PROGRESS REPORT

Mentor: Nilanjana Dutta Roy

As mentioned and decided earlier the entire team has been divided into three parts:

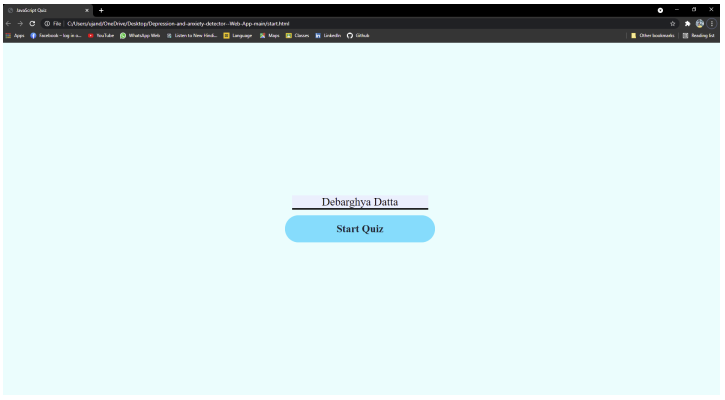
1. Data Processing and Handling Team (ML and Data Handling)
2. FrontEnd, User Interaction & User Experience
3. Cloud Execution and Cloud Computing

The teams were as follows:

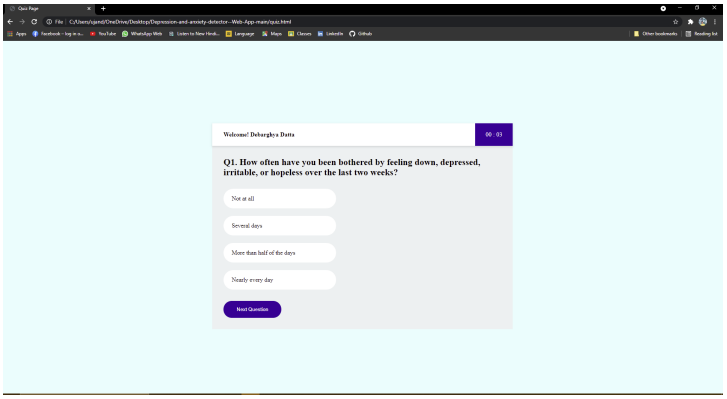
- Dhritesh Bhagat, Aritra Ray (Team 1)
- Debarghya Datta, Sneha Dasgupta (Team 2)
- Saptarshi Mondal, Sohini Chatterjee (Team 3)

We planned to execute the plan as the first and foremost thing needed would be the data collection and analysis for understanding human emotions. We hence began working on that, the most viable option was to get it using any form asking some questions, and the answers needed to be analyzed. We went for open-ended answers rather than using the MCQ-based versions, the reason being MCQ constraints our model to only those options and we don't get variations. Thus Team 1 started developing the emotion analysis using human interaction. Meanwhile, Team 2 worked on the UI for people to interact with that (we could have used Google Forms but then we wouldn't get the means to take the Rorschach test). We found a lot of algorithms for the Rorschach test as it is a very primary feature of basic health websites. We however had to work on our own version of text emotion analysis. We started by using basic classifier techniques like Naive Bayes, Log Reg, but we ended up getting an accuracy of 62%-69% at most. We then shifted our focus to more efficient algorithms with more accuracy in mind. Meanwhile, team3 has been working on the edge computing part where the data is just collected from sensors and then read directly in the cloud. Team2 since they have finished the basic UI structure shifted to web scraping and getting databases of words that can be used to train the model for team1. We first used bert by Google, i.e, an NLP developed by Google. But even in that, we got an accuracy of 72% giving the output in 0.06 seconds at its fastest.
We then started using NN for the model, we started with LSTM using the w2v method. In that, we convert each word into a vector, and then those vectors can be used to calculate the emotions. Even in LSTM, we got an accuracy of 74%. Finally, to our rescue, we figured out how to use CNN in our model, and with that, the present model stands at 75.66% accuracy with an output time rate of 0.01 seconds and the model can categorize data into five categories, namely:
joy, sadness, anger, neutral, fear.

Here are a few screenshots:



The form webpage we prepared:

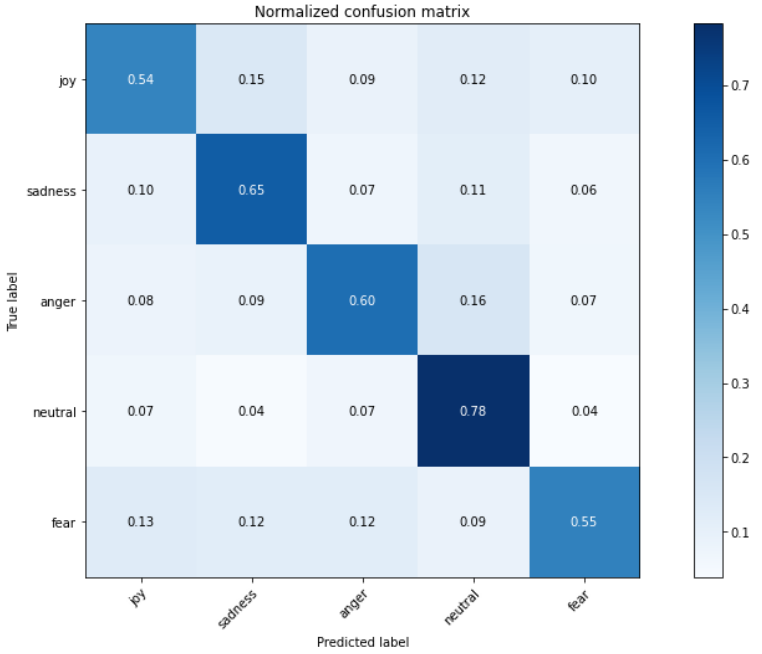


PROGRESS REPORT

Mentor: Nilanjana Dutta Roy

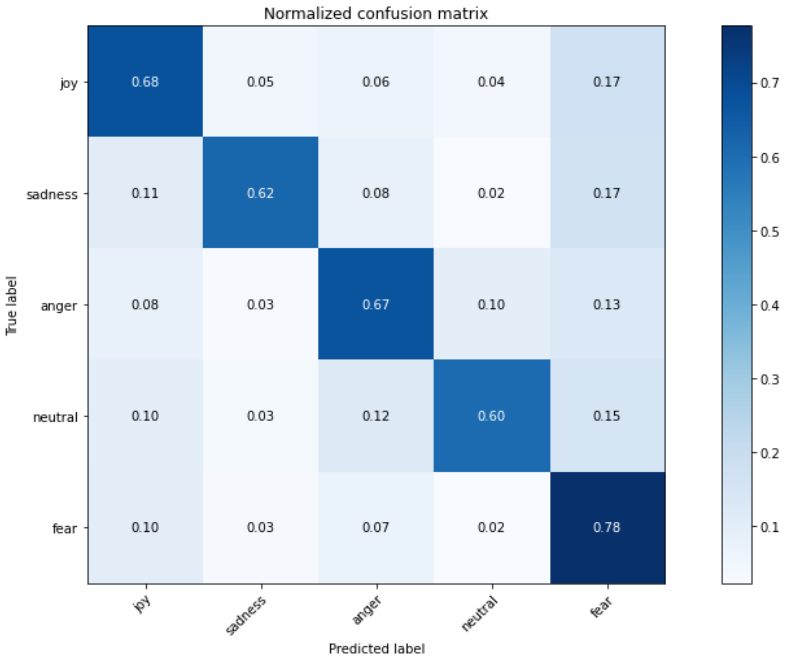
Random Forest:

```
Accuracy: 62.28%

F1 Score: 62.28

COnfusion Matrix:
 [[374 101  61  85  72]
 [ 65 443  50  77  44]
 [ 57  64 426 113  47]
 [ 43  27  44 499  25]
 [ 86  80  81  58 371]]
```



Naive Bayes:

```
Accuracy: 67.02%

F1 Score: 67.02

COnfusion Matrix:
 [[469  32  44  28 120]
 [ 73 420  55  16 115]
 [ 56  18 475  68  90]
 [ 61  20  76 385  96]
 [ 68  20  48  15 525]]
```

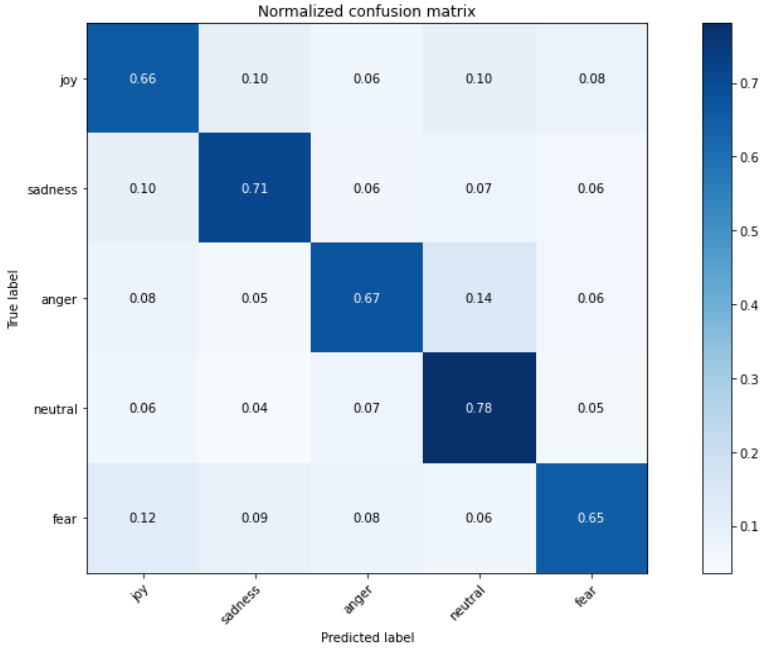


PROGRESS REPORT

Mentor: Nilanjana Dutta Roy

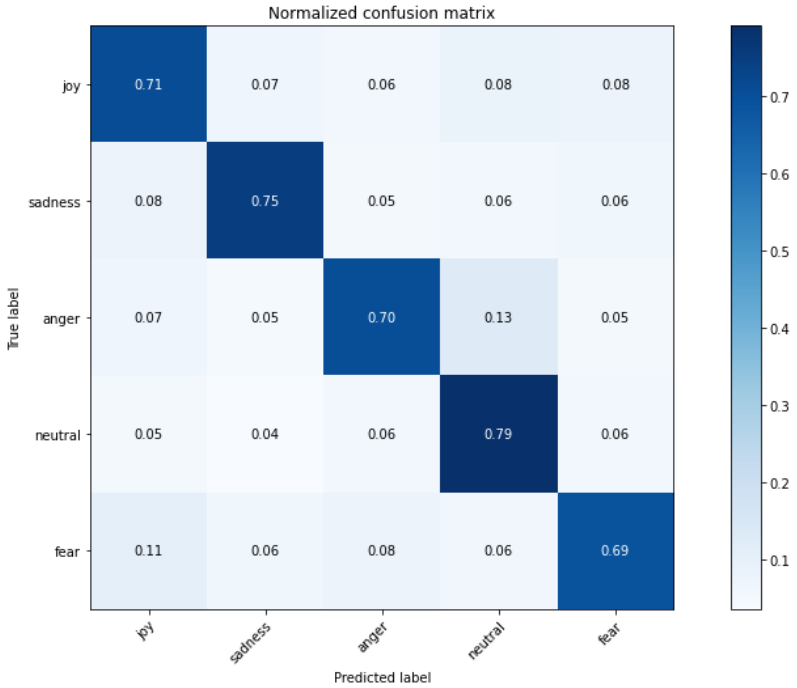
Log Reg:

```
Accuracy: 69.35%

F1 Score: 69.35

COnfusion Matrix:
 [[456  67  44  68  58]
 [ 65 483  42  50  39]
 [ 56  34 476 101  40]
 [ 41  23  42 498  34]
 [ 82  60  51  43 440]]
```



SVM

```
Accuracy: 72.71%

F1 Score: 72.71

COnfusion Matrix:
 [[490  49  41  58  55]
 [ 53 508  34  40  44]
 [ 50  33 498  91  35]
 [ 34  23  38 505  38]
 [ 72  43  53  42 466]]
```



PROGRESS REPORT

Mentor: Nilanjana Dutta Roy

BERT

```
[ ]  predictor = ktrain.get_predictor(learner.model, preproc)
     predictor.get_classes()
```

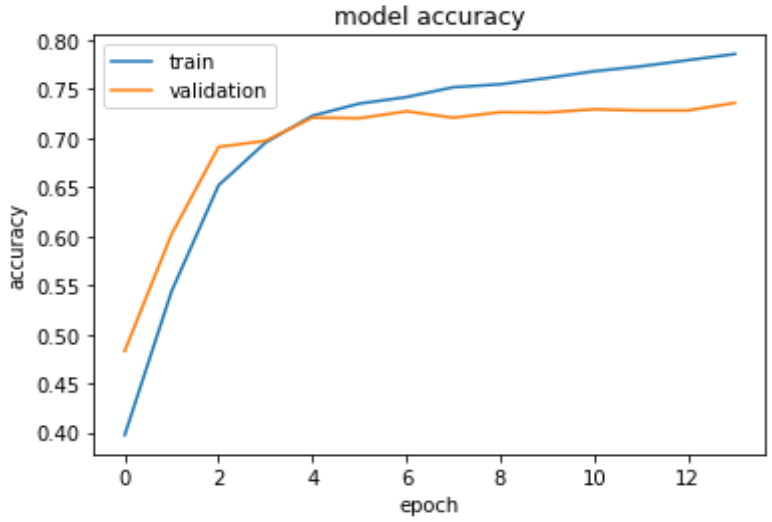
```
['joy', 'sadness', 'fear', 'anger', 'neutral']
```

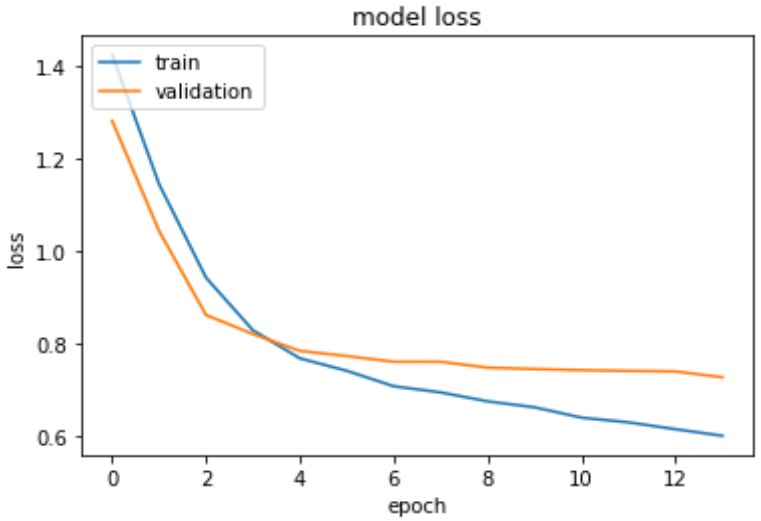
```
[ ]  import time

     message = 'I just broke up with my boyfriend'

     start_time = time.time()
     prediction = predictor.predict(message)

     print('predicted: {} ({:.2f})'.format(prediction, (time.time() - start_time)))
```

```
predicted: sadness (0.06)
```

LSTM using w2v

- **Model accuracy and loss graphs**

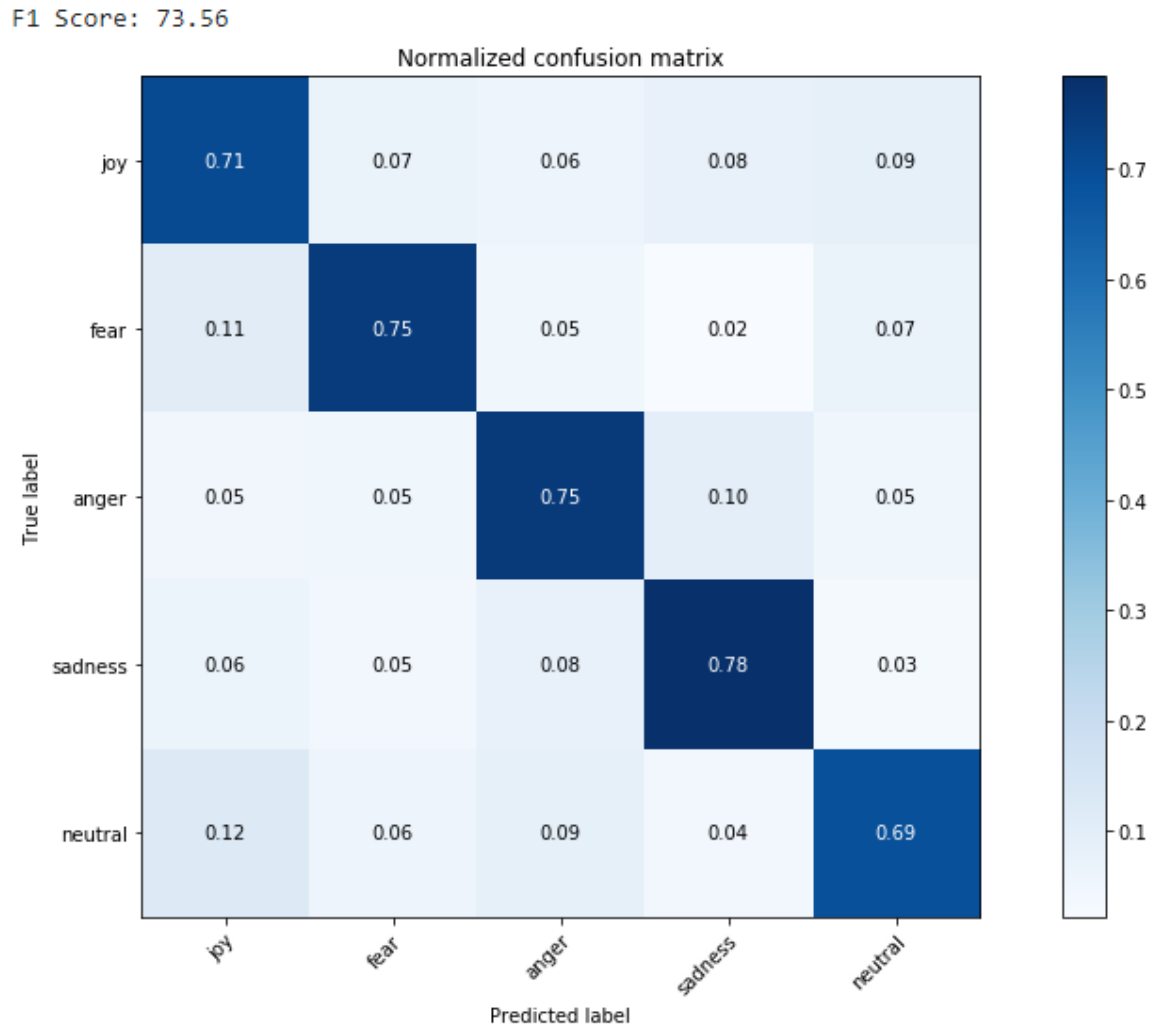




PROGRESS REPORT

Mentor: Nilanjana Dutta Roy

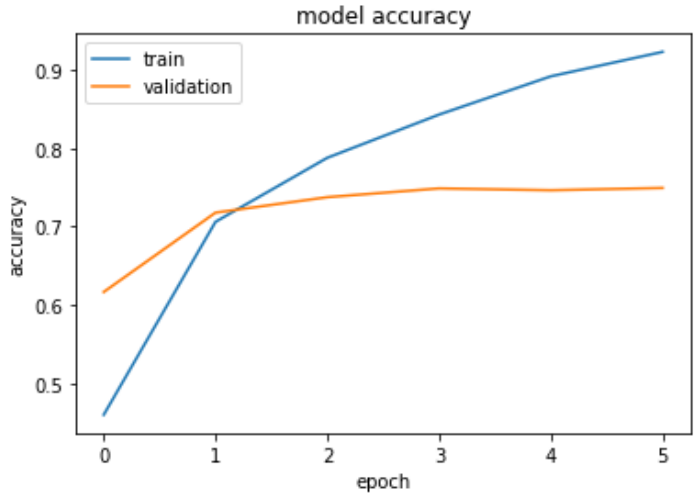
- **Matrix and f1 score**

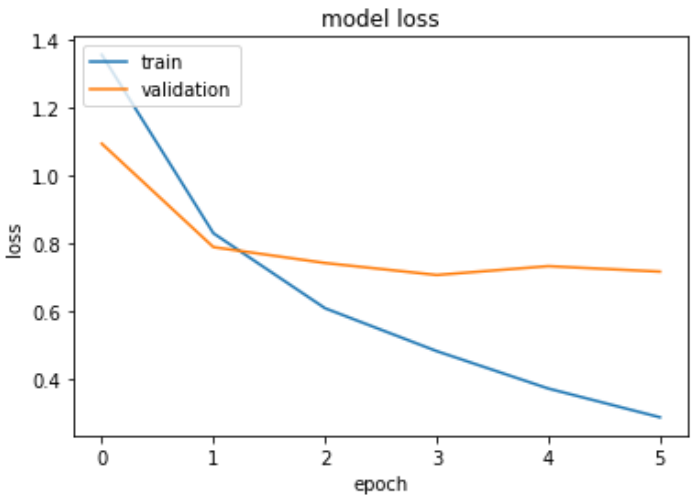


The time taken for prediction decreased from 0.1 to 0.05 seconds.

CNN using w2v:

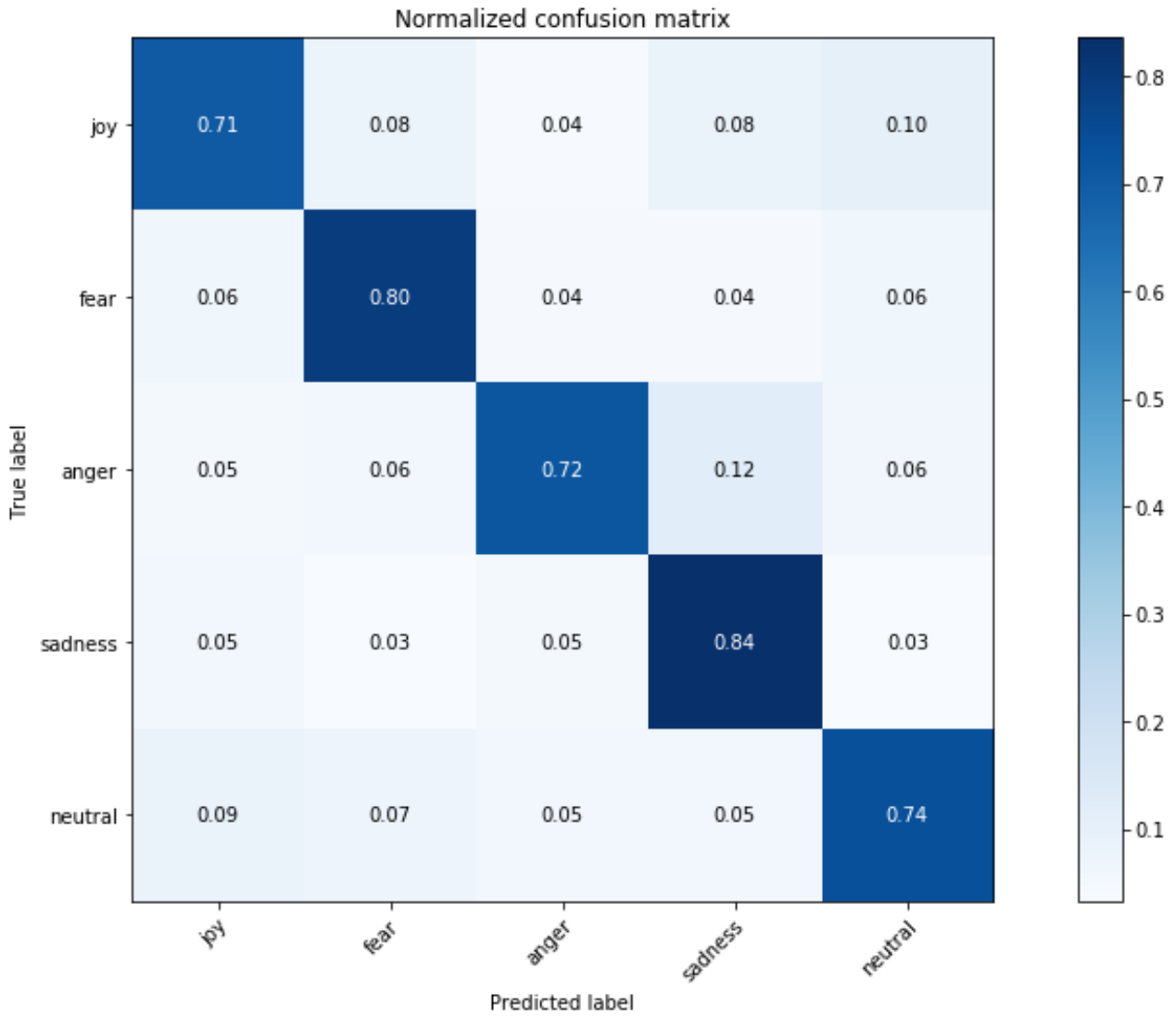
- **Model accuracy and loss graph:**





- **Matrix and f1 score:**

F1 Score: 75.66



PROGRESS REPORT

Mentor: Nilanjana Dutta Roy

Here are some screenshots of the text messages that we tested.

```
Message: ['I met my father after ten years!']
predicted: joy (0.19 seconds)

Message: ['Shit! Someone is following me.']
predicted: fear (0.22 seconds)

Message: ['People are dying because of COVID-19.']
predicted: sadness (0.22 seconds)

Message: ['delivery was hour late and my pizza was cold!']
predicted: sadness (0.06 seconds)

Message: ['I hate everyone around me.']
predicted: anger (0.20 seconds)

Message: ['I dont like going to the doctor.']
predicted: fear (0.21 seconds)

Message: ['Why are you not paying attention to me??']
predicted: anger (0.22 seconds)

Message: ['Bro! I got reaally good maarks.']
predicted: joy (0.20 seconds)

Message: ['I am confused about my life.']
predicted: sadness (0.21 seconds)

Message: ['Life has been toooooo eaasy.']
predicted: neutral (0.23 seconds)

Message: ['Why do we give exams']
predicted: anger (0.20 seconds)
```

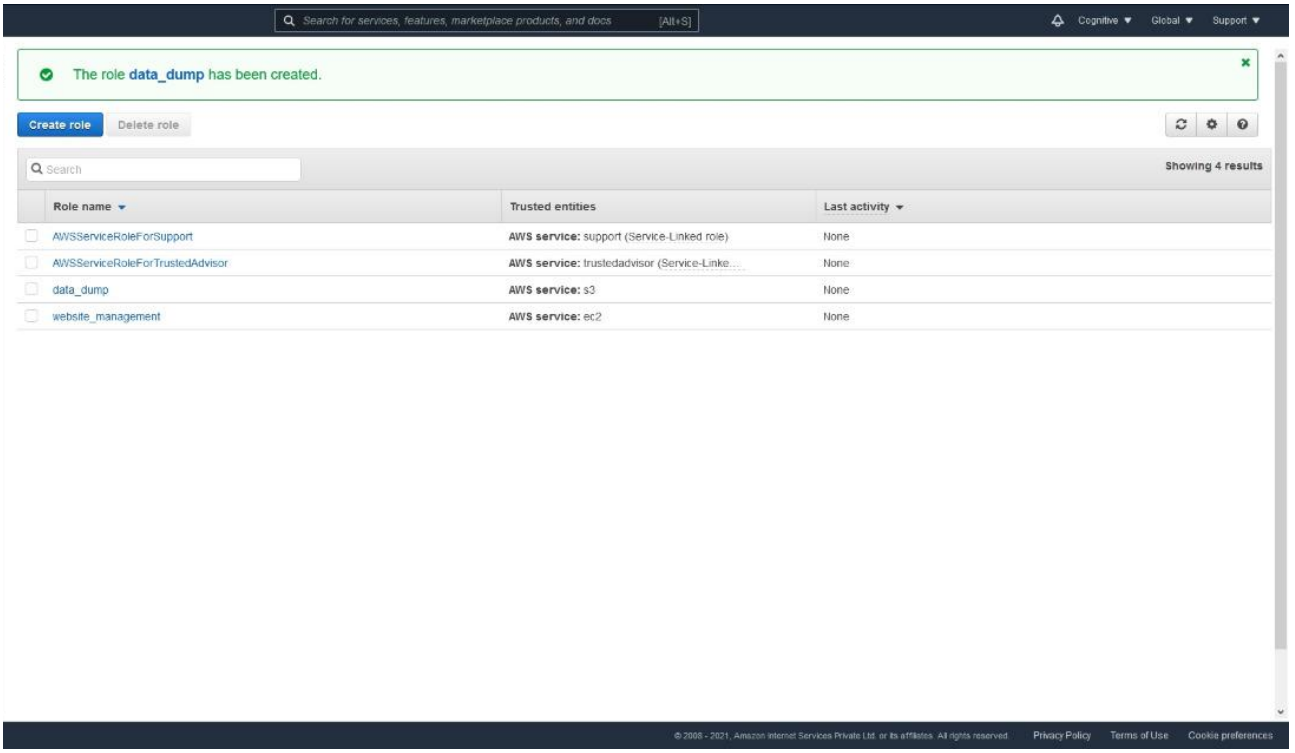
These messages are all in day-to-day text message format and the model can still predict them. These messages are all input-based and were not present in the training or testing set.

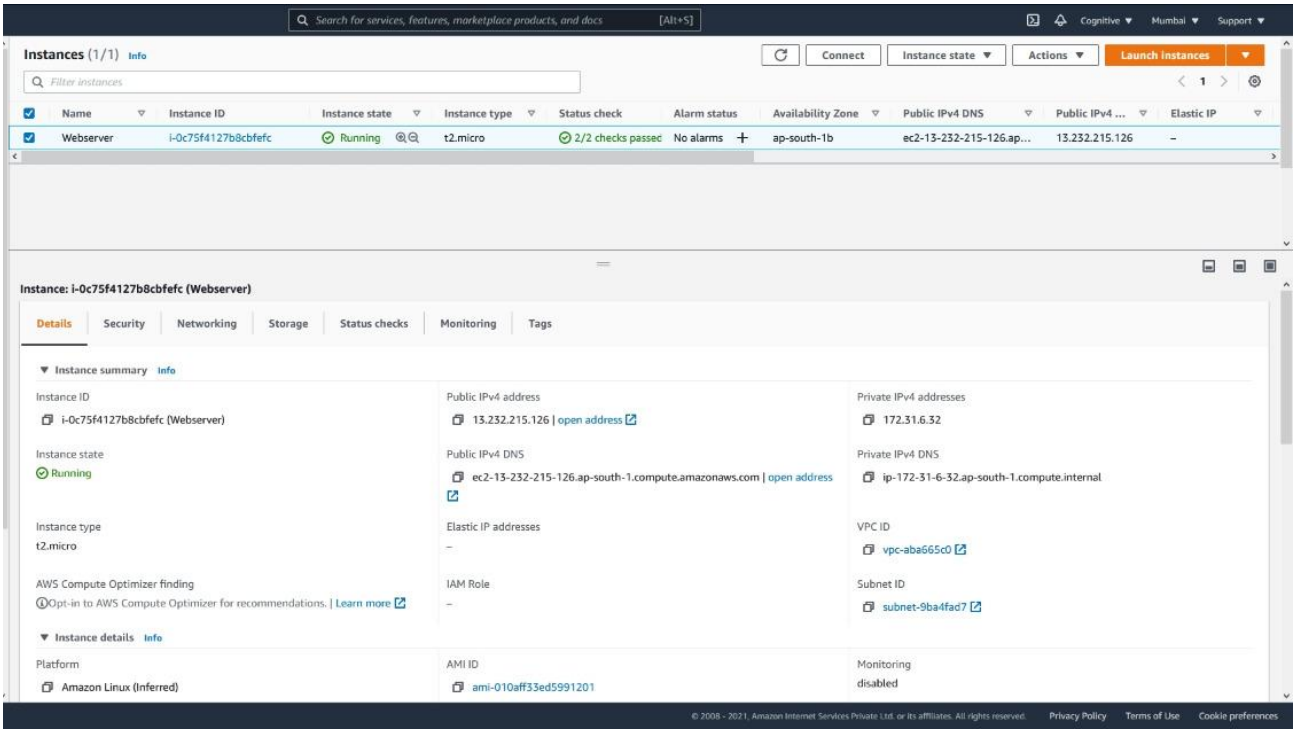
We can predict from these messages and other behavioral stimulations for further analysis.

PROGRESS REPORT

Mentor: Nilanjana Dutta Roy

Here are the screenshots of the cloud resources we have been using to read the data in real-time using sensors:





PROGRESS REPORT

Mentor: Nilanjana Dutta Roy

Here is the table of the work done!

Sl.no	Name	Third semester	Fourth semester
1.	Aritra Ray (Team1)	Documentation regarding the progress of similar projects and research papers giving up parameters to move forward with this project. Literature survey regarding the DSM-5 and hardware detection via EOG with ML incorporation.	Stressed on developing a Sentiment Analysis setup which will be based on the defined quiz provided to give us a combined score of multi-layered sentiment analysis. Worked on the traditional and NN methods for the development of the model providing us a visual of customer mental stability parameters after the quiz is taken by the same.
2.	Dhritesh Bhagat (Team1)	The study of various solutions to our problem which will be efficient in detecting real human emotions. Technological field study and existing tech resources moving forward with the project.Literature survey regarding in-depth resources like MEMO-box and psychodynamic treatment.	Stressed on developing a Sentiment Analysis setup which will be based on the defined quiz provided to give us a combined score of multi-layered sentiment analysis. Worked on the bert and NN methods for the development of the model providing us a visual of customer mental stability parameter after the quiz is taken by the same.
3.	Debarghya Datta (Team2)	Studied different approaches to our problem aimed with technological field help and existing tech resources moving forward with the project. Literature survey and prospects of app and web development for our project.	Developed the Front-end of the web app which would serve as the medium of interaction between users and our project. Added sample questions based on anxiety queries and subjects understudy for mental health concerns.
4.	Sneha Dasgupta (Team2)	Studied different approaches to our problem aimed with technological field help. Literature survey on the application of wristwatches and how they are useful.	Worked mainly on the web scraping part getting data for the CNNs to work. Developed the Front-end of the web app which would serve as the medium of interaction between users and our project.
5.	Sohini Chatterjee (Team3)	Studied various research papers, focussing on SVM, feature learning, and transfer learning, to collect data for the literature survey, and also worked on the documentation of the project.	Worked on the web scraping and also working on creating containers in the cloud that can store the entire application together. Worked on creating an AWS working environment by utilizing EC2, IAM, RDS.

6.	Saptarshi Mondal (Team3)	Studied the different services of AWS that can be utilized for cloud computing and reduce the load on the local peripheral. Studied the edge-level computing that can be used.	Worked on creating the pipeline for the data to travel and the sensor data. Worked on creating an AWS working environment by utilizing EC2, IAM, RDS.