MAHARAJA INSTITUTE OF TECHNOLOGY THANDAVAPURA
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

ANALYSING SPEECH DATA TO DETERMINE EMOTIONS

Presented by,

SUDHARSHAN V [4MN16CS040]
PUNITH NAYAKA K [4MN16CS032]
VARSHITHA S [4MN16CS048]
VINUTHA B [4MN16CS050]

Under the guidance of

Dr Mallikarjunaswamy N J
Assistant Professor,
Department of CSE,
MITT

Contents

- Abstract
- Introduction
- Literature Survey
- Methodology
- Existing system
- Proposed system
- Non functional Requirements

Continue…

- Training model
- Use case model
- Results
- Hardware and Software Requirements
- Conclusion
- Reference

Abstract

- The human voice is very versatile and carries a multitude of emotions. Emotion in speech carries extra insight about human actions.
- Through further analysis, we can better understand the motives of people, humans are easily able to determine the emotion of a speaker, but the field of emotion recognition through machine learning.
- The study of emotion in speech by detecting one emotion.

Introduction

The main objectives are following:

- The ability to understand vocal sounds.
- The human voice can be characterized by several attributes such as pitch, loudness, and vocal tone.
- humans express their emotions by varying different vocal attributes during speech generation.
- The human voice frequency is specifically a part of the human sound production mechanism.
- In which the vocal cords or folds are the primary source of generated sounds.

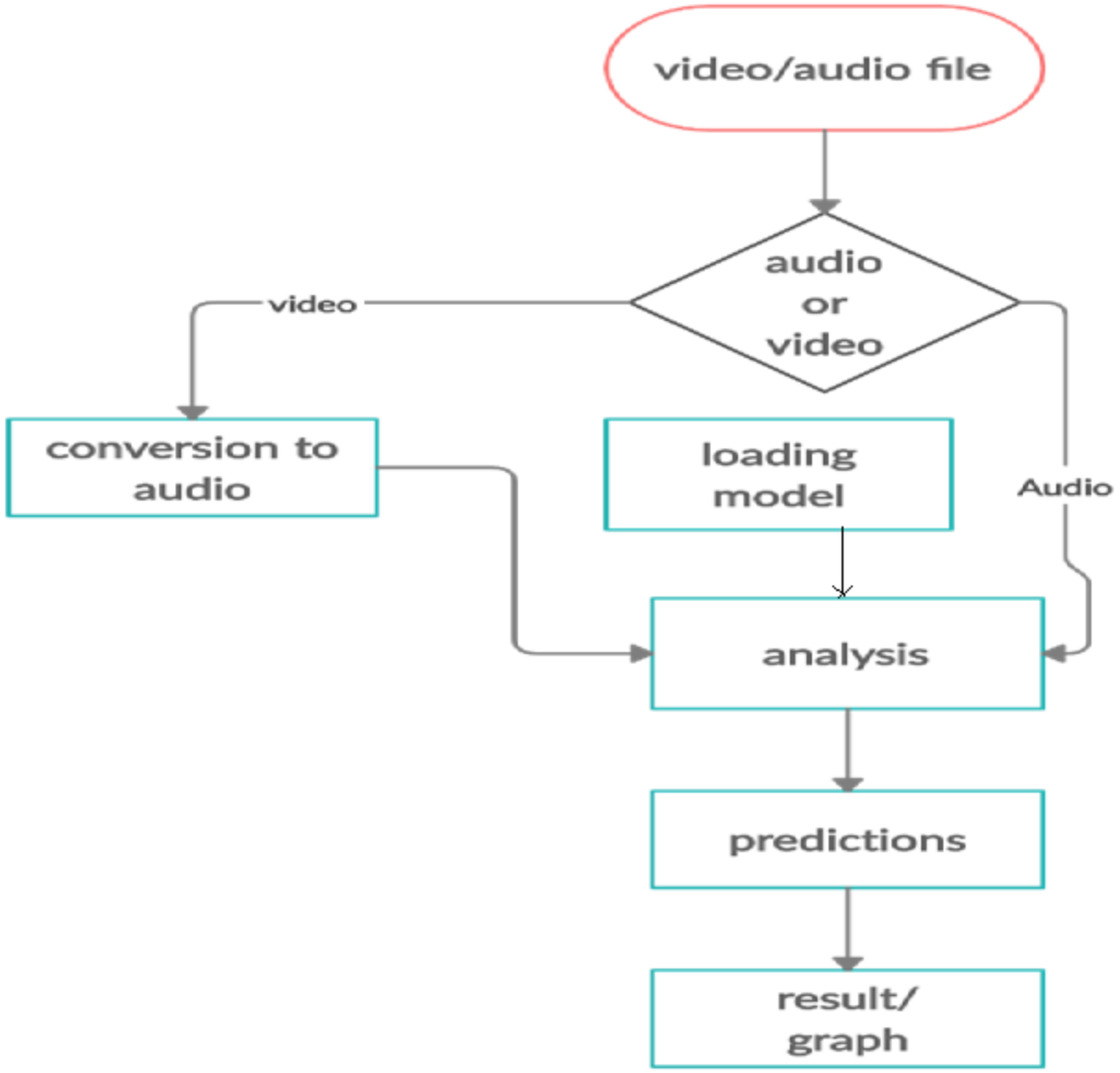
Literature Survey

Year	Author	Title	Method	Dataset	Results	Drawbacks
2018	Maisy Wieman, Andy Sun	Analyzing Vocal Patterns to Determine Emotion	We investigate the best algorithms to select features that are relevant to predicting emotion.	Stanford University Linguistics Department has an Emotional Prosody and Speech Corpus.	Predicted with relative accuracy. (70%).	Low accuracy, collection of data set is a tidiest task. Training takes a lot of time.
2018	Mithesh chouhan	Speech-Emotion-Analyzer,	Neural networks	Berlin voice dataset	Detect emotions with more than 70% accuracy.	Building the model was a challenging task as it involved lot of trail and error methods, tuning etc.
2007	Kamran Soltani Raja Noor Ainon	Speech emotion detection based on neural	Back propagation neural network as classifier.	Berlin Database	60% accuracy	Only 2 emotions were able to detect.

Continue…

2005	Chul Min Lee, Student Member, IEEE, and Shrikanth S. Narayanan, Senior Member, IEEE	Toward Detecting Emotions in Spoken Dialogs	Linear discriminant classifiers (LDC) with Gaussian class-conditional probability and k-nearest neighborhood classifiers (k-NN).	Real time dataset	classifier yielded promising results with 70% classification accuracy	Very time consuming
2011	Jangsik Cho ,Shohei Kato	Detecting emotion from voice	Bayesian pairwise classifiers	Stand ford voice dataset	Classifier with 60% prediction.	Taking probability into consideration the ratio of correctness mainly depends on datasets.

Methodology



Existing System

- The system that is currently present is google speech which also lacks in understanding the words when pronounced.
- It is time consuming.

Proposed System

- The system that is being proposed is to identify the modulation of the word in the speech data by understanding the tone in the voice that is being used.
- This helps in understanding whether the person is angry, happy, sad, neutral, calm, fearful, disgust, surprise etc.

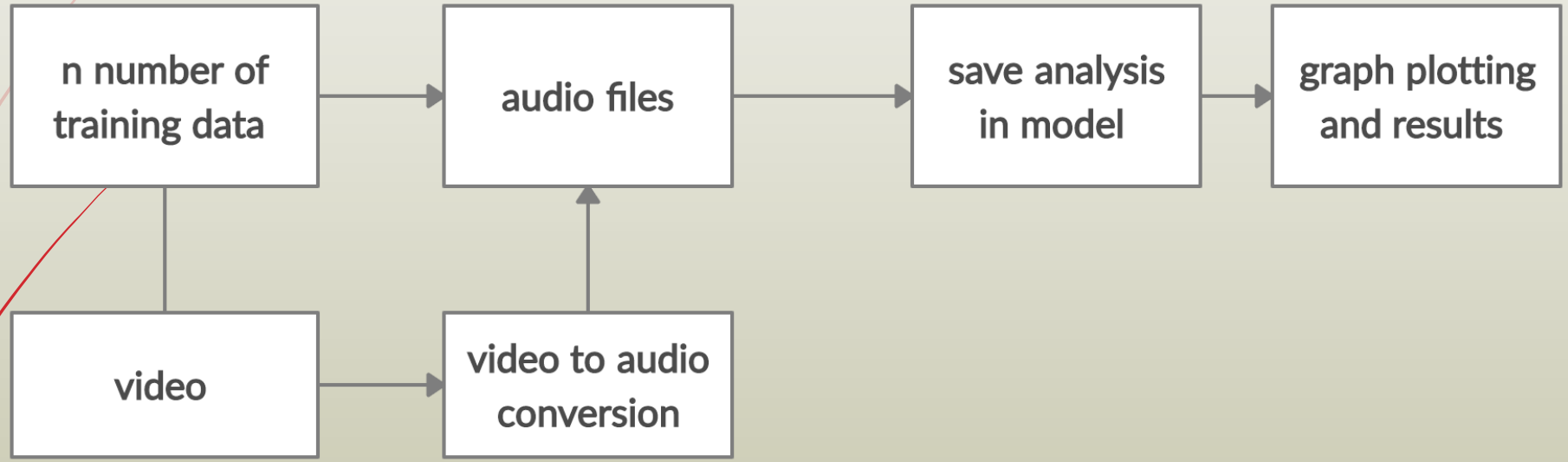
Functional Requirements

- Video input
- Audio input
- Predicted to emotion output

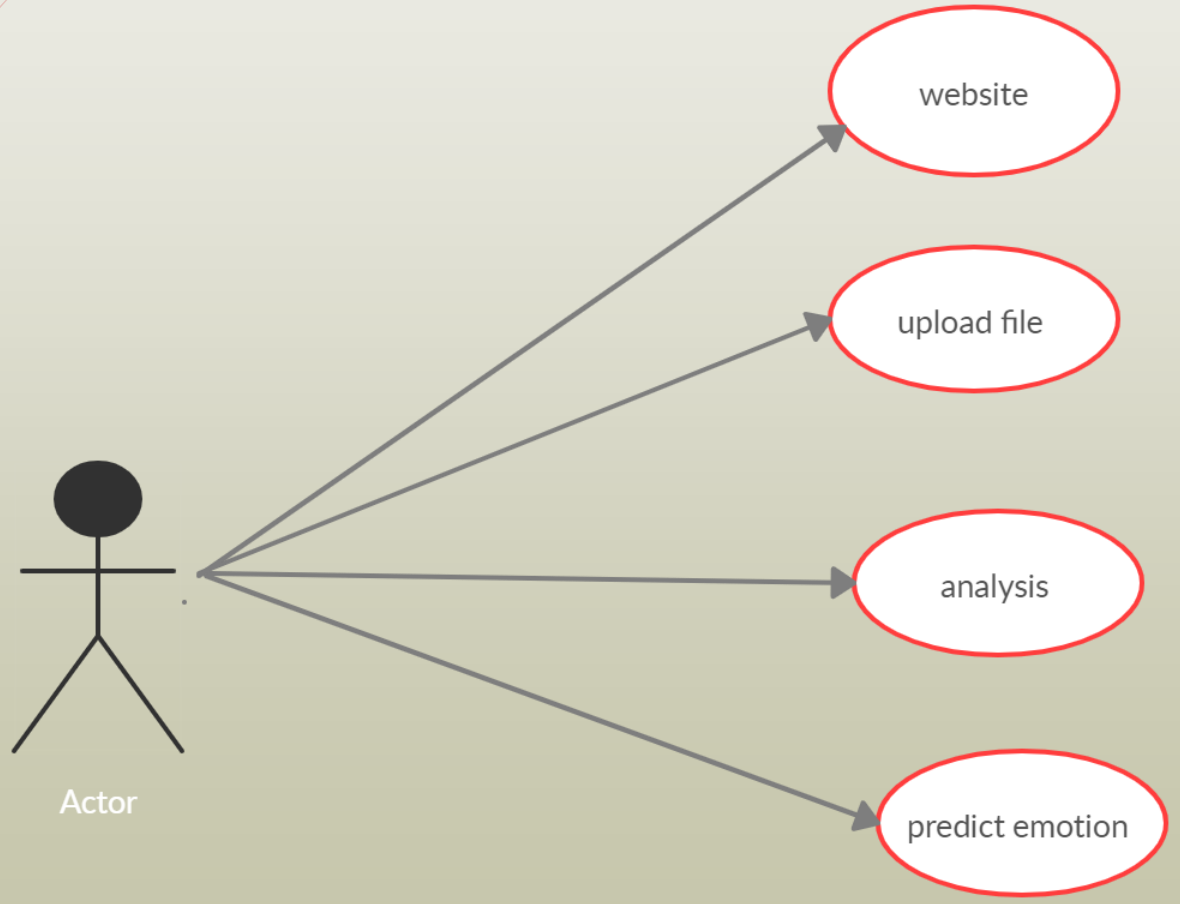
Non Functional Requirements

- Accessibility
- Documentation
- Maintainability
- Portability
- Reliability

Training model process

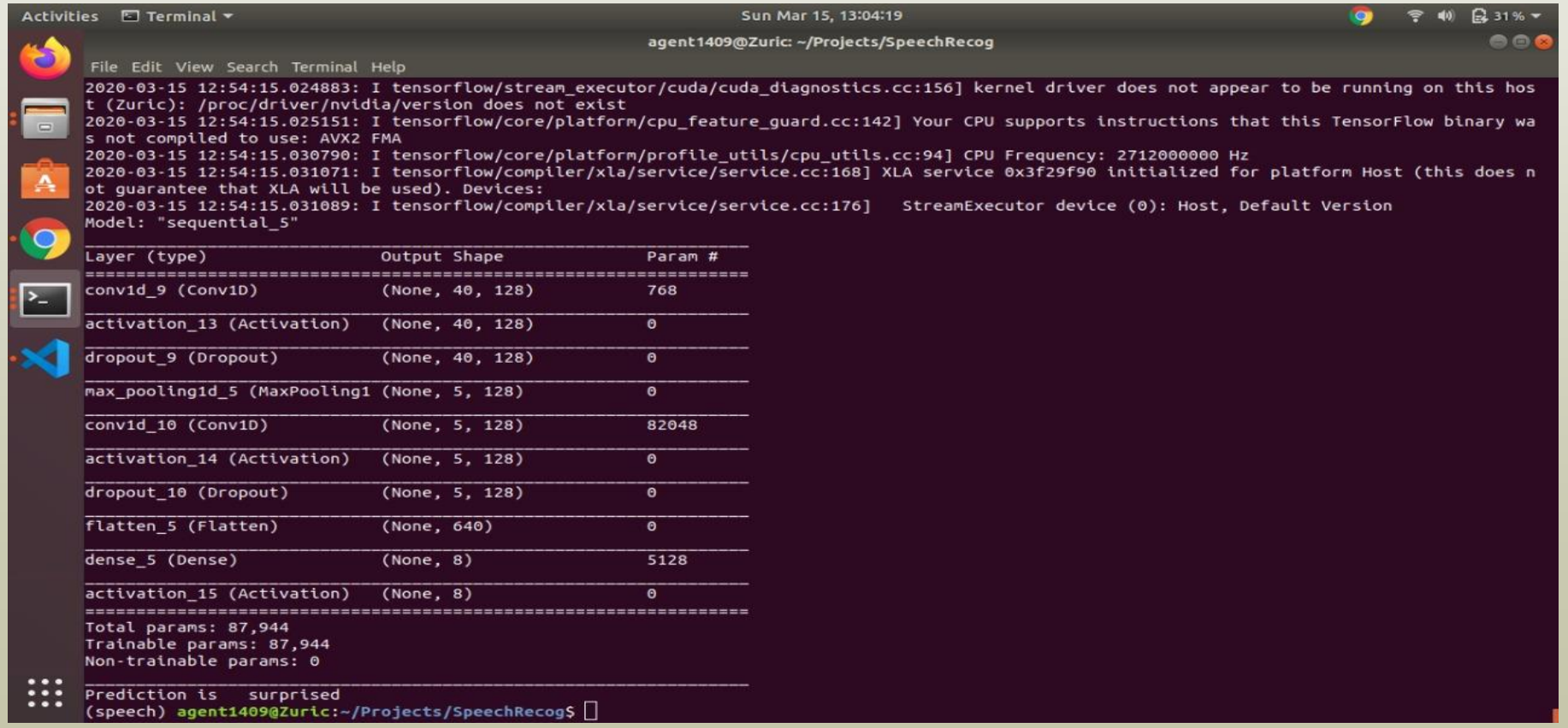


Use Case Model

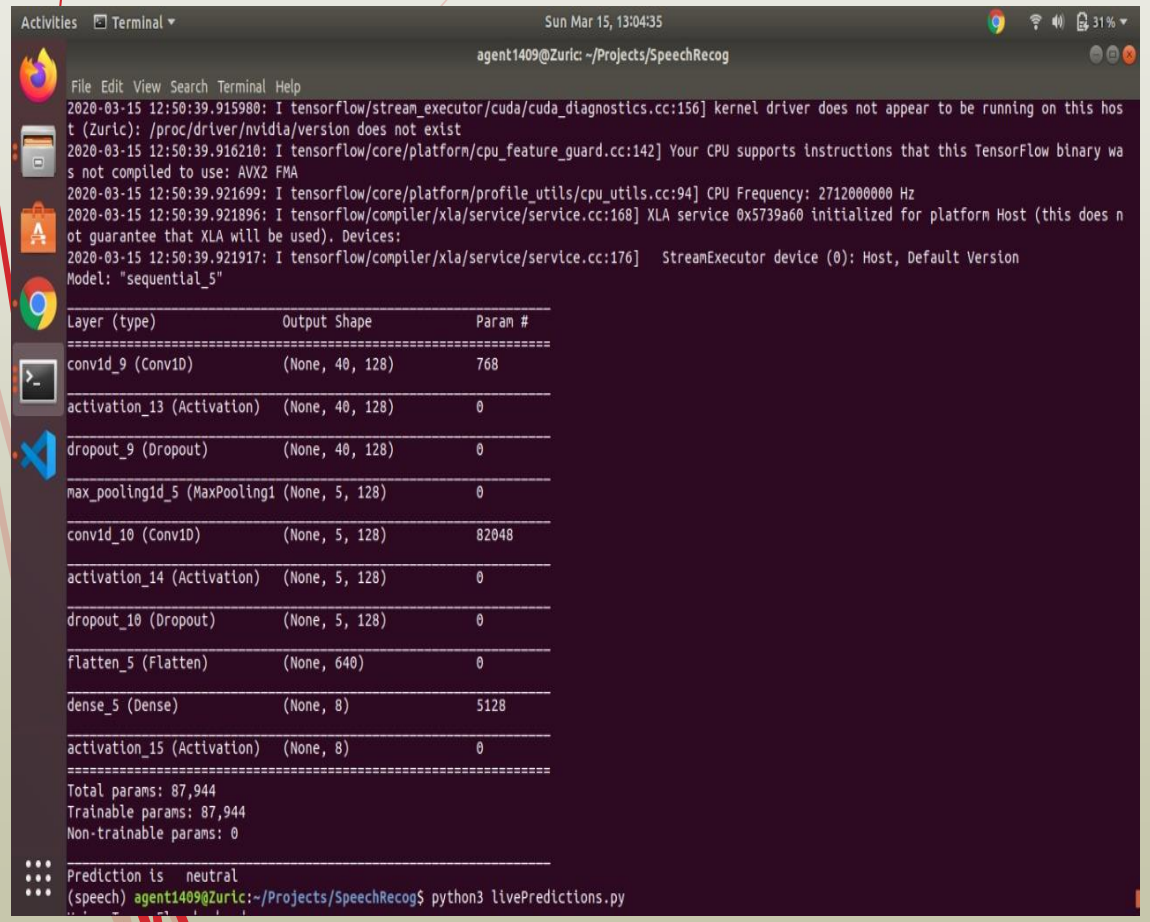


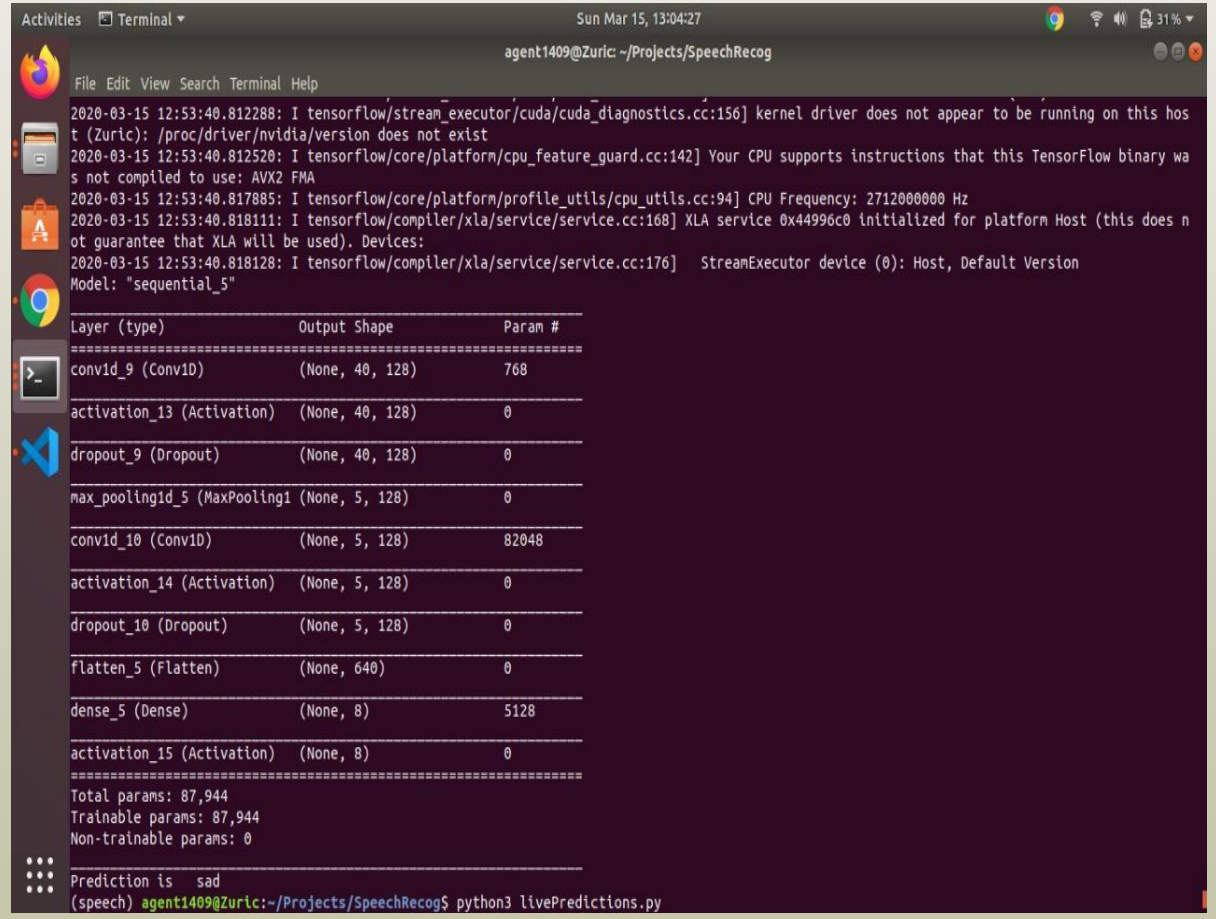
Results

- Snapshots

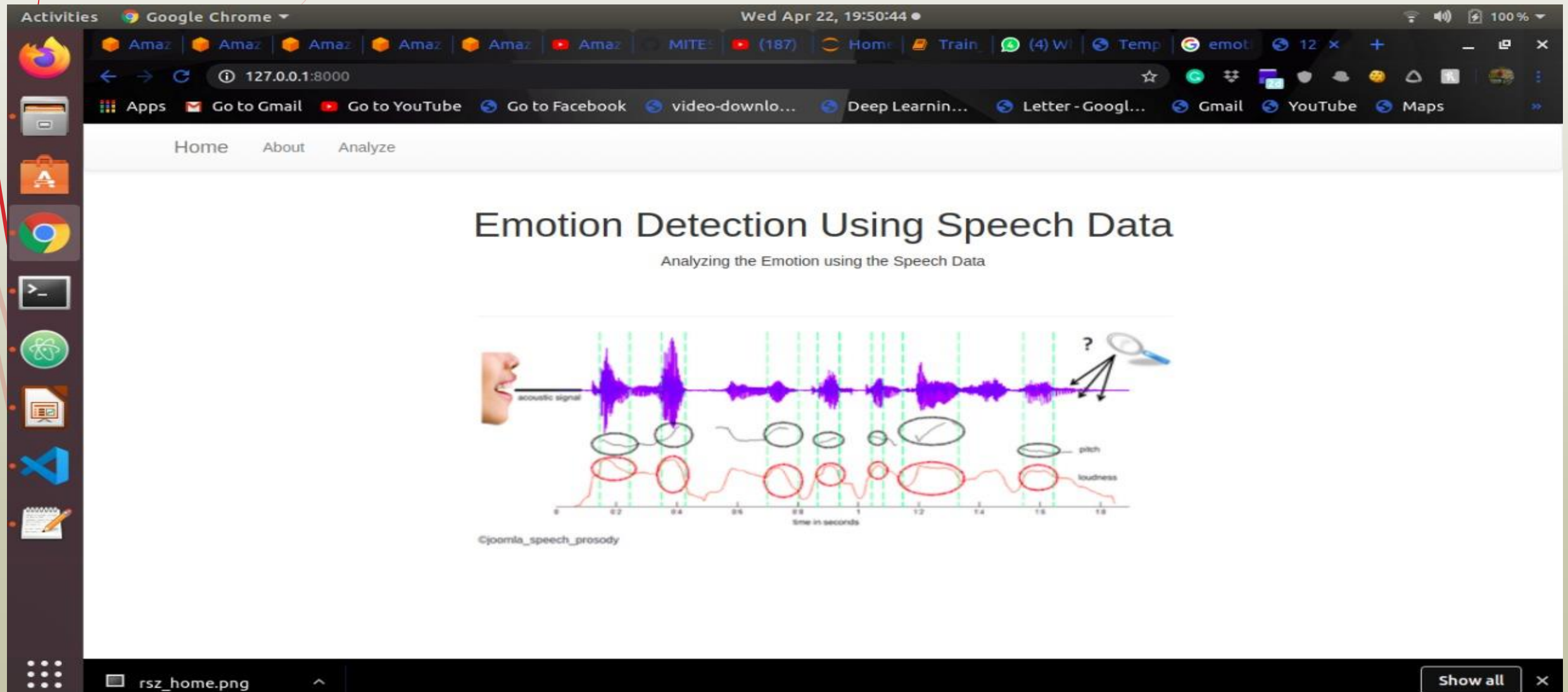


Continue…

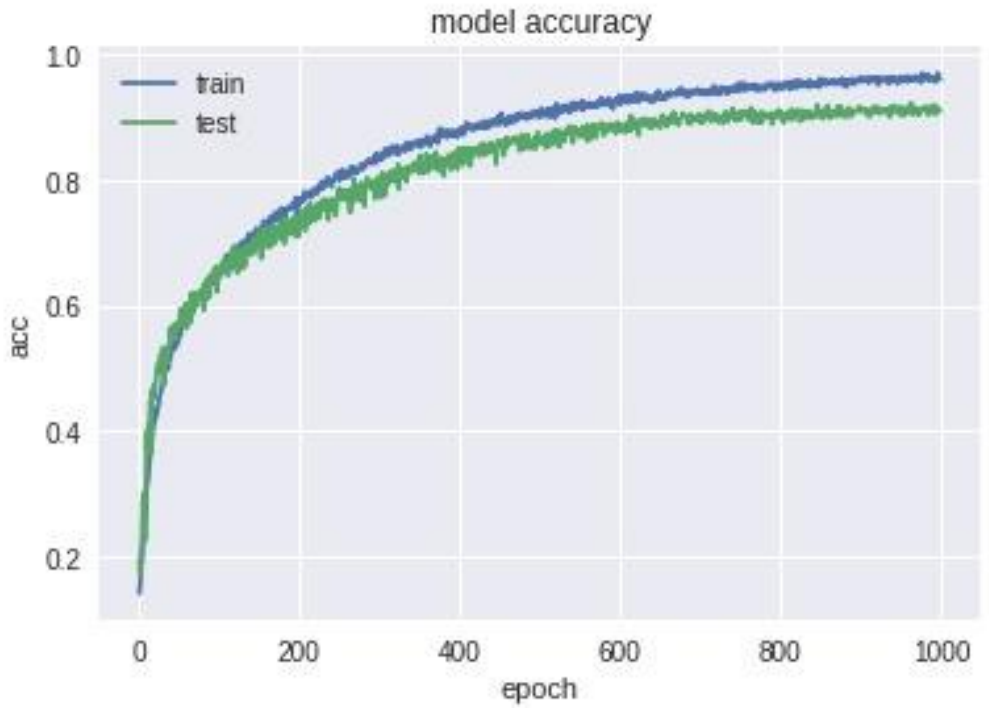

Activities Terminal Sun Mar 15, 13:04:35 31%
agent1409@Zuric: ~/Projects/SpeechRecog
File Edit View Search Terminal Help
2020-03-15 12:50:39.915980: I tensorflow/stream_executor/cuda/cuda_diagnostics.cc:156] kernel driver does not appear to be running on this host (Zuric): /proc/driver/nvidia/version does not exist
2020-03-15 12:50:39.916210: I tensorflow/core/platform/cpu_feature_guard.cc:142] Your CPU supports instructions that this TensorFlow binary was not compiled to use: AVX2 FMA
2020-03-15 12:50:39.921699: I tensorflow/core/platform/profile_utils/cpu_utils.cc:94] CPU Frequency: 2712000000 Hz
2020-03-15 12:50:39.921896: I tensorflow/compiler/xla/service/service.cc:168] XLA service 0x5739a60 initialized for platform Host (this does not guarantee that XLA will be used). Devices:
2020-03-15 12:50:39.921917: I tensorflow/compiler/xla/service/service.cc:176] StreamExecutor device (0): Host, Default Version
Model: "sequential_5"
Layer (type) Output Shape Param #
conv1d_9 (Conv1D) (None, 40, 128) 768
activation_13 (Activation) (None, 40, 128) 0
dropout_9 (Dropout) (None, 40, 128) 0
max_pooling1d_5 (MaxPooling1 (None, 5, 128) 0
conv1d_10 (Conv1D) (None, 5, 128) 82048
activation_14 (Activation) (None, 5, 128) 0
dropout_10 (Dropout) (None, 5, 128) 0
flatten_5 (Flatten) (None, 640) 0
dense_5 (Dense) (None, 8) 5128
activation_15 (Activation) (None, 8) 0
Total params: 87,944
Trainable params: 87,944
Non-trainable params: 0
Prediction is neutral
(speech) agent1409@Zuric:~/Projects/SpeechRecog$ python3 livePredictions.py


Activities Terminal Sun Mar 15, 13:04:27 31%
agent1409@Zuric: ~/Projects/SpeechRecog
File Edit View Search Terminal Help
2020-03-15 12:53:40.812288: I tensorflow/stream_executor/cuda/cuda_diagnostics.cc:156] kernel driver does not appear to be running on this host (Zuric): /proc/driver/nvidia/version does not exist
2020-03-15 12:53:40.812520: I tensorflow/core/platform/cpu_feature_guard.cc:142] Your CPU supports instructions that this TensorFlow binary was not compiled to use: AVX2 FMA
2020-03-15 12:53:40.817885: I tensorflow/core/platform/profile_utils/cpu_utils.cc:94] CPU Frequency: 2712000000 Hz
2020-03-15 12:53:40.818111: I tensorflow/compiler/xla/service/service.cc:168] XLA service 0x44996c0 initialized for platform Host (this does not guarantee that XLA will be used). Devices:
2020-03-15 12:53:40.818128: I tensorflow/compiler/xla/service/service.cc:176] StreamExecutor device (0): Host, Default Version
Model: "sequential_5"
Layer (type) Output Shape Param #
conv1d_9 (Conv1D) (None, 40, 128) 768
activation_13 (Activation) (None, 40, 128) 0
dropout_9 (Dropout) (None, 40, 128) 0
max_pooling1d_5 (MaxPooling1 (None, 5, 128) 0
conv1d_10 (Conv1D) (None, 5, 128) 82048
activation_14 (Activation) (None, 5, 128) 0
dropout_10 (Dropout) (None, 5, 128) 0
flatten_5 (Flatten) (None, 640) 0
dense_5 (Dense) (None, 8) 5128
activation_15 (Activation) (None, 8) 0
Total params: 87,944
Trainable params: 87,944
Non-trainable params: 0
Prediction is sad
(speech) agent1409@Zuric:~/Projects/SpeechRecog$ python3 livePredictions.py

Home Page


Activities
Google Chrome
Wed Apr 22, 19:50:44
100 %
127.0.0.1:8000
Apps
Go to Gmail
Go to YouTube
Go to Facebook
video-downlo...
Deep Learnin...
Letter - Googl...
Gmail
YouTube
Maps
Home
About
Analyze
Emotion Detection Using Speech Data
Analyzing the Emotion using the Speech Data
acoustic signal
pitch
loudness
time in seconds
©joomla_speech_prosody
rsz_home.png
Show all

Accuracy graph

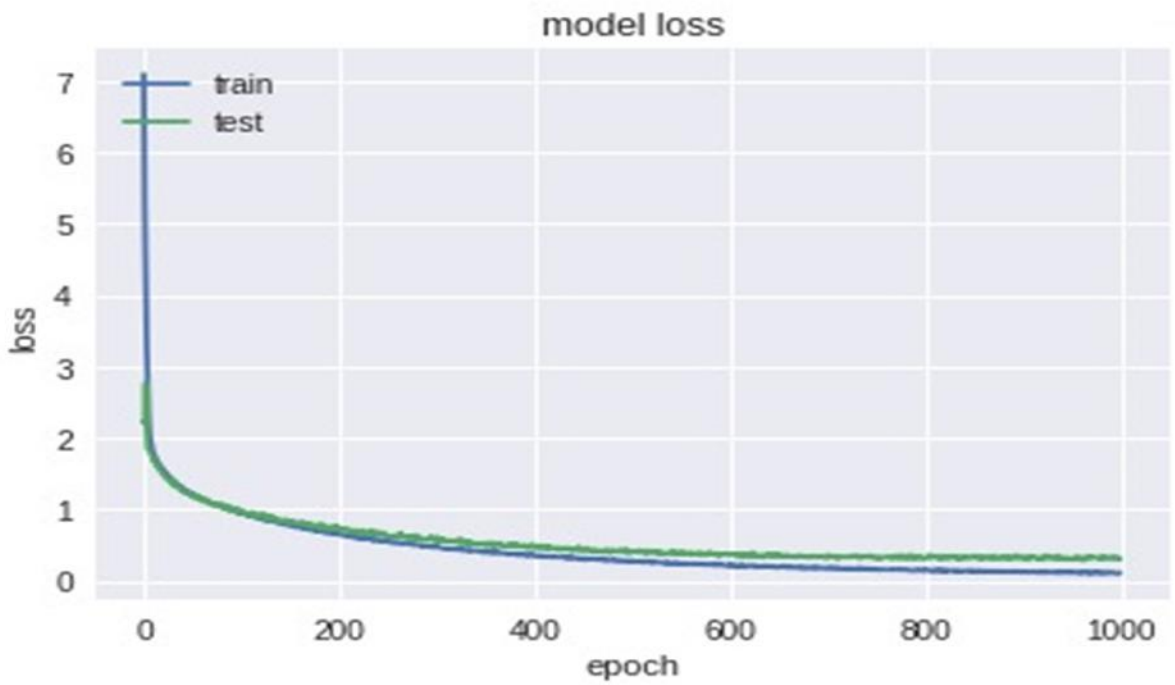


	precision	recall	f1-score	support
0	0.93	0.91	0.92	134
1	0.92	0.93	0.92	251
2	0.91	0.89	0.90	242
3	0.84	0.90	0.87	271
4	0.96	0.94	0.95	253
5	0.92	0.91	0.91	239
6	0.95	0.93	0.94	127
7	0.90	0.85	0.88	116
accuracy			0.91	1633
macro avg	0.92	0.91	0.91	1633
weighted avg	0.91	0.91	0.91	1633

```
from sklearn.metrics import confusion_matrix
matrix = confusion_matrix(new_Ytest, predictions)
print (matrix)

# 0 = neutral, 1 = calm, 2 = happy, 3 = sad, 4 = angry, 5 = fearful, 6 = disgust, 7 = surprised
```

```
[[128   2   0   3   0   0   1   0]
 [  2 226   7   8   0   0   8   0]
 [  4   1 220   4   5   6   0   2]
 [  2   2   2 241   3   5   7   9]
 [  2   0   2   2 244   0   1   2]
 [  1   0   2  19   0 214   2   1]
 [  0   0   2   2   2   0 121   0]
 [  0   2   4   0   0   2   2 106]]
```



Layer (type)	Output Shape	Param #
conv1d_3 (Conv1D)	(None, 40, 128)	768
activation_4 (Activation)	(None, 40, 128)	0
dropout_3 (Dropout)	(None, 40, 128)	0
max_pooling1d_2 (MaxPooling1	(None, 5, 128)	0
conv1d_4 (Conv1D)	(None, 5, 128)	82048
activation_5 (Activation)	(None, 5, 128)	0
dropout_4 (Dropout)	(None, 5, 128)	0
flatten_2 (Flatten)	(None, 640)	0
dense_2 (Dense)	(None, 8)	5128
activation_6 (Activation)	(None, 8)	0

Total params: 87,944
Trainable params: 87,944
Non-trainable params: 0

Hardware and Software Requirement

SOFTWARE REQUIREMENTS

- Operating system : Windows 7 &higher
- Coding Language : Python 3 (LibROSA)
- Tool : Anaconda

Continue….

HARDWARE REQUIREMENTS

- System : Pentium Dual Core.
- Input Devices : Keyboard, Mouse
- Ram : 8 GB

Conclusion

- We have Obtained a Voice analysing Model which not only determines the words in the Sentence but also determines the emotions in them.
- Energy and pitch are two important features for voice emotion recognition.

References

- Andy Sun, Maisy Wieman ,Analysing Vocal Patterns to Determine Emotions 2018
- Mithesh chouhan , Speech-Emotion-Analyzer,2018
- Kamran Soltani Raja,Noor Ainon, Speech Emotion Detection based on Neural networks,2007
- Chul Min Lee, Student Member,IEEE, and Shrikanth S.Narayanan,Senior Member,,Toward Detecting Emotions in Spoken Dialogs,2005
- Jangsik Cho,Shobei Kato,Detecting Emotions from Voice,2011

THANK YOU