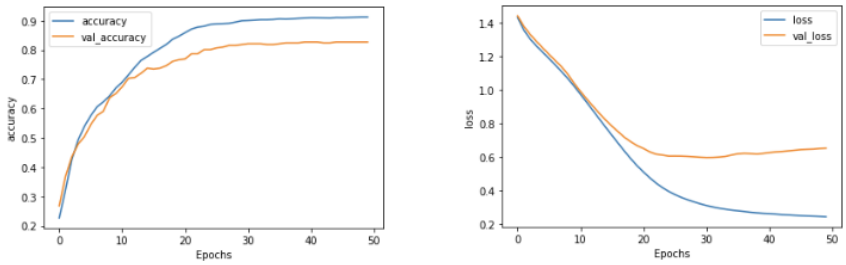


Figure 8: Best Dense Neural Network accuracy and loss

Neural Networks with embedding output a score of 0.724. From Figure 9 is possible to see how validation loss is reducing but still remain high with respect to the other Network Architectures.

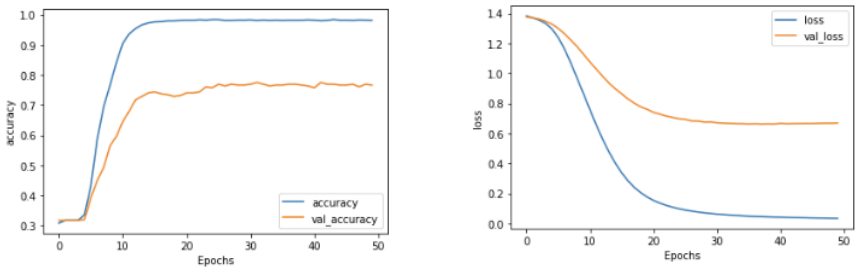


Figure 9: Best Neural Network with Embeddings accuracy and loss

Simple Feed Forward Neural Networks obtain a f1 score of 0.816. (Figure 10)

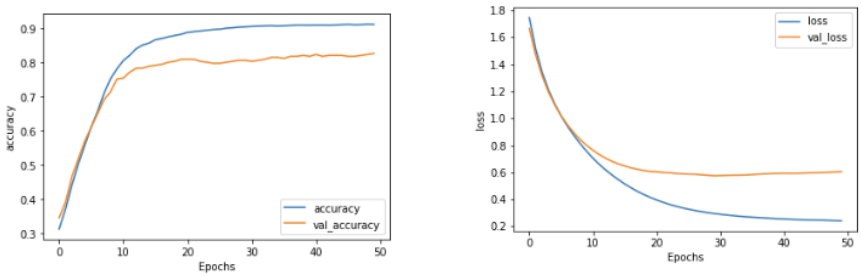


Figure 10: Best Feed Forward Neural Network accuracy and loss

Tuned Neural Networks with Grid Cross Validation output the best results. It gave a accuracy f1 score of 0.826. From Figure 11 is possible to see that after the 5^{th} epoc the validation loss is increasing.

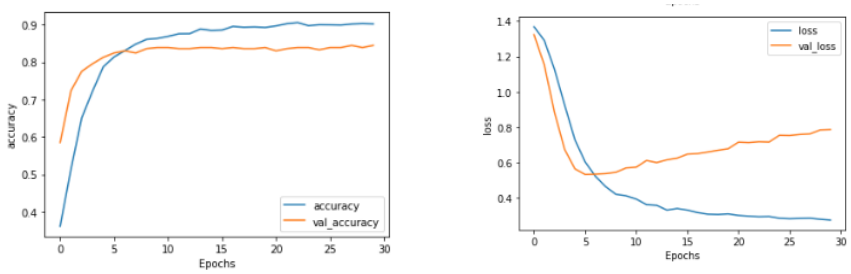


Figure 11: Best CV Neural Network accuracy and loss

From Figure 12 shows the emotion distribution predicted by the tuned Neural Network model. Sadness is the more predicted emotion by the model.

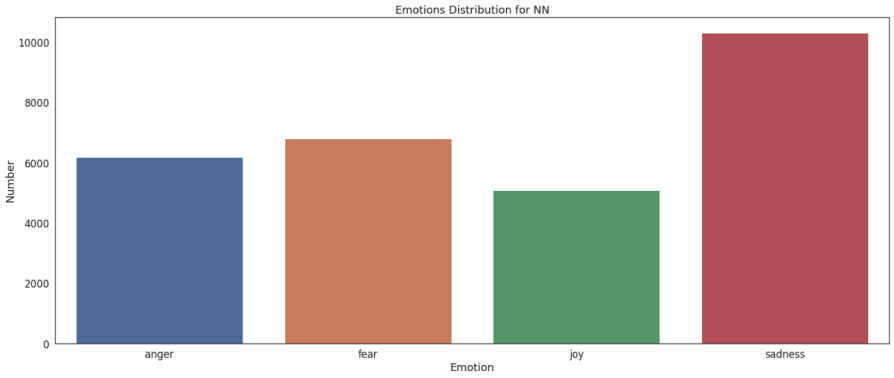


Figure 12: Emotion distribution predicted by Neural Networks

From Figure 22 is possible to see the emotion distribution of the top-ten different genre and different predicted emotions. Pop and sadness songs are the most common based on the prediction of the tuned Neural Networks.

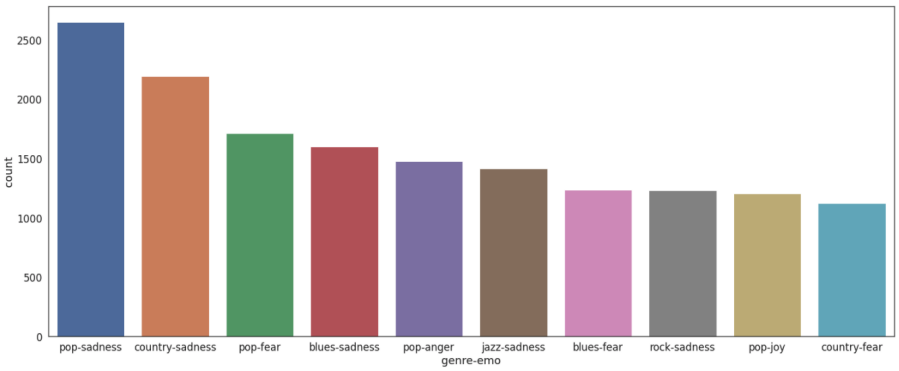


Figure 13: Top-ten genre-emotion distribution Neural Networks

3.2.5 Practical Results

Other lyrics example have been used to test and verify accuracy of the predictors in simple case and in real world scenarios.
Three simple sentence has been considered: *"Good day"*, *"Bad day"*, *"Evil"*.

They have been passed in input to the predictors to explore the possible prediction of the different models. Figure 14 shows that the results for the models are similar, but Random Forest is the only one who output a wrong result.
In Figure 15 the models output different results. Logistic and Neural Networks output Joy. This may be due to the fact the the word *day* is actually the one with more impact in the prediction. Random Forest again predicts Fear, while the one more closer is MLP with a sadness prediction.
In Figure 16 every models gave wrong results predicting joy.Random Forest, instead, predict Fear emotion. This could be due to the fact that the Random Forest predictor uses fear label to represent unknown test.

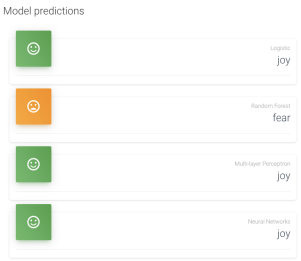


Figure 14: Predicted emotion for lyrics "Good day"

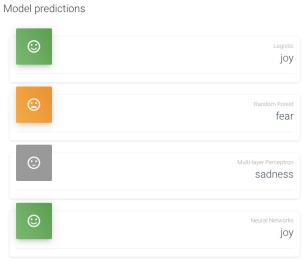


Figure 15: Predicted emotion for lyrics "Bad day"

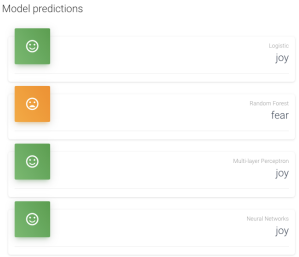


Figure 16: Predicted emotion for lyrics "Evil"

Figure 17: Simple practical examples using the Application

The second test focused in analysing the results for already existing songs: *"Blackbirds"*, *"Good Day Sunshine"* from The Beatles and *"Till I collapse"* by Eminem. The first song is an example of Sad song, the second is an example of Joyful song, the third one could be both fear or anger feeling. In Figure 18 the prediction for all the models is Sadness. This is a case where all models predict the right emotion. In Figure 19 the prediction is Joy for every models apart from the Logistic one that predicts Sadness. In Figure 20 the prediction is right for Neural Networks and MLP. Logistic is wrong but at least predicts a negative feeling, while Random Forest predicts Joy.

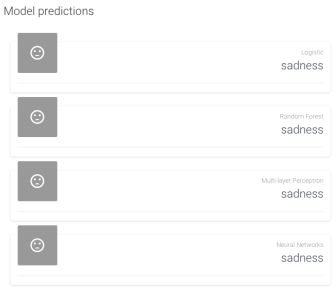


Figure 18: Predicted emotion for "Blackbird" by The Beatles

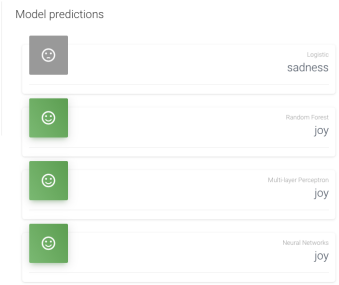


Figure 19: Predicted emotion for "Good Day Sunshine" by The Beatles

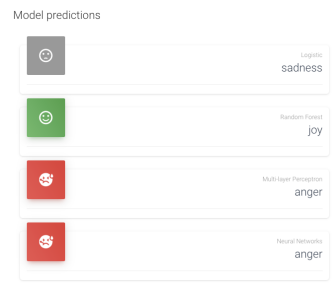


Figure 20: Predicted emotion for "Till I Collapse" by Eminem

Figure 21: Real song practical examples using the Application

3.2.6 Song Suggestions

For each model have been implemented a simple song suggestion algorithm based on the lyrics song dataset. The algorithm take as input the genre and a emotion of a song and outputs the information of 10 songs with the same predicted emotion and genre. The criteria of choices for the 10 suggested songs is random.

track_name	artist_name	genre	pred
anna lee, the healer	the beach boys	pop	joy
loving is easy	rex orange county	pop	joy
smoke gets in your eyes	the platters	pop	joy
jesus loves me	whitney houston	pop	joy
it's my life	no doubt	pop	joy
4th dimension	kids see ghosts	pop	joy
true believers	the bouncing souls	pop	joy
pay it back	elvis costello	pop	joy
do my thang	miley cyrus	pop	joy
all in the groove	blues traveler	pop	joy

Figure 22: Song suggestion example

4 Concluding remarks and future work

4.1 Conclusion

In conclusion, the two best models according to the f1 score are the tuned Neural Networks and tuned Logistic (Figure 23). Moreover all the model outputs a similar f1 score greater than 0.8 and differs only for 2 percentage points.
Considering practical examples done in the web application, the Neural Network models, together with the MLP are better in predicting simple text but also in predicting real world use cases. Neural Networks and MLP predicted the emotions balanced, while Random Forest and Logistic models predicted 60% of the time sadness .

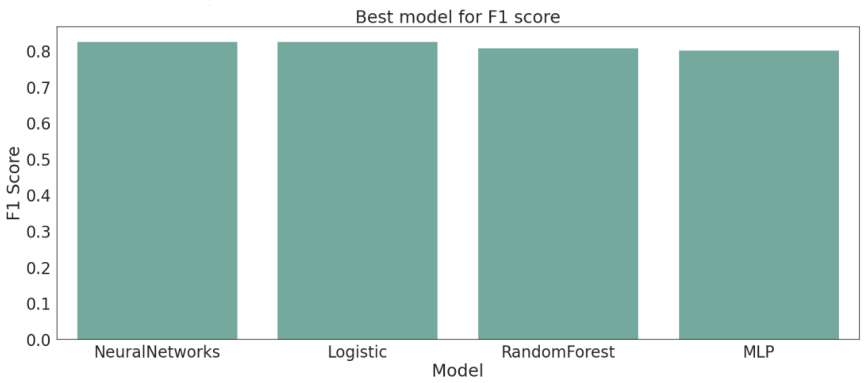


Figure 23: Final results

4.2 Future work

For future works a possible change in the prediction. Instead of considering all the text lyrics, it may be convenient to split the corpus in different sentence and predict the emotion for each one and then return the final results using a majority vote criteria. Another possible progression could be the integration of sound information. The inclusion of melodies and other information could lead to a better prediction.

5 Appendix

- Notebook code
- Repository link