uOttawa

Report

Emotion Detection Chatbot

Data Science Applications

**Group Number: G4**

Team Members:

Abdallah Ragab, Hossam Mahmoud

Ahmed Ali Ziada, Shaimaa Mamdouh Ahmed

1. Overview

Human can express their emotions by several ways one of them is texting but some people become afraid from express his emotions to people so our main objective in this project is to build a chatbot using google Dialog flow, that can understand the user emotions and recommend songs to the user based on the user mode.

2. System Architecture
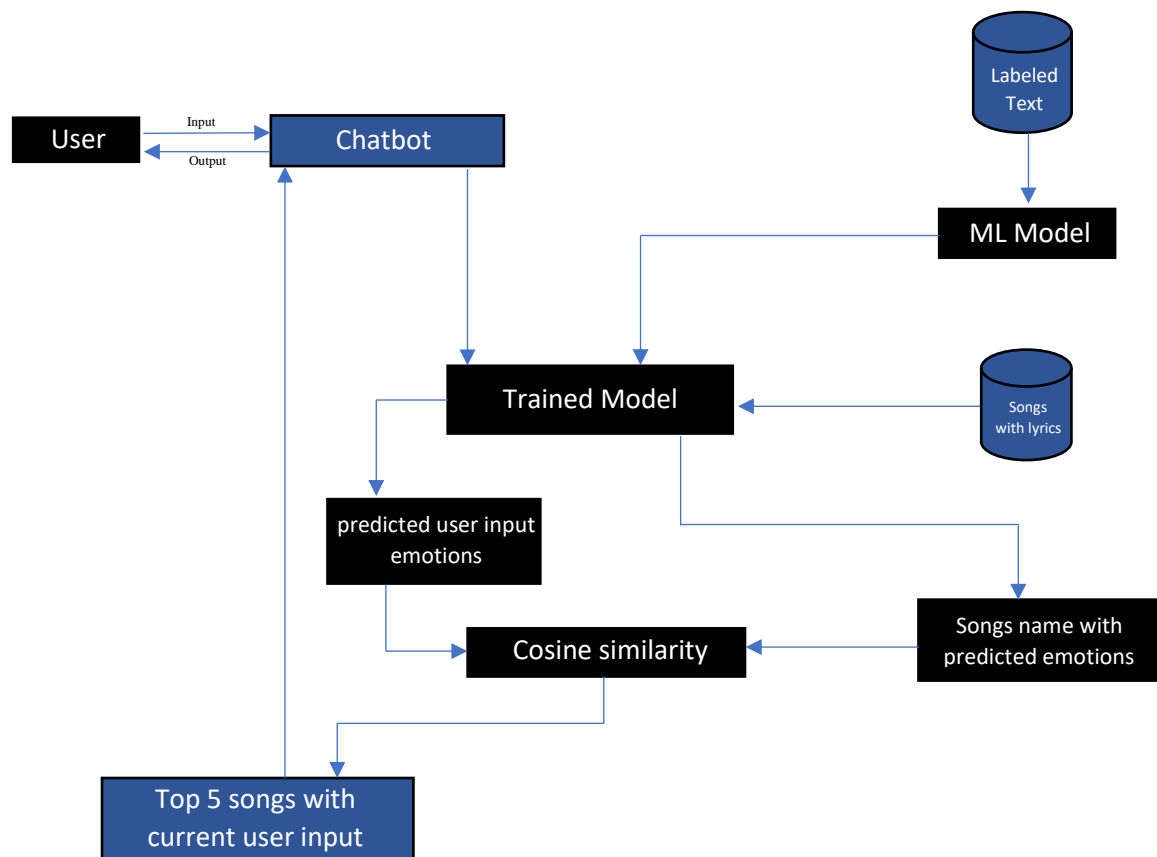
A machine learning model has been trained on labeled dataset and this model generate trained model, which we use it to predict the emotions of user's text input and song's lyrics.

calculate the cosine similarity between the user input emotion vector and songs emotion vectors and choose the most 5 similar songs.

The output will be text response from chatbot, and suggested songs.

3. Methodology

3.1 Firstly, chatbot

3.1.1 Chatbot creation

A chatbot that detects emotions and based on these emotions, the chatbot generate customize response to enhance mental health for users and recommend a song for them.
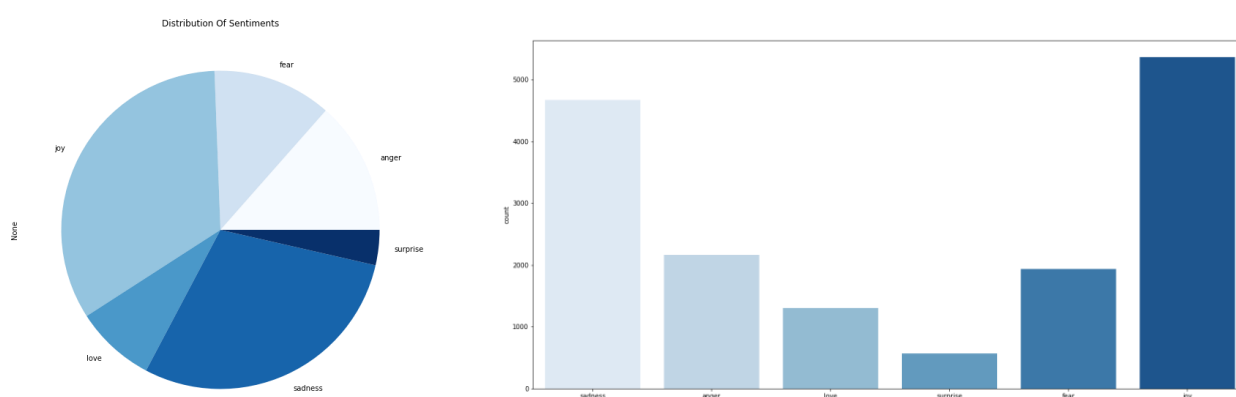
3.1.2 Chatbot tool

We used dialogflow from google to build the chatbot, using webhock to connect chatbot with our models.

3.2 Sentiment analysis

3.2.1 Dataset

We used dataset from Kaggle, and this data having sentences and their labels, which represent joy, love, surprise, sadness, fear, or angry.


3.2.2 Classification Models

We used several models to classify input sentence to their emotions class.

3.2.2.1 Support Vector Machine

SVM is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well its best suited for classification. The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points.
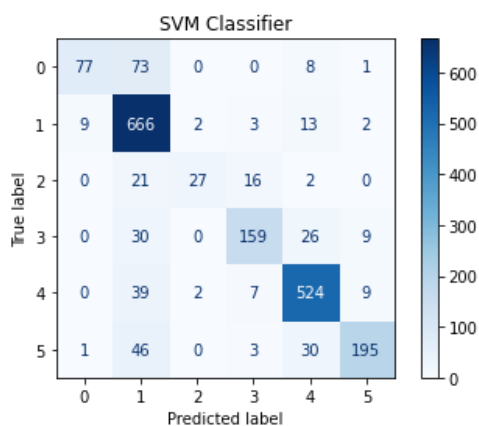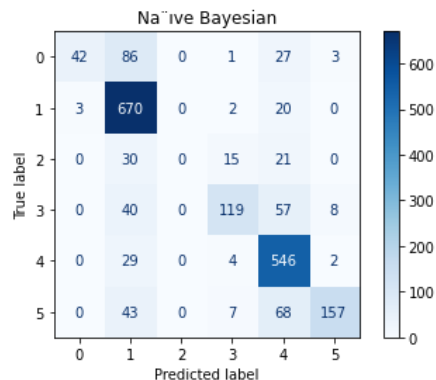

	precision	recall	f1-score	support
0	0.89	0.48	0.63	159
1	0.76	0.96	0.85	695
2	0.87	0.41	0.56	66
3	0.85	0.71	0.77	224
4	0.87	0.90	0.89	581
5	0.90	0.71	0.79	275
Accuracy	**82 %**			

3.2.2.2 Naive Bayes Classifier

Naïve Bayes (NB) Classifier: is Bayesian graphical model that has nodes corresponding to each of the columns or features. It is called naive because, it ignores prior distribution of parameters and assume independence of all features and all rows. Ignoring prior has both an advantage and disadvantage.
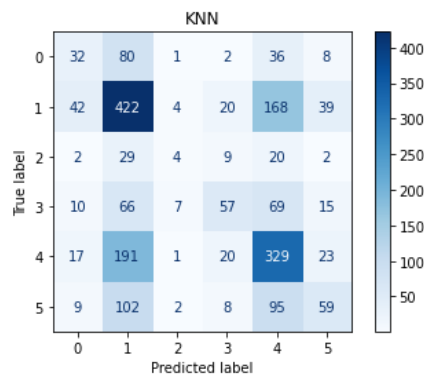

	precision	recall	f1-score	support
0	0.93	0.26	0.41	159
1	0.75	0.96	0.84	695
2	0.00	0.00	0.00	66
3	0.80	0.53	0.64	224
4	0.74	0.94	0.83	581
5	0.92	0.57	0.71	275
Accuracy	**77 %**			

3.2.2.3 K-Nearest Neighbors

KNeighbors Classifier: a user-defined function which accepts an array of distances, and returns an array of the same shape containing the weights, We used n_neighbors equal 5.
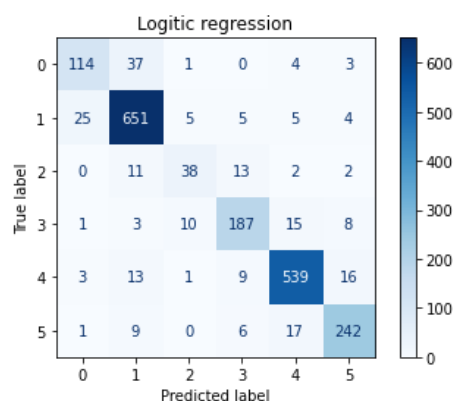

	precision	recall	f1-score	support
0	0.29	0.20	0.24	159
1	0.47	0.61	0.53	695
2	0.21	0.06	0.09	66
3	0.49	0.25	0.34	224
4	0.46	0.57	0.51	581
5	0.40	0.21	0.28	275
Accuracy	**45 %**			

3.2.2.4 Logistic Regression

Logistic Regression: is a classification algorithm rather than regression algorithm. Based on a given set of independent variables, it is used to estimate discrete value (0 or 1, yes/no, true/false).
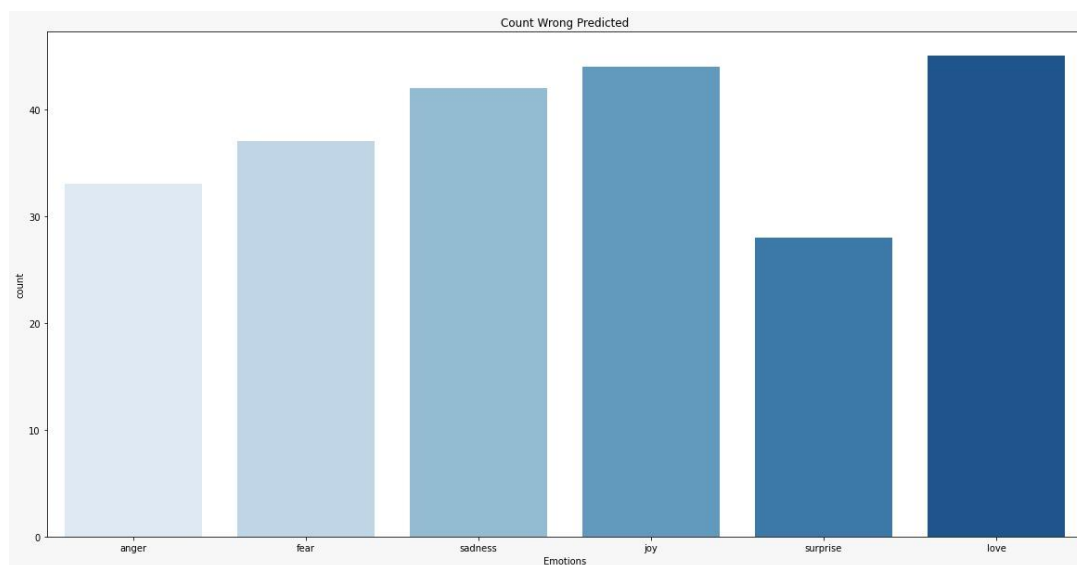

	precision	recall	f1-score	support
0	0.79	0.72	0.75	159
1	0.90	0.94	0.92	695
2	0.69	0.58	0.63	66
3	0.85	0.83	0.84	224
4	0.93	0.93	0.93	581
5	0.88	0.88	0.88	275
Accuracy	**89 %**			

Based on the accuracy we choose Logistic Regression Classifier to be our **champion model**

3.2.2.5 Error Analysis

Figure shown count of wrong predicted in every features.

3.2.3 Clustering Models

We used Kmean Classifier with same data but without labels.

3.2.3.1 Kmean

K-means clustering: is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. model evaluated with kappa score
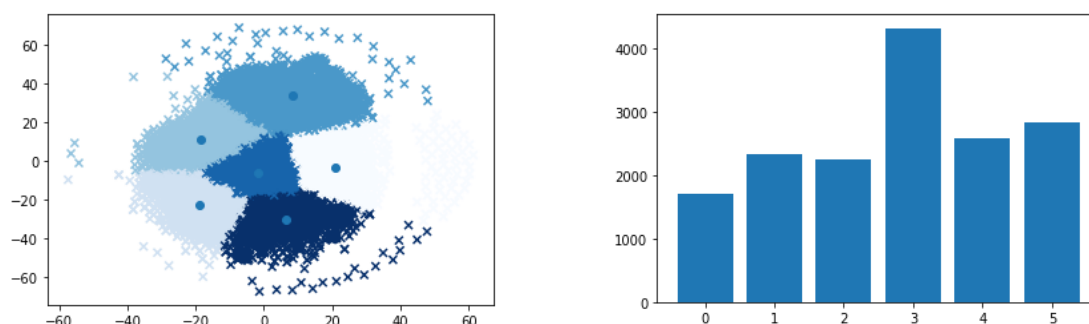
3.2.4 Recommendation system for music based on emotions

3.2.4.1 Emotion Detection on the songs

We download the songs dataset from Kaggle, which contains the song name and song lyrics. We applied our trained model from the previous step to get the vector of the probabilities of each emotion

Emotion Vector (love , joy , surprise , fear , sadness , anger)
E.g Ring song (0.035831 , 0.421128 , 0.015184 , 0.122859 , 0.281957 , 0.123042)

and we calculate the same vector for each song then combine the results with the song name, and save it as a .csv file to use it again to recommend the user with the closest song based on detected emotion from his input.

3.2.4.2 Detect the emotion from user input

we take the user input from the chatbot and predict emotion class and calculate the emotion vector

e.g. user input: "I am sad"
Emotion Vector (love   ,   joy   , surprise ,   fear   , sadness , anger)
Emotion Vector (0.01687 , 0.0992   , 0.011423 ,   0.08392 , 0.67880 ,   0.1097)
Predicted class = **sadness**

3.2.4.3 Choose the closest song to the user based on the emotion vector

Unlike making recommendations based on classified emotion e.g. (sadness,) only. we consider the rest of the emotions by calculating the emotion vector as shown previous section then calculate the cosine similarity between the user input emotion vector and songs emotion vectors and choose the most 5 similar songs.

3.2.5   Test cases from chatbot
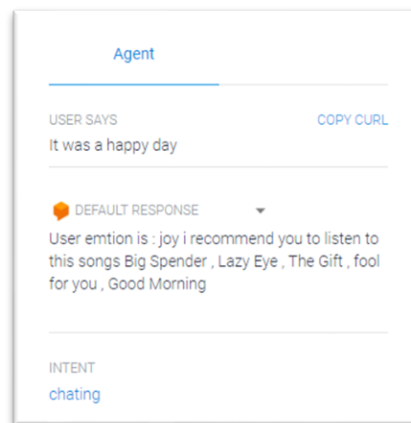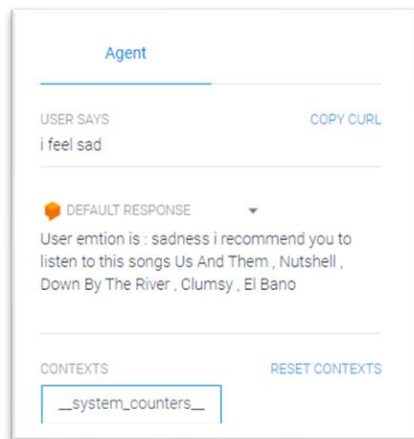
3.2.6   Error analysis for chatbot

Chatbot have an error in detecting mixed feelings
but it will recommend songs nearest to the mixed feelings.
As this because the recommendation is based on cosine similarity
Between user emotion vector and song emotion vector.
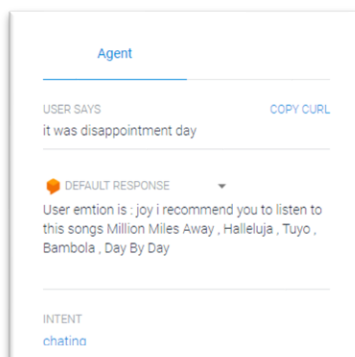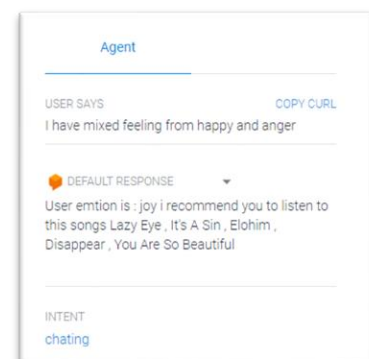
Chatbot have an error with some of expressions.

4. Conclusion

In the project, Emotion Detection Chatbot established showing satisfy results using classification model to detect the emotions from user's input and suggest a song based on detected user's emotions. The emotions classification model recorded accuracy of 89 % in testing phase. This chatbot can be used in various sectors specially in mental health sector.

5. References

- Dataset labeled with emotions
  https://www.kaggle.com/code/nataliakhol/emotion-detection-from-text-bilstm/data?select=val.txt

- Song's lyrics dataset
  https://www.kaggle.com/datasets/terminate9298/songs-lyrics?select=lyrics.csv

- https://iopscience-iop-org.proxy.bib.uottawa.ca/article/10.1088/1742-6596/1362/1/012039/pdf

- https://ieeexplore-ieee-org.proxy.bib.uottawa.ca/stamp/stamp.jsp?tp=&arnumber=9580069&tag=1