

INTERVIEW EMULATOR



BY

JAYANTA GHOSH

S.P. Mandali's
Ramnarain Ruia Autonomous College
Matunga, Mumbai-400019

Project Report
on
Interview Emulator

Project Guide
Prof. Abhijeet Gole

Project By
Jayanta Ghosh

Department of Computer Science Information Technology
2020-21

INDEX

SR. NO.	PARTICULARS	PAGE NO.
1	Preface	4
2	Project Overview <ul style="list-style-type: none"> • Project Title & Definition 	5
3	Introduction <ul style="list-style-type: none"> • Detailed explanation of study area 	5
4	Literature Review <ul style="list-style-type: none"> • Study undertaken on similar concept • Citation 	6
5	Research Methodology <ul style="list-style-type: none"> • Objective • Scope of study • Limitations • Significance of Research • Description of Algorithms / procedures / data collection methods / logic 	9
6	Experimental setup <ul style="list-style-type: none"> • Tools • Architecture/Framework • Software Language 	13
7	Results <ul style="list-style-type: none"> • Screen Layout • Testing report 	19
8	Summary and conclusion <ul style="list-style-type: none"> • Findings of Project work • Application area 	39
9	Further Research	40
10	Bibliography <ul style="list-style-type: none"> • Books • References (Conference Papers/ Journal Papers etc.) • Websites 	41

Preface

I, Mr. Jayanta Badal Ghosh, student of Ramnarain Ruia Autonomous College, acknowledge my sincere gratitude to all those who helped me in the project work of MSc. Computer Science. I extend my gratitude towards my project guide, Prof Abhijeet Gole Sir to help me to go ahead with my project and appreciating my work at every stage. The Project has been made by own effort that I have learned in our Academics duration. This Project is Big Data based Interview Emulator. Through I have acquired a knowledge about Big data and tried out to present this project properly. In preparing this project I have taken help from reference books.

Project Overview:-

Interview Emulator is a deep learning web application which provides Video Audio and text way through interviewing a candidate and find out analysis of result by comparing other candidates.

Introduction:-

Emotion Recognition is a vital area of research to enable effective human-computer interaction(HCI). Human-computer interaction is that the study, planning, and style of how people and computers work together in order that a person's needs are satisfied within the best way. Human emotions are often detected using speech signal, facial expressions, visual communication, and electroencephalography (EEG). Humans are capable of manufacturing thousands of facial actions during communication that change in complexity, intensity, and meaning. The addition or absence of 1 or more facial actions may alter its interpretation.

Affective computing is a field of Machine Learning and Computer Science that studies the recognition and the processing of human affects. It is an interdisciplinary field spanning computational & computation science, psychology, and cognitive science. Interview Emulator is a relatively new discipline that aims to include text inputs, as well as sound and video.

A Convolutional Neural Network (ConvNet/CNN) may be a Deep Learning algorithm which may absorb an input image, assign importance (learnable weights and biases) to varied aspects/objects within the image and be ready to differentiate one from the opposite

Literature Review:-

Emotion Recognition is been part of many fields now, by detecting the emotion of humans by using facial appearance, Program body shape, etc. The Facial image is periodically used to identify emotions. The algorithm is composed of three stages which include image processing stage, facial feature extraction stage, and emotion detection stage. The expected feature extraction method consists of three regions: eye region ancillary region and mouth region. Mainly Fuzzy classifier is used for detecting emotion stages and performing the Linear Matrix Inequality Optimization (LMI) Method which can be a very effective way to overcome this difficulty as it finds the vague patterns in the input emotion recognition system. Specifically detecting the whole face we have to extract the face region and that can be done by Virtual Face Model (VFM) based Histogram method to get information on face region and its components. [1]

Psychologists suggest that six basic emotions are extensive and occur in every art. Those are happiness, sadness, fear, anger, disgust, and surprise. Human Emotions convey by mimics and play an elementary role in everyday communication. Human Brain remembers to simulate in a split second. But for machine interaction, we had to use the natural dialog between use and computer system. Modern face emotion recognition systems receive as input both pictures and videos. Three main phases of face capturing, face features, and facial emotion recognition, face capturing include face detection and tracking that is linked with the evaluation of face position. Eradication of face features composes feature detection and tracking. Broadly the approaches used to solve the problem classified into local feature detection, global model detection, and hybrid systems. Emotion Recognition is a vital area of research to enable effective Human-Computer Interaction (HCI). [2]

By these Papers, it is found that video emotion recognition through face detection by various algorithms and regions of the face by eye, ancillary, and mouth. Human-computer interaction is that the study, planning, and style of how people and computers work together so that a person's needs are convinced within the best way. Human emotions are often detected using speech signals, facial expressions, visual communication, and electroencephalography (EEG). Humans are capable of manufacturing thousands of facial actions during communication that change in complexity, intensity, and meaning. The inclusion or omission of one or more facial actions may change its analysis.

Personality Psychology is the major aspect of text emotion analysis it's to uniquely identify person's text and their behaviour. The mechanisms should ensure one-to-one interactions with the social, anyone mind, or personality. It is used to depict individuals and each other searches using natural language and theoretically based personality models. Personality traits it's an alternative approach to individual linguistic markers. It showed that multiple correlations between linguistic markers and the Big Five traits like Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness. Neuroticism is positively correlated allow the use of negative emotion words and negatively correlated allow the use of Positive emotion words, Interaction between Grammatical markers and personality traits for the research in way of automated personality classification[3]. Extraversion is outstanding component in personality psychology as information by its presence in most personality measures, it typically identifies viscosity of geniality. Relates to the experience of positive emotions, and character are more likely to take on leadership roles and have more friends. Conscientiousness is linked to an individual's degree of self-discipline, as well as need for achievement order and tenacity. Openness defines by genius and disregard for the convention it seems possible that these traits are analogous to career success. Agreeableness is a drawback that means any person maybe

prone to job-hopping, immensely satisfying persons may sacrifice their success in amusing others. [8]

Personality traits are measured through various applications and surveys such as the Big five inventory however this is an option to interpret particular semantic indicators. Psychological effort subsists of organized comparably enduring traits and tools that signifies one's Interactions, with the physical and social environments. To extract any information from any Pretext, LWIC is applied. LWIC indicates as Grammatical analysis and Word results. It is a text analysis tool that implements a Profitable, and Comprehensive approach for studying the various affecting, mental and anatomical present in original samples. It targets words at a time by searching through lexicon file. [3]

Emotional Recognition aims at undoubtedly identifies the real state of a human voice. Correct disclosure of emotion from speech has clear gains for the design of natural human-machine interfaces. It's been seen that it plays a vital role-playing in many computer fields including health care, children education, etc. The Suggested structure aims at the identification of basic emotional states such as anger, joy, happiness, neutral, and sadness. [4]

Affective Computing has produced extreme suction over the past several layer insufficiency of widely accepted datasets for emotion recognition that can be used for criterion and ease fair comparisons. The text-based emotion recognition supports the emotional and business as a new path to measuring and illustrating individual and mutual emotional states. There have been many utilized methods from conventional robotics. It can found by a procedure like various indicators like crucial signs such as heart rate, muscle activity, or sweat production on the surface of the skin. It can handle both unimodal and modal analysis. Primarily Affective Computing handles both speech gesture and written report. [5]

Affective computing is a field of neural network and data processing that studies the understanding and plenty of human effects. It is an integrative field spanning commutation & calculation science, psychology, and cognitive science. Interview Emulator is an approximately new discipline that intent to include text inputs, as well as sound and video.

Various Applications of human-computer interaction from emotion recognition through speech for affective learning systems, mental health analysis, improvement of customer service, etc. To record the user's emotional states through customer service can help the agent to comply with his/her feedback to provide a better service. Speech Processing and fusion of the most common methods which were the detecting emotion from transcripts. It Spotlights on the systems using either acoustic or textual information and fusion of both which derived both. Speech emotion recognition can be done in two ways exploring emotion representatives features or building classifiers adapted to emotion recognition. Current Scenario it's even possible to work on both using a deep learning framework. [6]

The emotion could be analyzed by the variations between the speaking rate, pitch tone, and tone intensity and the quality of the voice also matters a lot. In particular pitch and intensity seem to be correlated to activation, so that high frequency and intensity values imply high, low frequency and intensity values low activation.[7]

In addition to the paper research works, it is also evident to note that, In these Pandemic Situations the whole world is facing crisis and the most hard-hit are the companies who have been experiencing uncertainty due to the COVID-19. The companies are left to a dilemma such as they cannot recruit candidates even though there are vacancies in the companies, because of the COVID-19 contraction fear and dilemma. More over the candidate could use the time productively by

preparing for the interviews through this tool, which would be expected to analyse the emotions accordingly and prepare the interviewee for the same.

Citations

- [1] Y. H. J. J. B. P. Moon Hwan Kim, "ResearchGate," Kunsan National University, 19 February 2014. [Online]. Available: https://www.researchgate.net/publication/228870902_Emotion_Detection_Algorithm_Using_Frontal_Face_Image. [Accessed July 2015].
- [2] P. N. Elżbieta Kukla, "Facial Emotion Recognition Based on Cascade of Neural Networks," *Springer Link*, vol. 314, pp. 67-78, 2015.
- [3] J. C. U. B. A. L. P. C. K. C. a. R. d. D. B. Edward P. Tighe, *Personality Trait Classification of Essays with the Application of Feature*, Manila: Semantic Scholar, 2016.
- [4] M. A. P. Wanare, "Human Emotion Recognition From Speech," *Int. Journal of Engineering Research and Applications*, vol. 4, no. 7, pp. 74-78, 2014.
- [5] S. I. M. K. S. F. a. H. P. Bernhard Kratzwald, *Deep learning for affective computing: text-based emotion recognition in decision support*, Tokyo: <https://arxiv.org/abs/1803.06397>, 2018.
- [6] R. P. P. K. J. V. Y. C. a. N. D. Jaejin Cho, *Deep neural networks for emotion recognition combining audio and transcripts*, Hyderabad: <https://arxiv.org/abs/1911.00432v1>, 2019.
- [7] E. A. a. J. W. Thuriid Vogt, "Automatic Recognition of Emotions from Speech: A Review of the Literature and Recommendations for Practical Realisation," *Affect and Emotion in Human-Computer Interaction*, vol. 4868, pp. 75-91, 2008.
- [8] C. A. H. C. J. T. a. M. R. B. Timothy A. Judge, "THE BIG FIVE PERSONALITY TRAITS, GENERAL MENTAL ABILITY, AND CAREER SUCCESS ACROSS THE LIFE SPAN," *Personnel Psychology*, vol. 52, no. 3, pp. 621-652, 1999.

Research Methodology

Objective:

The main objective is to develop state of the art technology in order to analyze and recognise emotion of candidates appearing for Interview and visualize the outcome by comparing with another candidates.

Scope of Study:

The main and foremost motive of the prototype is to analyze the emotions of the emotions during the time course of interview and deciding whether the candidate is eligible to join the company or not by providing the data analysis of the current candidate in comparison to other counterparts. Means for e.g Companies because of COVID - 19 are being converted to work from anywhere, so first and foremost advantage is that remote work, let's say a person who is located in Czech Republic and he is highly talented company is Indian based so company could hire from different continents also Second is that it would lower the cost of the whole hiring life cycle but at the same time hiring cycle would be fast.

Limitations :-

While this emulator/simulator really lowers the costs involved in the hiring process and the lame manual work that could be reduced to a greater extent, it comes with some cost that has been illustrated below :-

1. the accuracy for text model is bit low, since there is a great amount of ambiguities which is a major challenge of NLP.
2. Since Model are based on neural networks amount of memory and CPU requirement is pretty high.

Description of Algorithm

The prototype should be flexible as well as more faster to compute tasks, so selection of the right architecture is very important, as it is directly in proportional with the quality and robustness of the prototype. The prototype will be expected to house three types of models in it.

Our aim is to develop a model able to provide a live sentiment analysis with a visual user interface. Therefore, we have decided to separate three types of inputs :

a.Video Analysis

The Xception architecture is based on DepthWise Separable convolutions that allow to train much fewer parameters.

When it comes to applying CNNs in real life application, being able to explain the results is a great challenge. We can indeed plot class activation maps, which display the pixels that have been activated by the last convolution layer. We notice how the pixels are being activated differently depending on the emotion being labeled. The happiness seems to depend on the pixels linked to the eyes and mouth, whereas the sadness or the anger seem for example to be more related to the eyebrows.

b.Audio Analysis

The model we have chosen is a Time Distributed Convolutional Neural Network.

The main idea of a Time Distributed Convolutional Neural Network is to apply a rolling window (fixed size and time-step) all along the log-mel-spectrogram.

Each of these windows will be the entry of a convolutional neural network, composed by four Local Feature Learning Blocks (LFLBs) and the output of each of these convolutional networks will be fed into a recurrent neural network composed by 2 cells LSTM (Long Short Term Memory) to learn the long-term contextual dependencies. Finally, a fully connected layer with softmax activation is used to predict the emotion detected in the voice.

SVM

Classical approach for Speech Emotion Recognition consists in applying a series of filters on the audio signal and partitioning it into several windows (fixed size and time-step). Then, features from time domain (Zero Crossing Rate, Energy and Entropy of Energy) and frequency domain (Spectral entropy, centroid, spread, flux, rolloff and MFCCs) are extracted for each frame. We compute then the first derivatives of each of those features to capture frame to frame changes in the signal. Finally, we calculate the following global statistics on these features: *mean, median, standard deviation, kurtosis, skewness, 1% percentile, 99% percentile, min, max* and *range* and train a simple SVM classifier with rbf kernel to predict the emotion detected in the voice.

To limit overfitting, tuned the model with :

- Audio data augmentation
- Early stopping
- And kept the best model

c.Text Analysis

Model

We have chosen a neural network architecture based on both one-dimensional convolutional neural networks and recurrent neural networks. The one-dimensional convolution layer plays a role comparable to feature extraction : it allows finding patterns in text data.

The Long-Short Term Memory cell is then used in order to leverage on the sequential nature of natural language : unlike regular neural network where inputs are assumed to be independent of each other, these architectures progressively accumulate and capture information through the sequences. LSTMs have the property of selectively remembering patterns for long durations of time.

Our final model first includes 3 consecutive blocks consisting of the following four layers : one-dimensional convolution layer - max pooling - spatial dropout - batch normalization. The numbers of convolution filters are respectively 128, 256 and 512 for each block, kernel size is 8, max pooling size is 2 and dropout rate is 0.3.

Following the three blocks, we chose to stack 3 LSTM cells with 180 outputs each. Finally, a fully connected layer of 128 nodes is added before the last classification layer.

Data Collection:-

Video Analysis

The data set used for training is the Kaggle FER2013 emotion recognition data set :
<https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data>.

The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centered and occupies about the same amount of space in each image. The task is to categorize each face based on the emotion shown in the facial expression in to one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

Audio Analysis

The data set used for training is the Ryerson Audio-Visual Database of Emotional Speech and Song:

<https://zenodo.org/record/1188976#.XA48aC17Q1J>

RAVDESS contains 24 professional actors (12 female, 12 male), vocalizing two lexically-matched statements in a neutral North American accent. Speech includes calm, happy, sad, angry, fearful, surprise, and disgust expressions, and song contains calm, happy, sad, angry, and fearful emotions. Each expression is produced at two levels of emotional intensity (normal, strong), with an additional neutral expression.

Text Analysis

Stream-of-consciousness

<https://drive.google.com/file/d/1bbbn8kSBmcVObafdzAQEipRBc4SVqwtb/view?usp=sharing>

We are using data that was gathered in a study by Pennebaker and King [1999]. It consists of a total of 2,468 daily writing submissions from 34 psychology students (29 women and 5 men whose ages ranged from 18 to 67 with a mean of 26.4). The writing submissions were in the form of a course unrated assignment. For each assignment, students were expected to write a minimum of 20 minutes per day about a specific topic. The data was collected during a 2-week summer course between 1993 to 1996.

Each student completed their daily writing for 10 consecutive days. Students' personality scores were assessed by answering the Big Five Inventory (BFI). The BFI is a 44-item self-report questionnaire that provides a score for each of the five personality traits. Each item consists of short phrases and is rated using a 5-point scale that ranges from 1 (disagree strongly) to 5 (agree strongly). An instance in the data source consists of an ID, the actual essay, and five classification labels of the Big Five personality traits. Labels were originally in the form of either yes ('y') or no ('n') to indicate scoring high or low for a given trait. It is important to note that the classification labels have been applied according to answers to a rather short self-report questionnaire : there might be a non-negligible bias in the data due to both the relative simplicity of the BFI test compared to the complexity of psychological features, and the cognitive biases preventing users from providing a perfectly accurate assessment of their own characteristics.

Experimental Setup

Software languages and Tools:-

Web Application Micro-Framework for designing Visual Interface:-

Flask

Flask is a lightweight WSGI web application framework. It is designed to make getting started quick and easy, with the ability to scale up to complex applications. It began as a simple wrapper around Werkzeug and Jinja and has become one of the most popular Python web application frameworks.

Flask is often referred to as a micro framework. It aims to keep the core of an application simple yet extensible. Flask does not have built-in abstraction layer for database handling, nor does it have form a validation support. Instead, Flask supports the extensions to add such functionality to the application.

For Encoding/Decoding The Video :-

PyAudio

PyAudio provides Python bindings for PortAudio, the cross-platform audio I/O library. With PyAudio, you can easily use Python to play and record audio on a variety of platforms. PyAudio is inspired by:

- pyPortAudio/fastaudio: Python bindings for PortAudio v18 API.
- tkSnack: cross-platform sound toolkit for Tcl/Tk and Python.

For Image Processing:-

Pillow (Python Image Library)

PIL is the Python Imaging Library by Fredrik Lundh and Contributors.

The Python Imaging Library adds image processing capabilities to your Python interpreter.

This library provides extensive file format support, an efficient internal representation, and fairly powerful image processing capabilities.

The core image library is designed for fast access to data stored in a few basic pixel formats. It should provide a solid foundation for a general image processing tool.

For Data Analysis :-

Pandas

pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

Pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

For Data Visualization:-

Seaborn And Matplotlib

Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

Seaborn helps you explore and understand your data. Its plotting functions operate on dataframes and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots. Its dataset-oriented, declarative API lets you focus on what the different elements of your plots mean, rather than on the details of how to draw them.

Matplotlib

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+.

Neural Network Creation :-

Keras

Keras is a high-level neural networks library, written in Python and capable of running on top of either TensorFlow or Theano. It was developed with a focus on enabling fast experimentation. Being able to go from idea to result with the least possible delay is key to doing good research.

Keras if you need a deep learning library that:

- Allows for easy and fast prototyping (through total modularity, minimalism, and extensibility).
- Supports both convolutional networks and recurrent networks, as well as combinations of the two.
- Supports arbitrary connectivity schemes (including multi-input and multi-output training).
- Runs seamlessly on CPU and GPU.

Tensorflow

TensorFlow has APIs available in several languages both for constructing and executing a TensorFlow graph. The Python API is at present the most complete and the easiest to use, but other language APIs may be easier to integrate into projects and may offer some performance advantages in graph execution.

Front End:-

Javascript

JavaScript is a programming language commonly used in web development. It was originally developed by Netscape as a means to add dynamic and interactive elements to websites. ... Like server-side scripting languages, such as PHP and ASP, **JavaScript** code can be inserted anywhere within the HTML of a webpage.

CSS

Cascading Style Sheets is a style sheet language used for describing the presentation of a document written in a markup language such as HTML. CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript.

HTML

Hypertext Markup Language is the standard markup language for documents designed to be displayed in a web browser. It can be assisted by technologies such as Cascading Style Sheets and scripting languages such as JavaScript.

D3.js

D3.js is a JavaScript library for producing dynamic, interactive data visualizations in web browsers. It makes use of Scalable Vector Graphics, HTML5, and Cascading Style Sheets standards. It is the successor to the earlier Protovis framework.

D3 combines powerful visualization and interaction techniques with a data-driven approach to DOM manipulation, giving you the full capabilities of modern browsers and the freedom to design the right visual interface for your data.

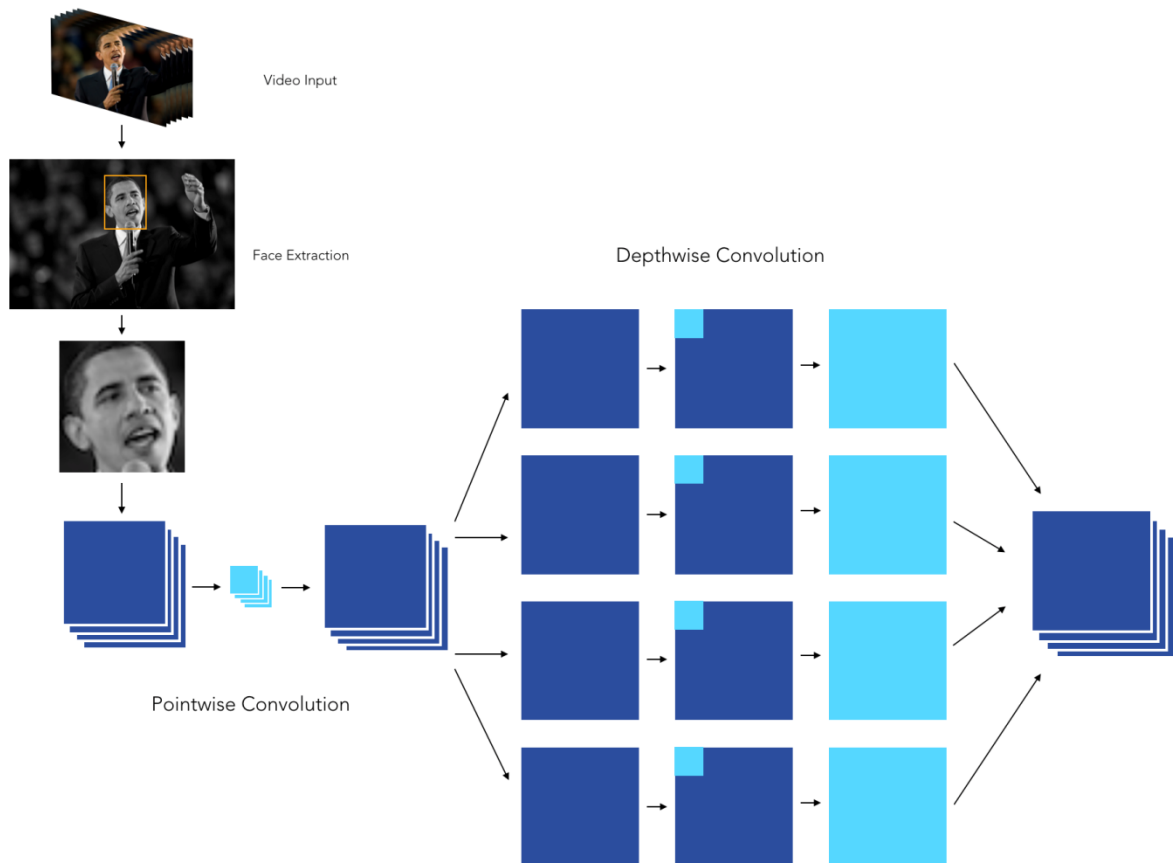
Architecture Framework

Video Framework Setup:-

Pipeline:-

The video processing pipeline was built the following way :

- Launch the webcam
- Identify the face by Histogram of Oriented Gradients
- Zoom on the face
- Dimension the face to 48 * 48 pixels
- Make a prediction on the face using our pre-trained model
- Also identify the number of blinks on the facial landmarks on each picture



Audio Framework Setup:-

Pipeline

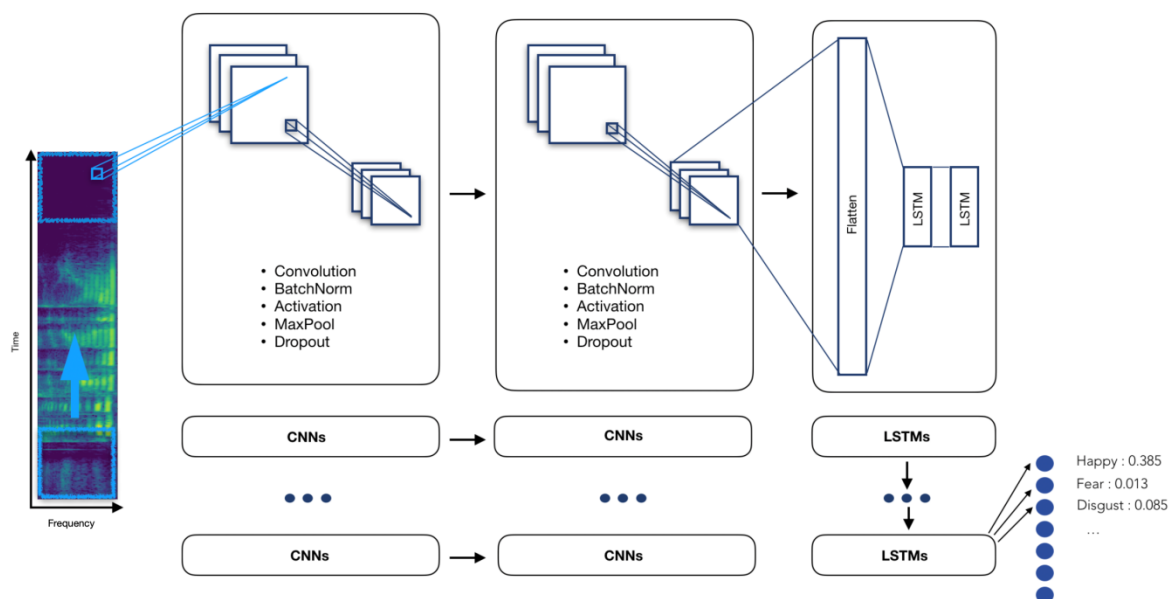
The speech emotion recognition pipeline was built the following way :

- Voice recording

- Audio signal discretization
- Log-mel-spectrogram extraction
- Split spectrogram using a rolling window
- Make a prediction using our pre-trained model

SVM classification pipeline:

- Voice recording
- Audio signal discretization
- Apply pre-emphasis filter
- Framing using a rolling window
- Apply Hamming filter
- Feature extraction
- Compute global statistics
- Make a prediction using our pre-trained model

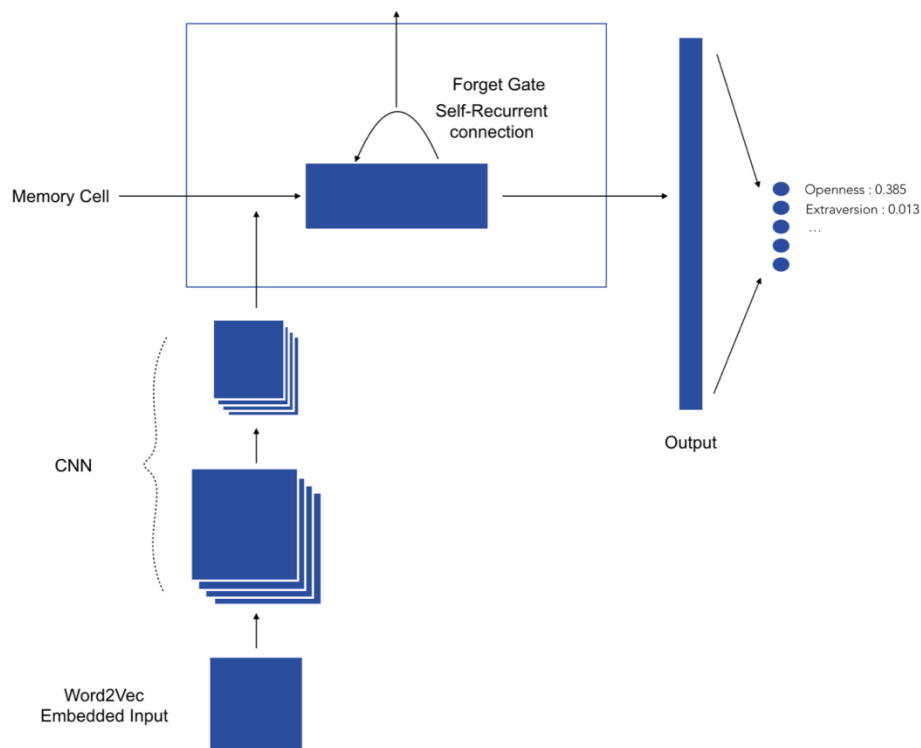


Text Framework Setup:-

The text-based personality recognition pipeline has the following structure :

- Text data retrieving
- Custom natural language preprocessing :

- Tokenization of the document
- Cleaning and standardization of formulations using regular expressions
- Deletion of the punctuation
- Lowercasing the tokens
- Removal of predefined *stopwords*
- Application of part-of-speech tags on the remaining tokens
- Lemmatization of tokens using part-of-speech tags for more accuracy.
- Padding the sequences of tokens of each document to constrain the shape of the input vectors.
- 300-dimension Word2Vec trainable embedding
- Prediction using our pre-trained model



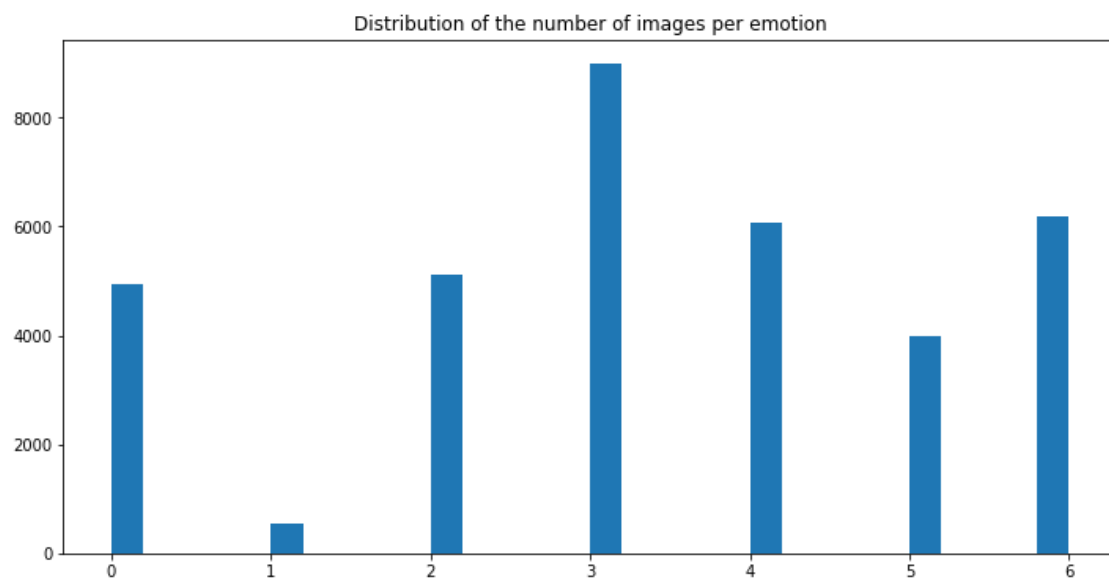
Hardware Specifications:-

- OS: Windows 10.
- Code Editor: VS Code, Python Idle, Jupyter Notebook.
- RAM: 4GB

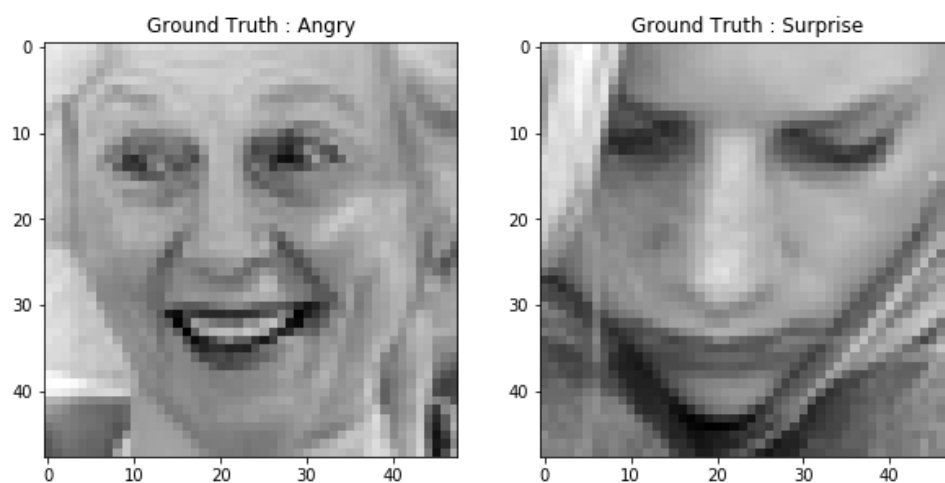
- HDD:500MB(Minimum)
- GPU: Recommended Nvidia for ML Training Model

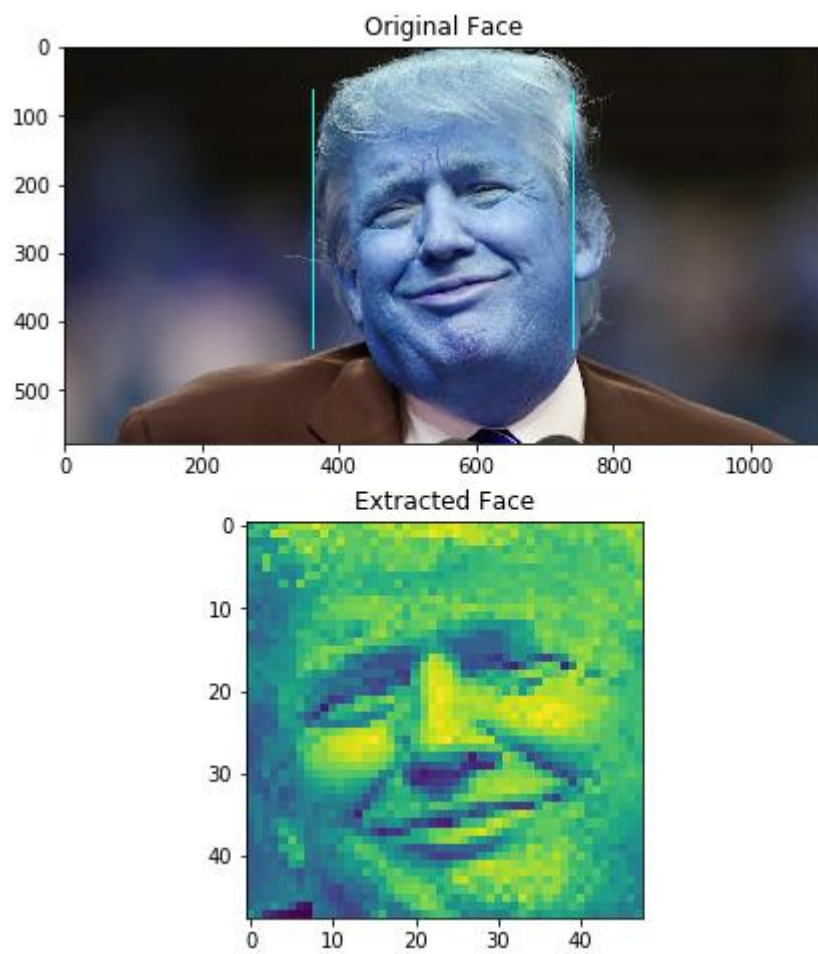
Results:-

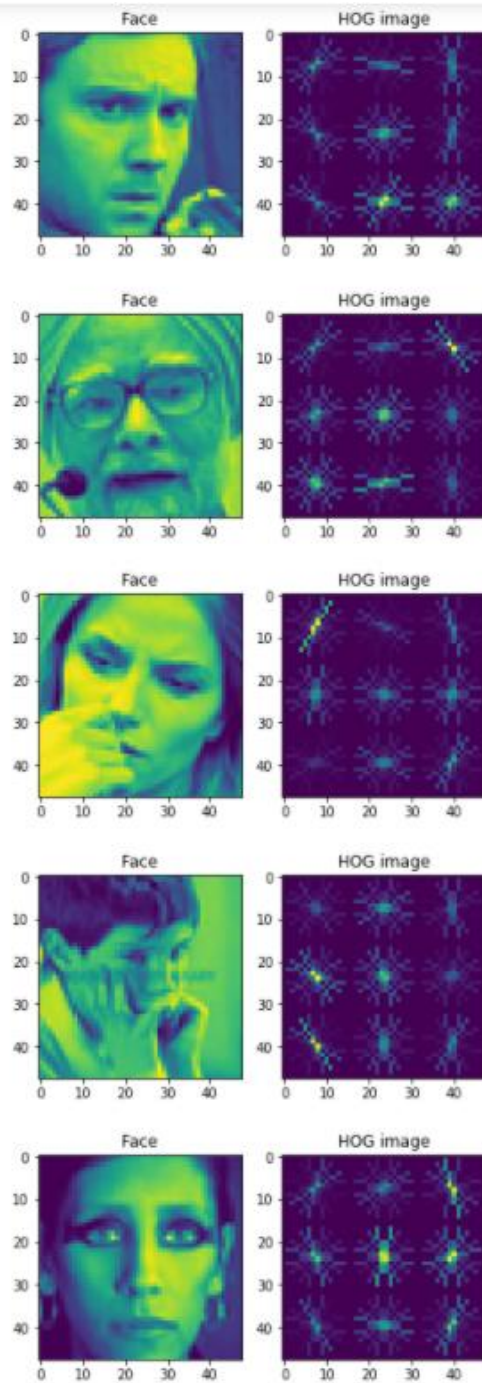
Preprocessing and Training of Video analysis



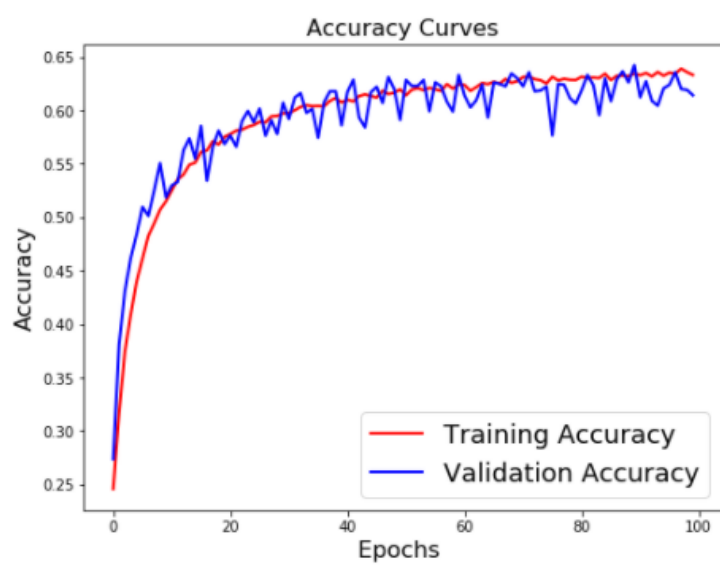
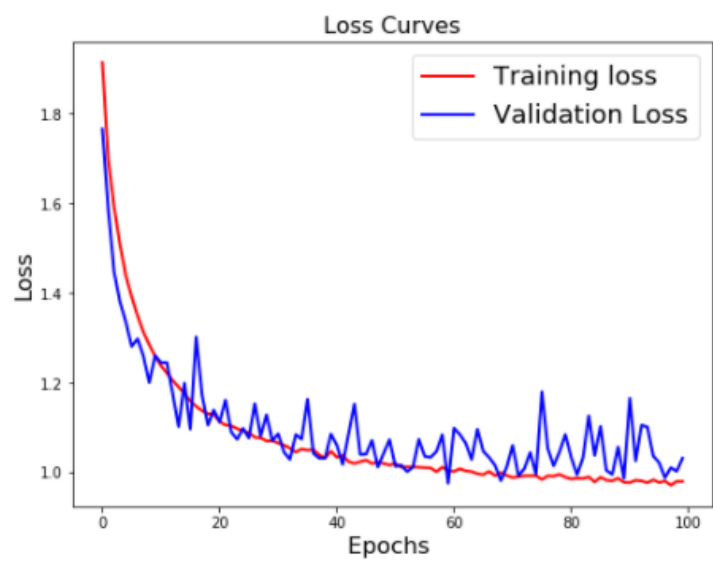
```
Out[45]: Text(0.5, 1.0, 'Ground Truth : Surprise')
```

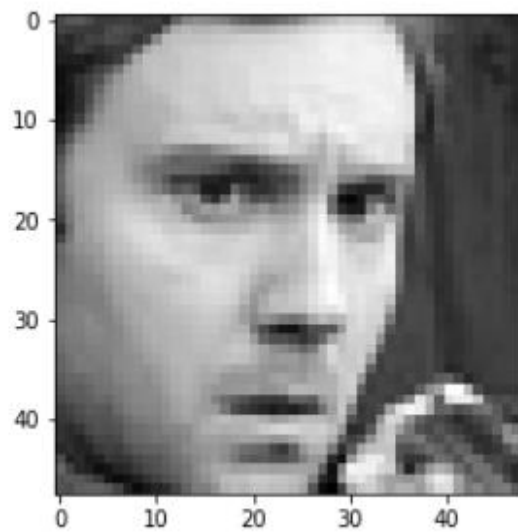






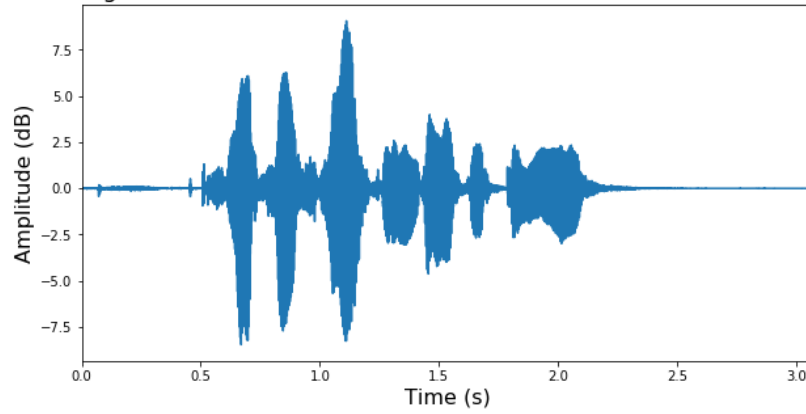
Out[136]: Text(0.5, 1.0, 'Accuracy Curves')





PreProcessing and Training of Audio Analysis

Signal wave of file '03-01-07-01-02-01-22.wav' with label SAD

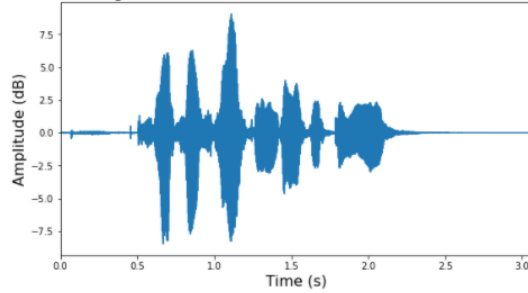


Audio file '03-01-07-01-02-01-22.wav':

Out[4]:



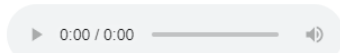
Signal wave of file '03-01-07-01-02-01-22.wav'



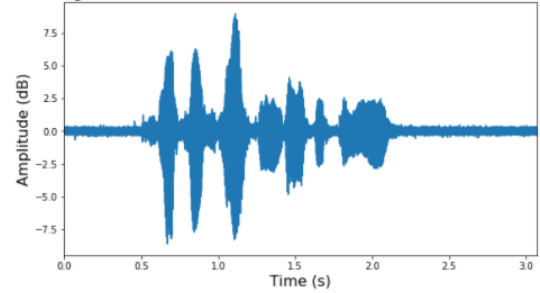
Audio file '03-01-07-01-02-01-22.wav':



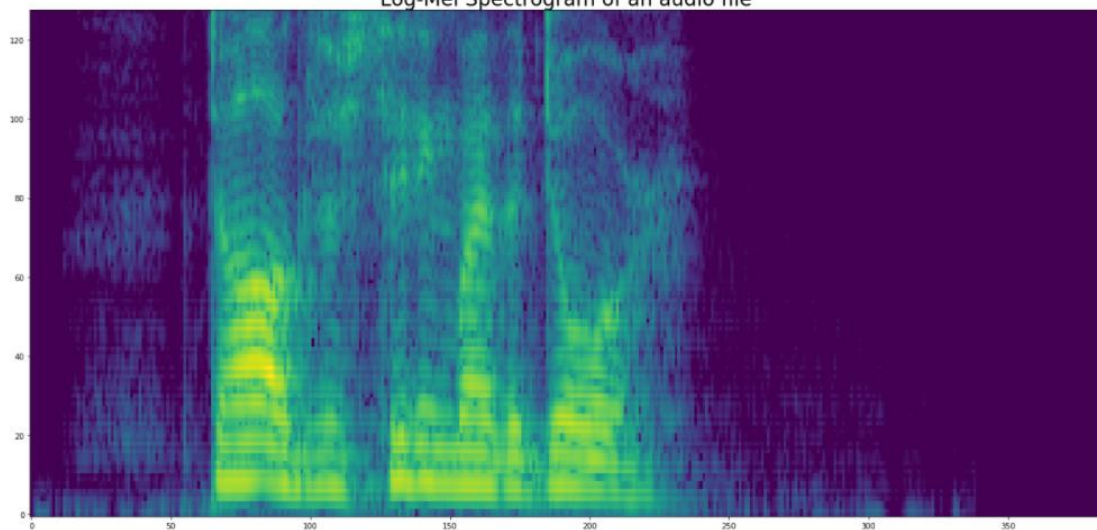
Audio file '03-01-07-01-02-01-22.wav' with noise:

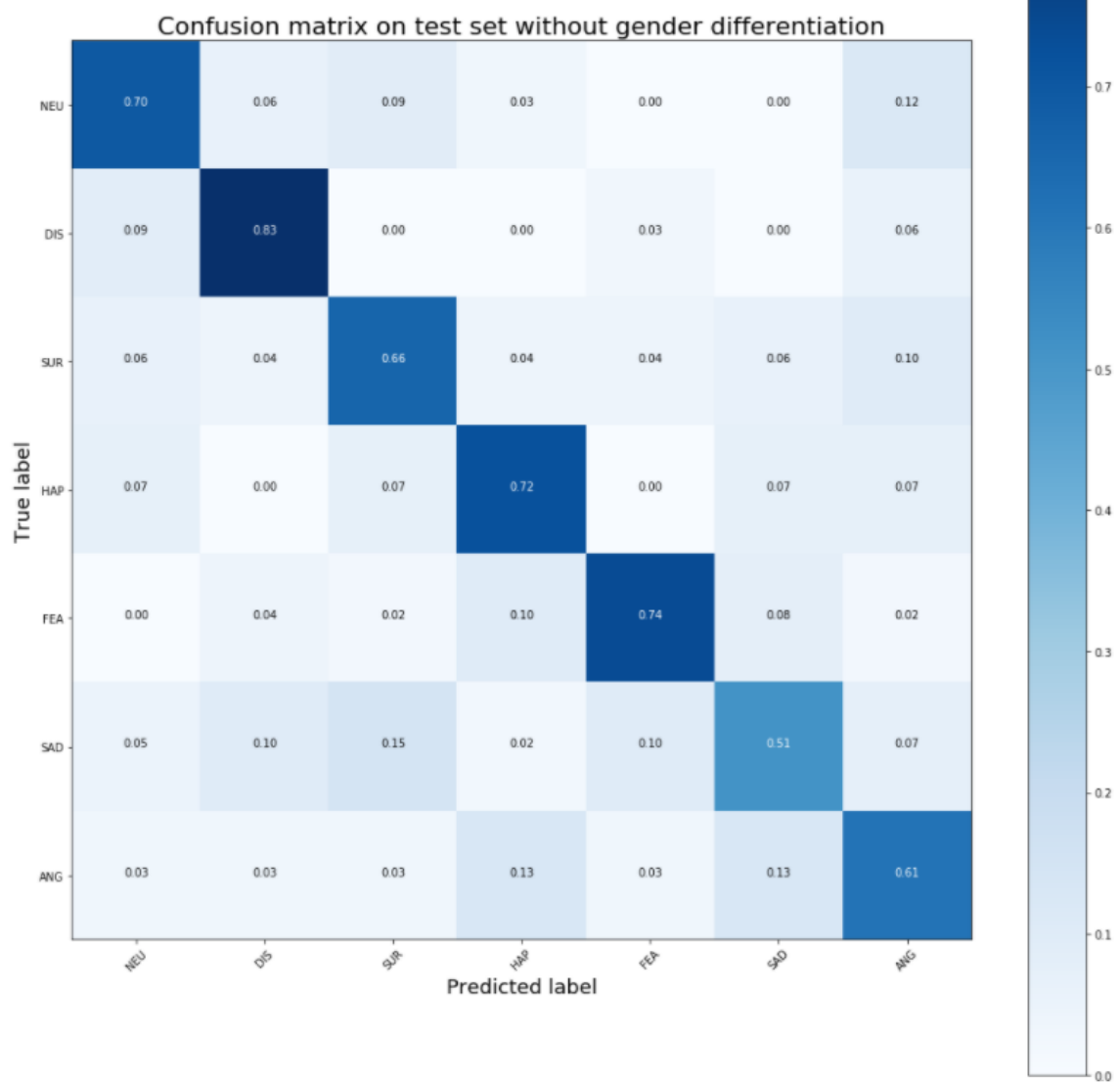


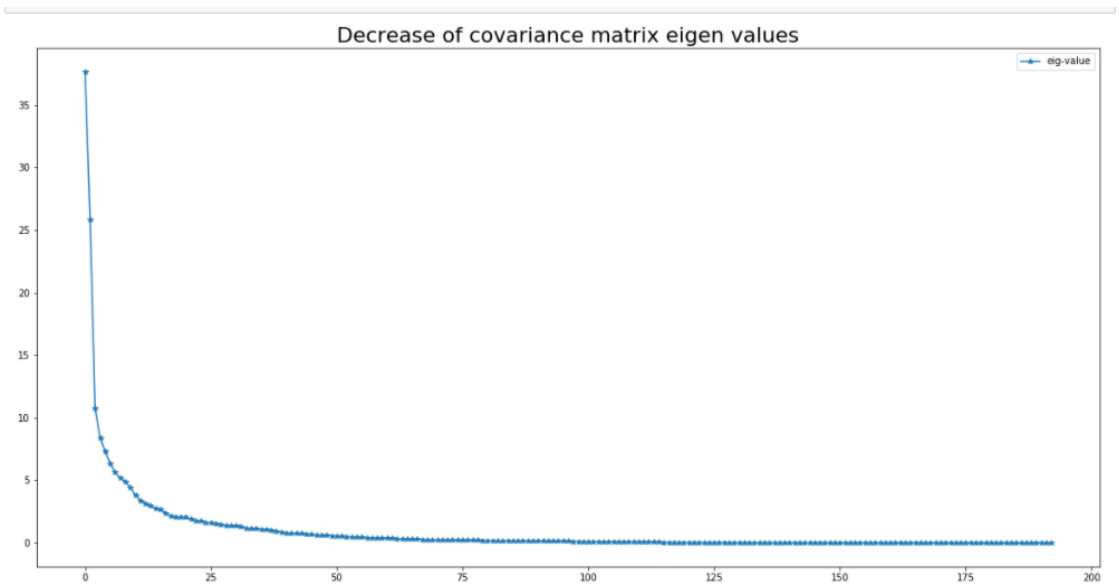
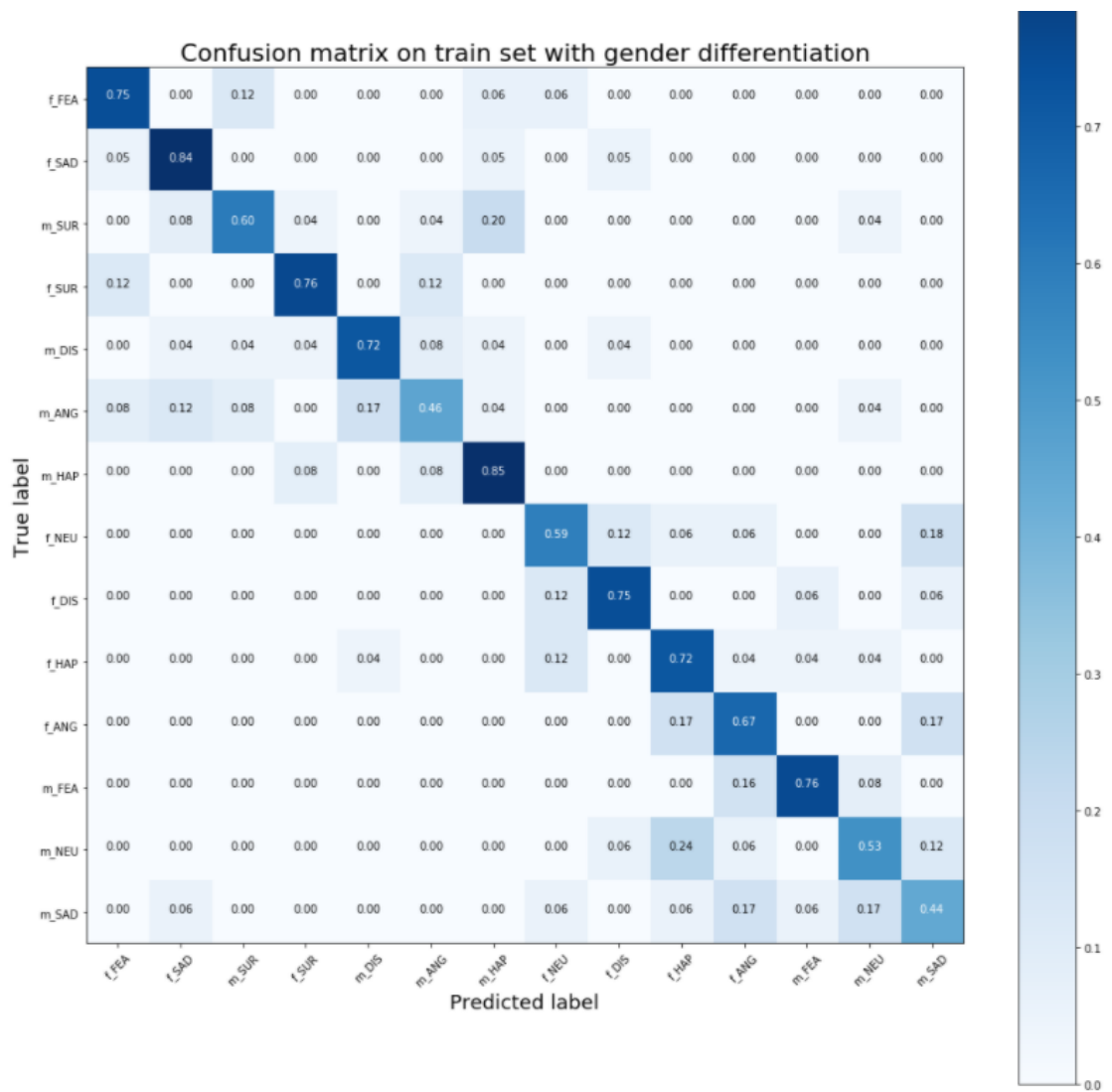
Signal wave of file '03-01-07-01-02-01-22.wav' with Noise

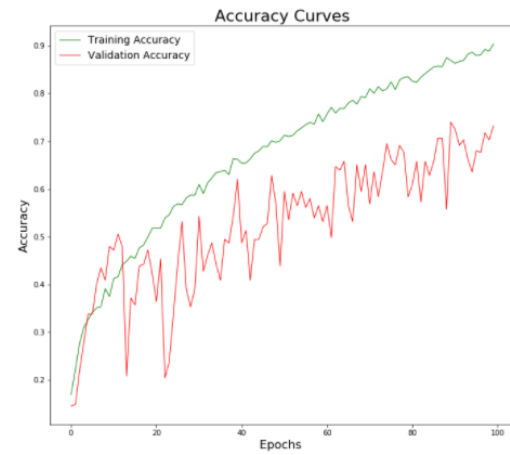
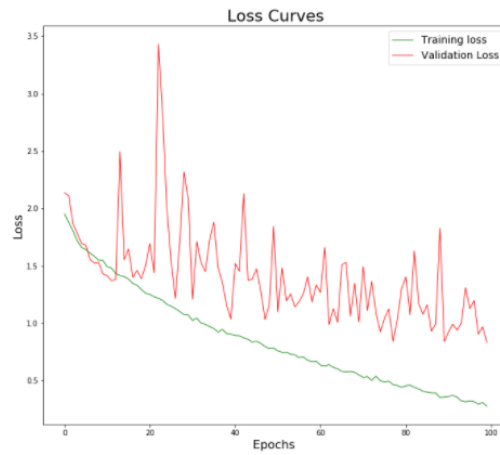


Log-Mel Spectrogram of an audio file



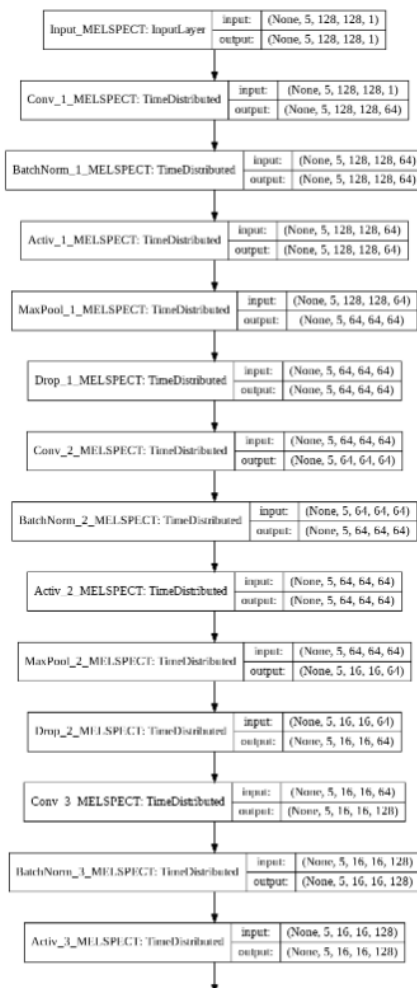


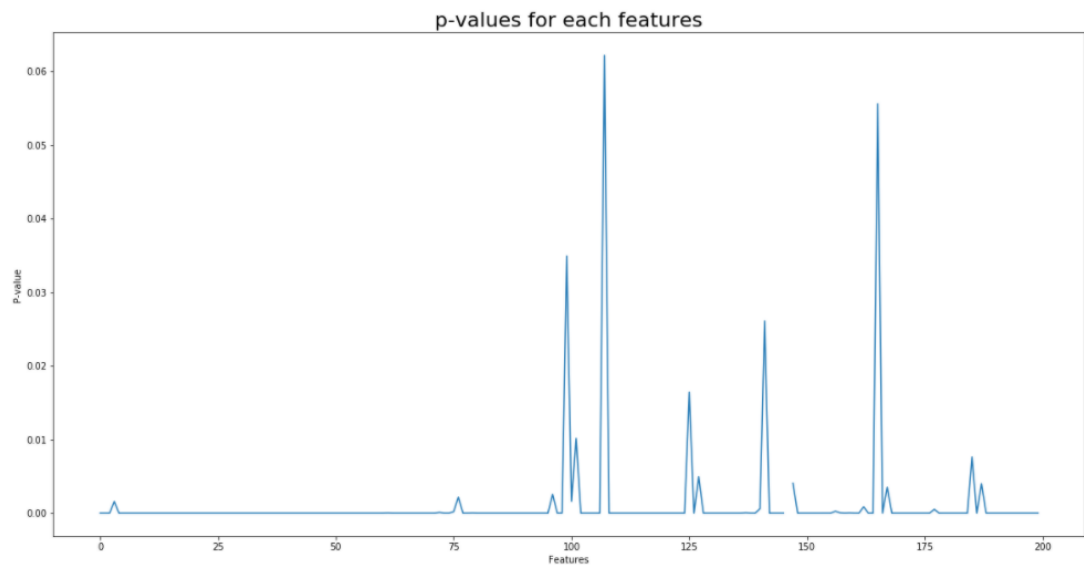
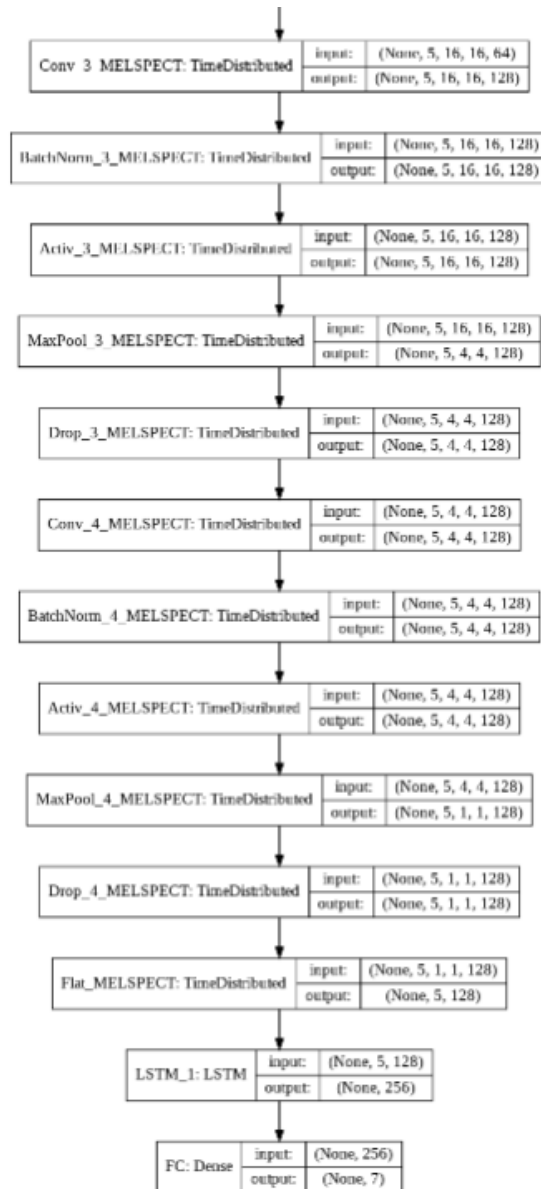




Training of LSTM

Out[6]:

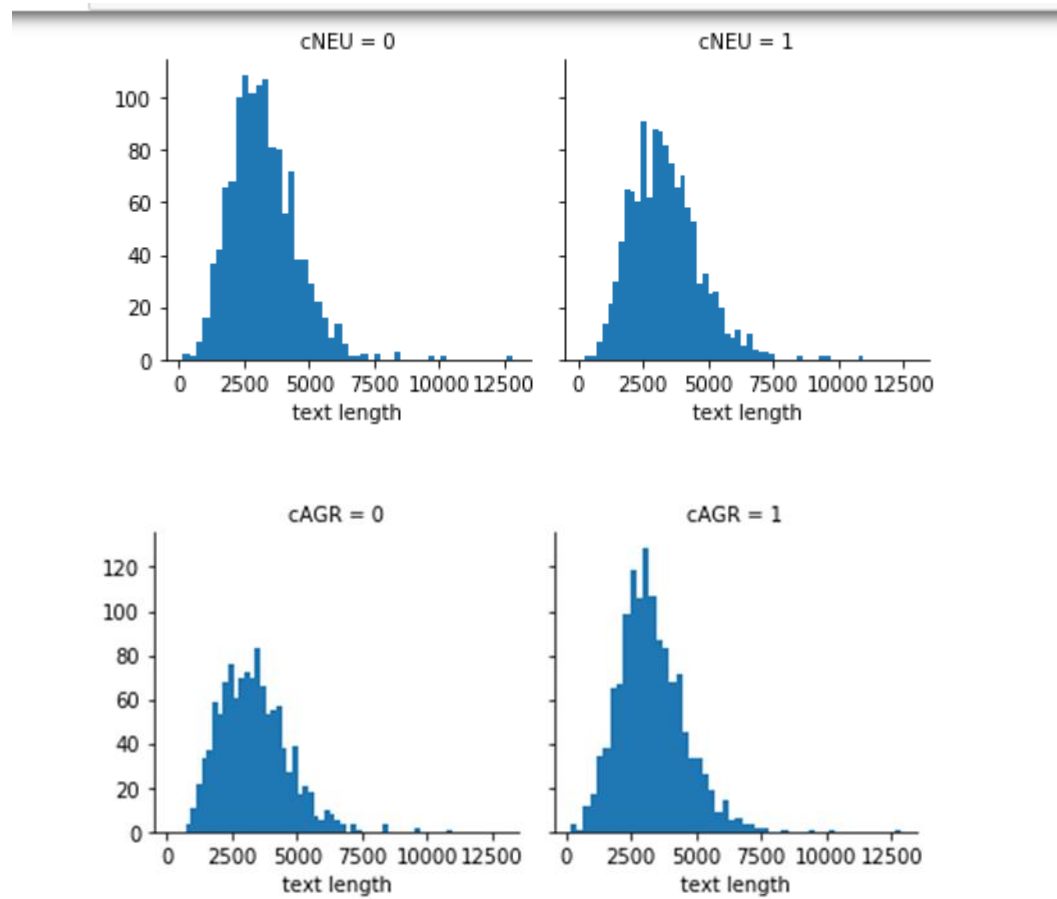
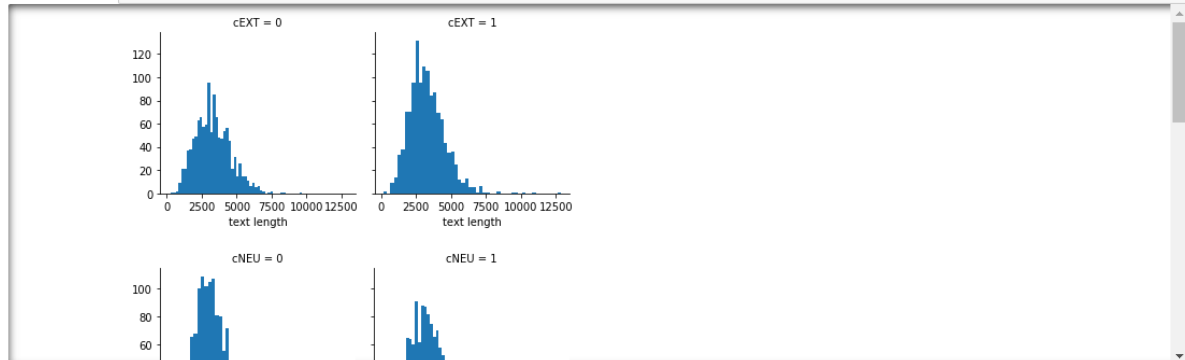




Preprocessing and Training of Text Analysis

Histograms of text length distribution for the different labels

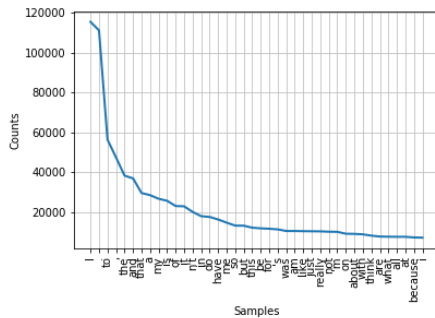
```
In [89]: viz = visualize(data_essays, X_essays, labels)
viz.textlength_vs_labels_histogram()
```



Most frequent words in the corpus

In [91]: viz.most_frequent_words()

```
List of 100 most frequent words/counts
[('I', 115468), ('.', 111178), ('to', 56263), (',', 47355), ('the', 38232), ('and', 36810), ('that', 29456), ('a', 28408), ('m', 26580), ('is', 25576), ('of', 22939), ('it', 22781), ('n't', 19996), ('in', 17828), ('do', 17448), ('have', 16166), ('me', 14588), ('so', 13099), ('but', 13060), ('this', 12054), ('be', 11724), ('for', 11520), ('s', 11198), ('was', 10392), ('am', 10378), ('like', 10308), ('just', 10250), ('really', 10207), ('not', 10015), ('m', 9973), ('on', 9015), ('about', 8941), ('with', 8708), ('think', 8061), ('are', 7602), ('what', 7517), ('all', 7475), ('at', 7469), ('because', 7144), ('i', 7047), ('know', 6959), ('get', 6875), ('he', 6605), ('now', 6154), ('would', 6077), ('you', 6013), ('if', 6001), ('time', 5966), ('out', 5923), ('they', 5905), ('up', 5743), ('or', 5733), ('going', 5621), ('go', 5576), ('she', 5556), ('?', 5539), ('want', 5483), ('will', 5420), ('can', 5276), ('l', 4959), ('as', 4939), ('people', 4898), ('her', 4879), ('when', 4867), ('we', 4836), ('much', 4702), ('it', 4626), ('one', 4465), ('how', 4308), ('feel', 4239), ('there', 4218), ('him', 3988), ('good', 3964), ('here', 3837), ('more', 3810), ('some', 3762), ('had', 3759), ('from', 3684), ('need', 3580), ('been', 3560), ('right', 3428), ('them', 3346), ('did', 3323), ('ca', 3261), ('too', 3213), ('has', 3200), ('could', 3131), ('things', 3117), ('My', 3075), ('well', 3012), ('school', 3010), ('class', 2983), ('wonder', 2959), ('see', 2882), ('should', 2836), ('guess', 2811), ('friends', 2804), ('back', 2684), ('something', 2657), ('very', 2648)]
```



GUI Setup Results :-

```
(multi_mode) C:\Users\welcome\OneDrive\Desktop\Interview_Emulator\Interview_Emulator_Project\04-WebApp>python main.py
2020-12-04 09:08:47.789003: W tensorflow/stream_executor/platform/default/dso_loader.cc:59] Could not load dynamic library 'cudart64_101.dll'; dlderror: cudart64_101.dll not found
2020-12-04 09:08:47.793975: I tensorflow/stream_executor/cuda/cudart_stub.cc:29] Ignore above cudart dlerror if you do not have a GPU set up on your machine.
* Serving Flask app "main" (lazy loading)
* Environment: production
  WARNING: This is a development server. Do not use it in a production deployment.
  Use a production WSGI server instead.
* Debug mode: on
* Restarting with stat
2020-12-04 09:08:54.259680: W tensorflow/stream_executor/platform/default/dso_loader.cc:59] Could not load dynamic library 'cudart64_101.dll'; dlderror: cudart64_101.dll not found
2020-12-04 09:08:54.264205: I tensorflow/stream_executor/cuda/cudart_stub.cc:29] Ignore above cudart dlerror if you do not have a GPU set up on your machine.
* Debugger is active!
* Debugger PIN: 277-967-615
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
```

After Running the Localhost :-

INTERVIEW EMULATOR

Video Interview

Use the video Interview Emulator and get a feedback on how our algorithm interprets your facial emotions compared to other candidates.

You will be provided a feedback on your facial emotions such as :

- Anger
- Happiness
- Fear
- Sadness
- Surprise
- Disgust

[Video Interview](#)

Audio Interview

Use the audio Interview Emulator and get a feedback on how our algorithm interprets your vocal emotions compared to other candidates.

You will be provided a feedback on your vocal emotions such as :

- Anger
- Happiness
- Fear
- Sadness
- Surprise
- Disgust

[Audio Interview](#)

Text Interview

Use the text Interview Emulator and get a feedback on how our algorithm interprets your psychological traits through compared to other candidates.

You will be provided a feedback on your Big Five Psychological traits, which include :

- Openness
- Conscientiousness
- Extraversion
- Agreeableness
- Neuroticism

[Text Interview](#)

By clicking Video interview Button:-

Video Interview

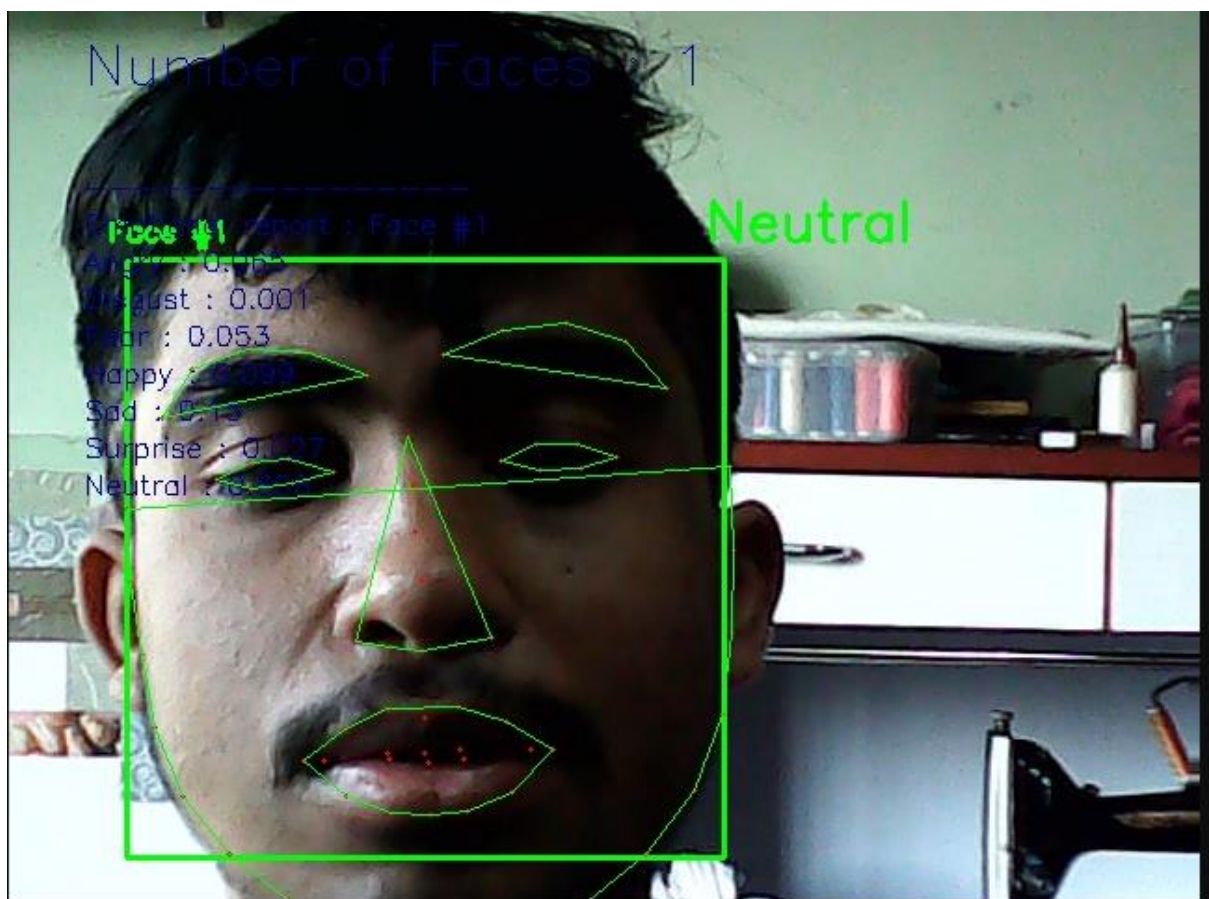
What can you tell me about yourself?
What is your greatest weakness?
Why should we hire you?
Why do you want this job?

Start Recording

You will have 120 seconds to discuss the topic mentioned above. Due to restrictions, we are not able to redirect you once the video is over. Please move your URL to /video_dash instead of /video_1 once over. You will be able to see your results then.

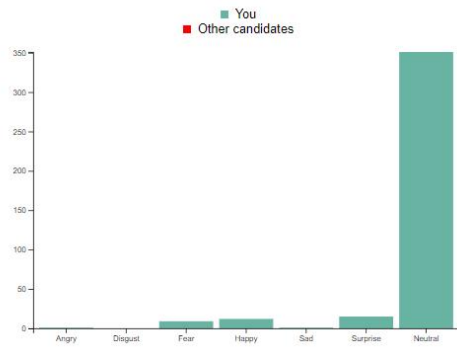
Back

Emotional Report Generated after 120 seconds



Video Interview

Perceived emotions



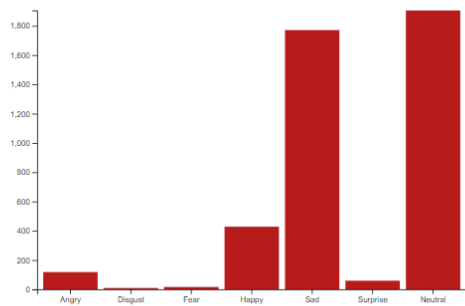
Facial Emotions

Your most frequent emotion is :

Neutral

- Anger : 0%
- Disgust : 0%
- Fear : 2%
- Happiness : 3%
- Sadness : 0%
- Surprise : 3%
- Neutrality : 90%

Other candidates



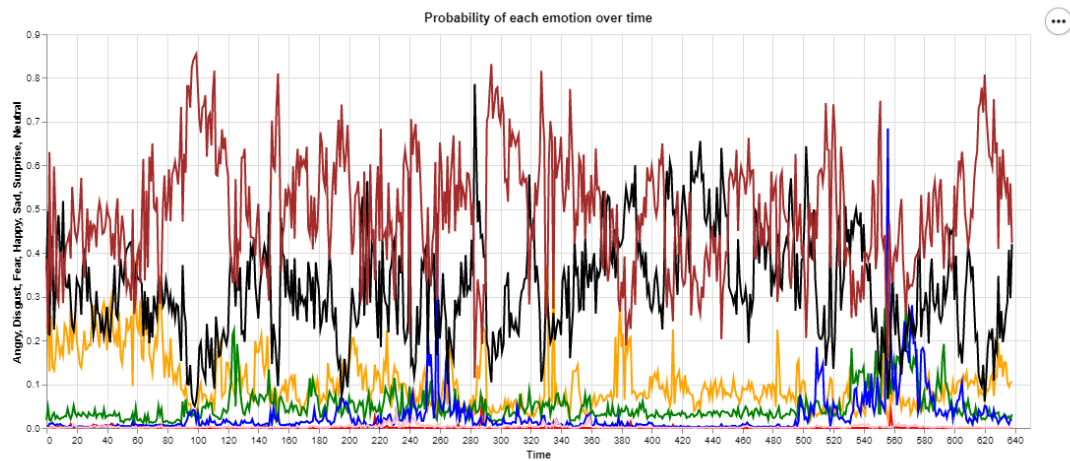
Other candidates most frequent emotion is :

Neutral

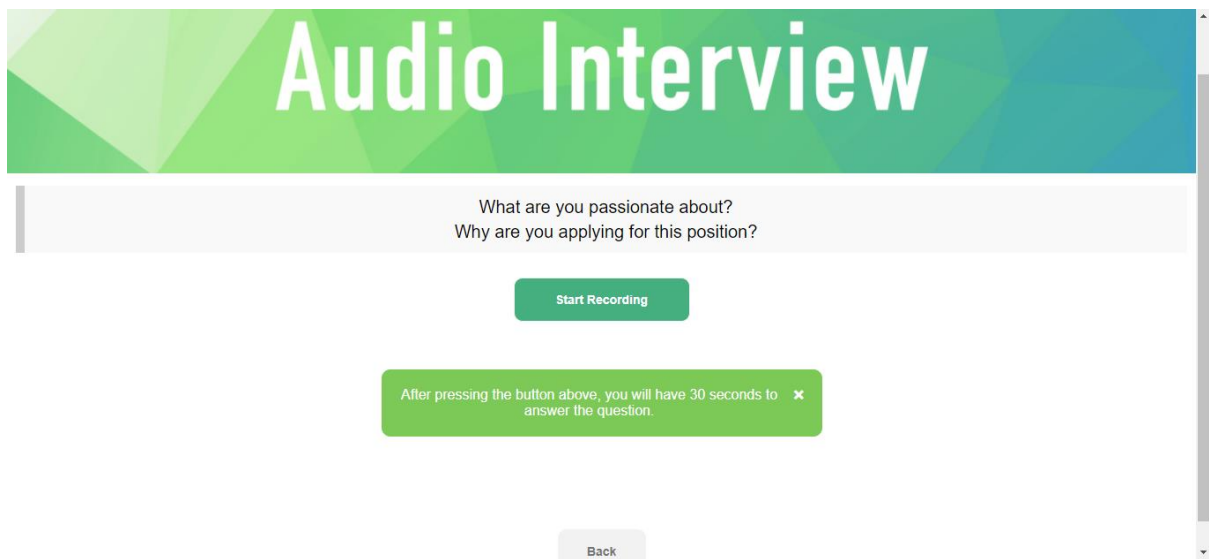
- Angry : 2%
- Disgust : 0%
- Fear : 0%
- Happy : 9%
- Sadness : 41%
- Surprise : 1%
- Neutrality : 44%

Probabilities of emotions over time of Video Emotions through SVG chart/graphs

Over time

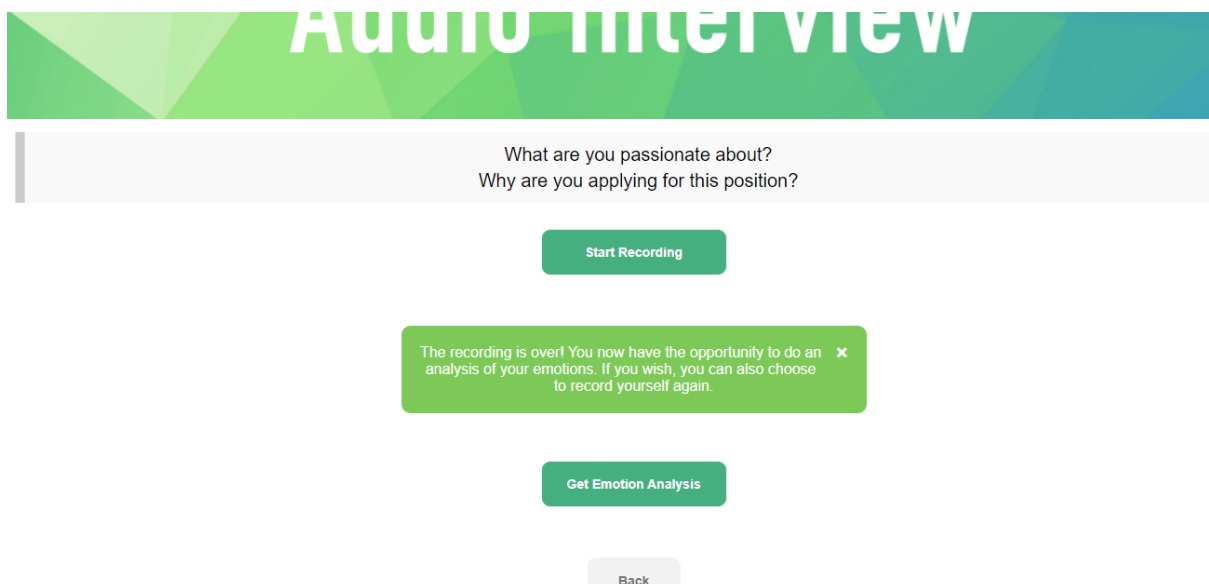


By clicking Audio Button:-



The screenshot shows the 'Audio Interview' interface. At the top is a green and blue geometric header with the text 'Audio Interview'. Below this is a light gray box containing the question: 'What are you passionate about? Why are you applying for this position?'. A green button labeled 'Start Recording' is centered below the question. A green notification box with a close icon (x) contains the text: 'After pressing the button above, you will have 30 seconds to answer the question.' At the bottom center is a small gray button labeled 'Back'.

After starting recording of 30 seconds

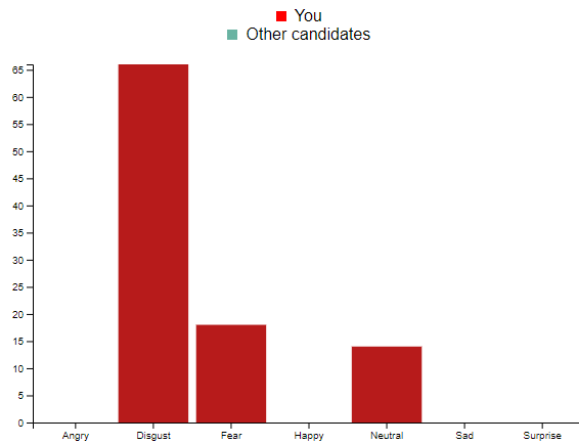


The screenshot shows the 'Audio Interview' interface after 30 seconds of recording. The header and question box remain the same. The 'Start Recording' button is still present. A green notification box with a close icon (x) contains the text: 'The recording is over! You now have the opportunity to do an analysis of your emotions. If you wish, you can also choose to record yourself again.' Below this, a green button labeled 'Get Emotion Analysis' is centered. At the bottom center is a small gray button labeled 'Back'.

By clicking Emotion Analysis Button:-

Dashboard of Audio Analysis

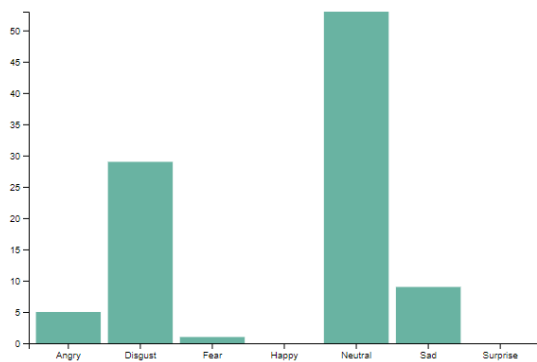
Perceived emotions



Your most frequent emotion is : **Disgust**

- Angry : 0%
- Disgust : 66%
- Fear : 18%
- Happy : 0%
- Neutral : 14%
- Sad : 0%
- Surprise : 0%

Other candidates



Other candidates most frequent emotion is : **Neutral**

- Angry : 5%
- Disgust : 29%
- Fear : 1%
- Happy : 0%
- Neutral : 53%
- Sad : 9%
- Surprise : 0%

Back

Temporary files which will be stored in tmp files

Name	Status	#	Title	Contributing
 t				
 voice_recording				

By clicking Text Interview Button

Text Interview

What questions do you have for me?

Or upload your Cover Letter :

Start Analysis

Choose File | No file chosen

Start Analysis

By Asking Questions to HR:-

What questions do you have for me?

Or upload your Cover Letter :

How has the position changed over time?
what does my potential team role's hierarchy look like?
what are the main things HR does to assist the department i would be working for?

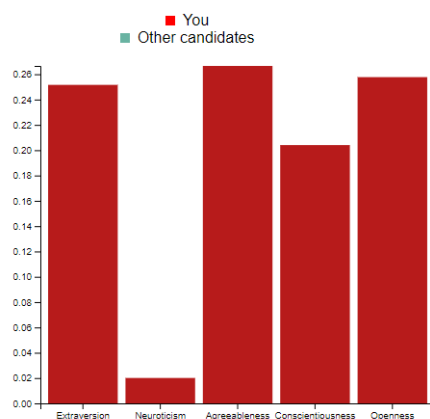
Start Analysis

Choose File | No file chosen

Start Analysis

Emotional Analysis Report :-

Perceived Psychological Traits



Your most visible trait is :

Agreeableness

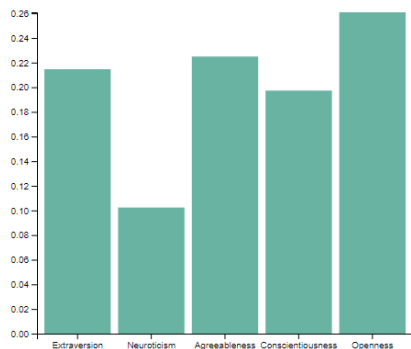
Psychological Traits :

- Extraversion : 25%
- Neuroticism : 2%
- Agreeableness : 26%
- Conscientiousness : 20%
- Openness : 25%

Most common words :

- position
- change
- time
- potential
- team
- role
- hierarchy
- look
- like
- main
- thing
- hr
- assist
- department
- would

Other candidates



Their most visible trait is :

Openness

Psychological Traits :

- Extraversion : 21%
- Neuroticism : 10%
- Agreeableness : 22%
- Conscientiousness : 19%
- Openness : 26%

Most common words :

- 0
- network
- data
- use
- emotion
- work
- system
- 1
- recognition
- computer
- c
- social
- 00
- analysis
- 7

Back

We can upload files of Resume or CV first we have to setup Tika Server :-

```
C:\Users\welcome\Downloads>java -jar tika-server-1.24.1.jar
Dec 04, 2020 9:09:22 AM org.apache.tika.config.InitializableProblemHandler$3 handleInitializableProblem
WARNING: J2KImageReader not loaded. JPEG2000 files will not be processed.
See https://pdfbox.apache.org/2.0/dependencies.html#jai-image-io
for optional dependencies.

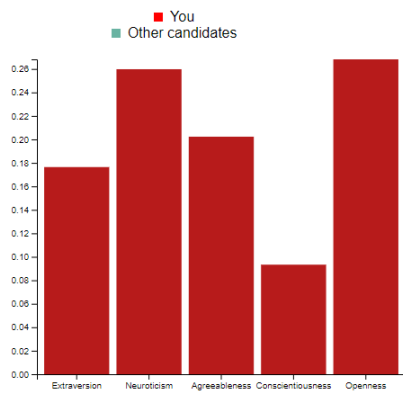
Dec 04, 2020 9:09:22 AM org.apache.tika.config.InitializableProblemHandler$3 handleInitializableProblem
WARNING: org.xerial's sqlite-jdbc is not loaded.
Please provide the jar on your classpath to parse sqlite files.
See tika-parsers/pom.xml for the correct version.
INFO Starting Apache Tika 1.24.1 server
INFO Setting the server's publish address to be http://localhost:9998/
INFO Logging initialized @985ms to org.eclipse.jetty.util.log.Slf4jLog
INFO jetty-9.4.27.v20200227; built: 2020-02-27T18:37:21.340Z; git: a304fd9f351f337e7c0e2a7c28878dd536149c6c; jvm 15+36-1562
INFO Started ServerConnector@24faea88{HTTP/1.1, (http/1.1)}{localhost:9998}
INFO Started @1095ms
WARN Empty contextPath
INFO Started o.e.j.s.h.ContextHandler@22c86919{/,null,AVAILABLE}
INFO Started Apache Tika server at http://localhost:9998/
INFO rmeta/text (autodetecting type)
INFO rmeta/text (autodetecting type)
```

Or upload your Cover Letter :

Choose File Jayant_Resume.pdf

Start Analysis

Perceived Psychological Traits



Your most visible trait is :

Openness

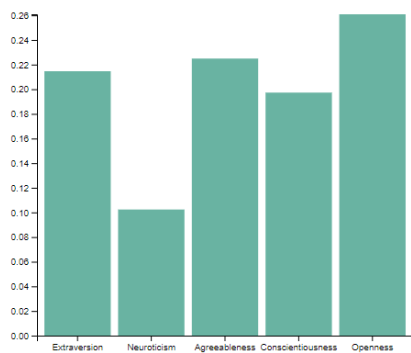
Psychological Traits :

- Extraversion : 17%
- Neuroticism : 25%
- Agreeableness : 20%
- Conscientiousness : 9%
- Openness : 26%

Most common words :

- c
- college
- ghosh
- jayanta
- work
- badal
- mr
- name
- got
- certification
- 80
- 78
- english
- computer
- thakurli

Other candidates



Their most visible trait is :

Openness

Psychological Traits :

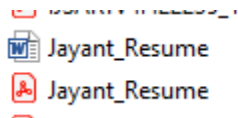
- Extraversion : 21%
- Neuroticism : 10%
- Agreeableness : 22%
- Conscientiousness : 19%
- Openness : 26%

Most common words :

- 0
- network
- data
- use
- emotion
- work
- system
- 1
- recognition
- computer
- c
- social
- 00
- analysis
- 7

Back

Temporary resume files which will be stored in WebApp folder



29-11-2020 08:35
04-12-2020 09:25

Conclusion:-

Finally concluding the idea extract, by As, in the prevailing and ever growing world, where the technology is booming in each and every field this prototype would be used for recruiting the candidates from all over the world using a single tool. Also, here depicting the use cases of this future Prototype. The tool can be used as an alternative (as an assistive tool not a replacement of the manual interview process) of manual interview in the unprecedented times, so that, the company can be safeguard the health and safety of both their employees and the interviewee candidate. The tool can also be used to train the candidates in order to outperform their budding competitive opposition candidates.

Further Research:-

Interview Emulator is a relatively new discipline that aims to include text inputs, as well as sound and video. This field has been rising with the development of social network that gave researchers access to a vast amount of data. The project consists of emotion recognition using audio video and textual base. This platform helps to analyze the job candidates in an effective way and enables corporates to choose a good candidate among the appearing ones. This project can be used in diversified fields to assess the candidate. It can be used to automate the entire hiring process in near future.

Bibliographies

- ❖ <https://uu.diva-portal.org/smash/get/diva2:952138/FULLTEXT01.pdf>
- ❖ <https://arxiv.org/abs/1205.4944>
- ❖ <https://www.sciencedirect.com/science/article/pii/S235291482030201X>
- ❖ **pyAudioAnalysis: An Open-Source Python Library for Audio Signal Analysis**
- ❖ Mohit Shrivastava and Anupam Agarwal. Classification of emotions from speech using implicit features. In 9th International Conference on Industrial and Information Systems, 1-6, Dec. 2014
- ❖ Personality Trait Classification of Essays with the Application of Feature Reduction
Edward P. Tighe, Jennifer C. Ureta, Bernard Andrei L. Pollo, Charibeth K. Cheng, and Remedios de Dios Bulos De La Salle Univeristy, Manila, Philippine
- ❖ https://www.tensorflow.org/api_docs
- ❖ <https://faroit.com/keras-docs/1.2.0/>
- ❖ <https://flask.palletsprojects.com/en/1.1.x/>

