”SPEECH EMOTION RECOGNITION USING MLP CLASSIFIER IN MACHINE LEARNING”

Minor project report submitted
in partial fulfillment of the requirement for award of the degree of

Bachelor of Technology
in
Computer Science & Engineering

By

SANTOSH MEHATA (19UECS0873)
SUJIT YADAV (19UECS0937)
SHYAM KUMAR SAH (19UECS0909)

Under the guidance of
DR.C CHANDRU VIGNESH,M.E.,Ph.D
ASSOCIATE PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
SCHOOL OF COMPUTING

VEL TECH RANGARAJAN DR. SAGUNTHALA R&D INSTITUTE OF SCIENCE AND TECHNOLOGY

(Deemed to be University Estd u/s 3 of UGC Act, 1956)
Accredited by NAAC with A Grade
CHENNAI 600 062, TAMILNADU, INDIA

December,2021

CERTIFICATE

It is certified that the work contained in the project report titled ”SPEECH EMOTION RECOGNITION USING MLP CLASSIFIER IN MACHINE LEARNING” by SANTOSH MEHATA (19UECS0873) SUJIT YADAV (19UECS0937) SHYAM KUMAR SAH (19UECS0909) has been carried out under my/our supervision and that this work has not been submitted elsewhere for a degree.

Signature of Supervisor
Dr.C Chandru Vignesh
Associate professor
Computer Science & Engineering
School of Computing
Vel Tech Rangarajan Dr.Sagunthala R&D
Institute of Science and Technology
December,2021

Signature of Head of the Department
Dr. V. Srinivasa Rao
Professor & Head
Computer Science & Engineering
School of Computing
Vel Tech Rangarajan Dr.Sagunthala R&D
Institute of Science and Technology
December,2021

DECLARATION

We declare that this written submission represents my ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Signature)

SANTOSH MEHATA

Date: / /

(Signature)

SUJIT YADAV

Date: / /

(Signature)

SHYAM KUMAR SAH

Date: / /

APPROVAL SHEET

This project report entitled SPEECH EMOTION RECOGNITION USING MLP CLASSIFIER IN MACHINE LEARNING by (SANTOSH MEHATA (19UECS0873), (SUJIT YADAV (19UECS0937), (SHYAM KUMAR SAH (19UECS909) is approved for the degree of B.Tech in Computer Science & Engineering.

Examiners **Supervisor**

Dr C Chandru Vignesh M.Tech.,

Date: / /

Place:

ACKNOWLEDGEMENT

We express our deepest gratitude to our respected **Founder Chancellor and President Col. Prof. Dr. R. RANGARAJAN B.E. (EEE), B.E. (MECH), M.S (AUTO). DSc., Foundress President Dr. R. SAGUNTHALA RANGARAJAN M.B.B.S.** Chairperson Managing Trustee and Vice President.

We are very much grateful to our beloved **Vice Chancellor Prof. S. SALIVAHANAN,** for providing us with an environment to complete our project successfully.

We record indebtedness to our **Dean & Head, Department of Computer Science and Engineering Dr. V. SRINIVASA RAO, M.Tech., Ph.D.,** for immense care and encouragement towards us throughout the course of this project.

We also take this opportunity to express a deep sense of gratitude to Our Internal Supervisor **Supervisor name,degree.,(in capital letters)** for her cordial support, valuable information and guidance, he helped us in completing this project through various stages.

A special thanks to our **Project Coordinators Mr. V. ASHOK KUMAR, M.Tech., Ms. C. SHYAMALA KUMARI, M.E., Ms.S.FLORENCE, M.Tech.,** for their valuable guidance and support throughout the course of the project.

We thank our department faculty, supporting staff and friends for their help and guidance to complete this project.

SANTOSH MEHATA (19UECS0873)
SUJIT YADAV (19UECS0937)
SHYAM KUMAR SAH (19UECS0909)

ABSTRACT

Our aim is to design and develop a web based system for a car rental company. This system enables admin to rent a car that can be used by a customer on payment basis .The car information can be added to the system or existing car information can be edited or deleted too by the administrator .Hence this system enhances the car and customer management and provides customer satisfaction their by maintaining customer retention.The motivation behind this research is the growing popularity of web-based systems and the need to explore that industries could tap into to enhance their services to the customers. This paper described a notification-based content alert and web-based system using . It was specifically developed for the alert notification to the customers about the car rental information, and the availability of the car reserved. The main purpose of developing car rental system is to reduce the cost and time consumed, which is beneficial to the car rental agencies and customers

Keywords:
MLP-Classifier, MFCC,Neural Networks, Prediction

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

abbr	Abbreviation
CHROMA	A Chroma vector is typically a 12-element feature vector indicating how much energy of each pitch class is present in the signal in a standard chromatic scale.
MLP	MLPClassifier stands for Multi-layer Perceptron classifier which in the name itself connects to a Neural Network. Unlike other classification algorithms such as Support Vectors or Naive Bayes Classifier, MLPClassifier relies on an underlying Neural Network to perform the task of classification.
MFCC	MFCC was by far the most researched about and utilized feature in this dataset. It represents the short-term power spectrum of a sound.
MEL SPECTOGRAM	This is just a spectrogram that depicts amplitude which is mapped on a Mel scale.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

1.1 Introduction

speech emotion recognition (SER) is becoming increasingly important in various applications. At present, speech emotion recognition is an emerging crossing field of artificial intelligence and artificial psychology. The research is widely applied in human-computer interaction, interactive teaching, entertainment, security fields, and so on. Speech emotion processing and recognition system is generally composed of three parts, the first being speech signal acquisition, then comes the feature extraction followed by emotion recognition. The most propitious technique for speech recognition is the neural network based approach. Artificial Neural Networks. Speech recognition is the process of converting an acoustic signal, captured by microphone or a telephone, to a set of characters. There are also many models which were proposed earlier to improve the predicting accuracy of the SERS. For example, we have Support Vector Machine (SVM), which is a classifier that mathematically computes the parameters of the audio signal to be able to predict the emotion. This model has been very successful in the domain of SER. But the main disadvantage with SVM's is that it can only classify the data into two classes; either class 1 or 2. And other disadvantages include processing time, noise leading to errors in prediction

1.2 Aim of the project

The audio files are given as the input. The data sets travels through a number of blocks of processes which makes it executable to help for the analysis of the speech parameters. The data is preprocessed to change it to the suitable format and the respective features from the audio files are extracted using various steps such as framing, hamming, windowing, etc.

1.3 Scope of the Project

To enter the unknown test dataset as an input, it will retrieve the parameters and predict the emotion as per training dataset values. The accuracy of the system is displayed in the form of percentage which is the final result of our project.

1.4 Methodology

EXISTING SYSTEMS:
The speech emotion detection system is implemented as a Machine Learning (ML) model.The first step is data collection, which is of prime importance. The model being developed will learn from the

data provided to it and all the decisions and results that a developed model will produce is guided by the data.

PROPOSE SYSTEM:

proposed a system that considered that the emotion expressed by humans are mostly a result of mixed feeling. Therefore, they suggested an improvement over the SVM algorithm that would consider mixed signals and choose the most dominant one. For this purpose, a ranking SVM algorithm was chosen. The ranking SVM takes all predictions from individual binary classification SVM classifiers also called as rankers, and applies it to the final multi-class problem. Using the ranking SVM algorithm, an accuracy of 65.458 percent was achieved in their system

FEATURE EXTRACTION:

Speech is a varying sound signal. Humans are capable of making modifications to the sound signal using their vocal tract, tongue, and teeth to pronounce the phoneme. The features are a way to quantify data. A better representation of the speech signals to get the most information from the speech is through extracting features common among speech signals

Chapter 2

LITERATURE REVIEW

.[1]Jerry Joy, Aparna Kannan, Shreya Ram, S. Rama
In this paper the emotions in the speech are predicted using neural network. Multi-Layer Perceptron Classifier (MLP Classifier) is used for the classification of emotions , 2020 Positive emotion recognition rate is higher than other approaches but neutral and negative emotions are often confused with each other.Then features learned by the 1D CNN and 2D CNN are transferred to the Merged CNN. Then, the merged deep CNN is initialised with the transferred features. Two hyperparameters of these architectures were chosen from Bayesian optimisation in the training. H.K. Palo et al. in their research use MLP network for Emotion Recognition
.[2]Navya Damodar, Vani H Y, Anusuya M
A Voice Emotion Recognition using CNN and Decision Tree .This model can be improved by making the dataset times three times the original size to achieve a greater accuracy 2020.The underlying emotion in our speech is reflected in our voice through tone and pitch. In this paper we aim to classify elicit types of emotions such as sad, happy, neutral, angry, disgust,surprised, fearful and calm. In this paper the emotions in the speech are predicted using neural networks.
.[3] Chaitanya Singlaa, Sukhdev Singhb,Monika Pathak,
Automatic Audio Based Emotion Recognition System: Scope and Challenges,Common sense Affect-based approach (Realworld Knowledgeconcept models) 2020.Since our focus is on recognise emotions from speech,our model is trained on Audio-only data. Two fixed statements are vocalised by all the 24 actors for all the 8 emotions, with each statement repeated twice. All emotional expressions are uttered at two levels of intensity: normal and strong, except for the ‘neutral’ emotion, it is produced only in normal intensity.

Chapter 3

PROJECT DESCRIPTION

3.1 Existing System

Speech emotion recognition using CNN,deep learning etc.

3.2 Proposed System

speech emotion recognition using meachine learning in different language such as python programming.

Advantages

3.3 Feasibility Study

our goal is to study the feasibility of an end-to end emotion recognition system and to compare BLSTM models with and without the attention mechanism in the formation of the utterance representations.

3.3.1 Economic Feasibility

It predict the emotion of people and make our work easier. for example, suppose people are calling in call center when staff of call center understand the people emotion then they deal with them easily and the market of this industries grow efficiency.

3.3.2 Technical Feasibility

IN this system, we take the voice sample of some people, this voice sample pass through different parameter where the noise from the voice simple remove and then it pass through MLP classifier where it predict the emotion with accuracy result.

3.3.3 Social Feasibility

It help to understand people behaviour. it also help our security people to understand the emotion of criminal.it also helps to in education system.

3.4 System Specification

3.4.1 Hardware Specification

- hardware spec used database for voice sample storage. used mlp classifier to predict emotion.

3.4.2 Software Specification

- software spec We need one programming language like python.
  TO run the code we need one compiler like Anconda.

Chapter 4

MODULE DESCRIPTION

4.1 General Architecture

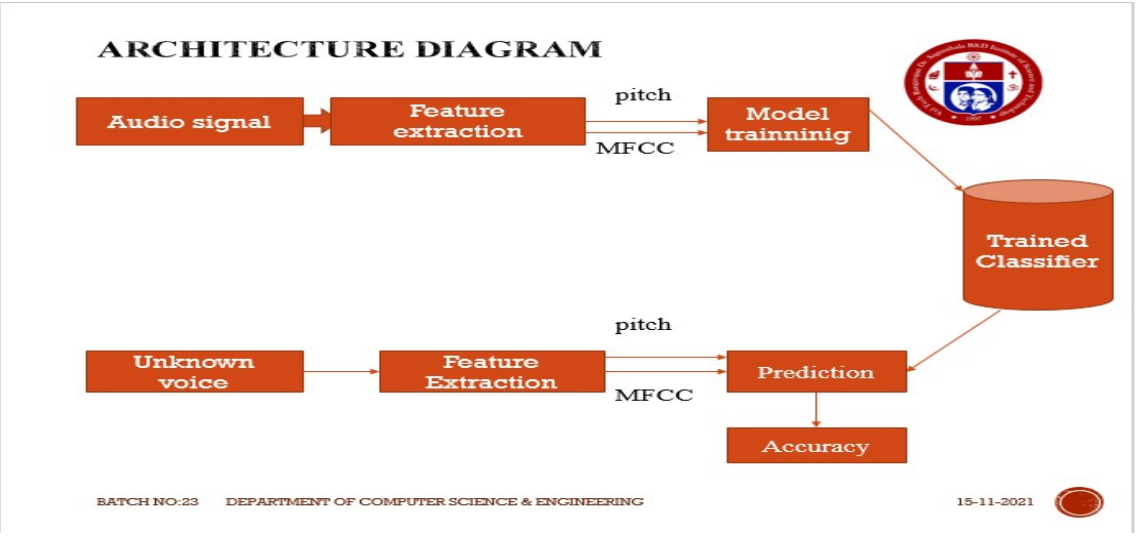


Figure 4.1: **Architecture Diagram**

Description

TO provide audio voice sample that audio voice sample passes through different parameters and there is feature extraction through which noise is remove from audio sample then after it goes into MLC classifier where MLC classifier predict the emotion with accuracy and it provide the result.

4.2 Design Phase

4.2.1 Data Flow Diagram

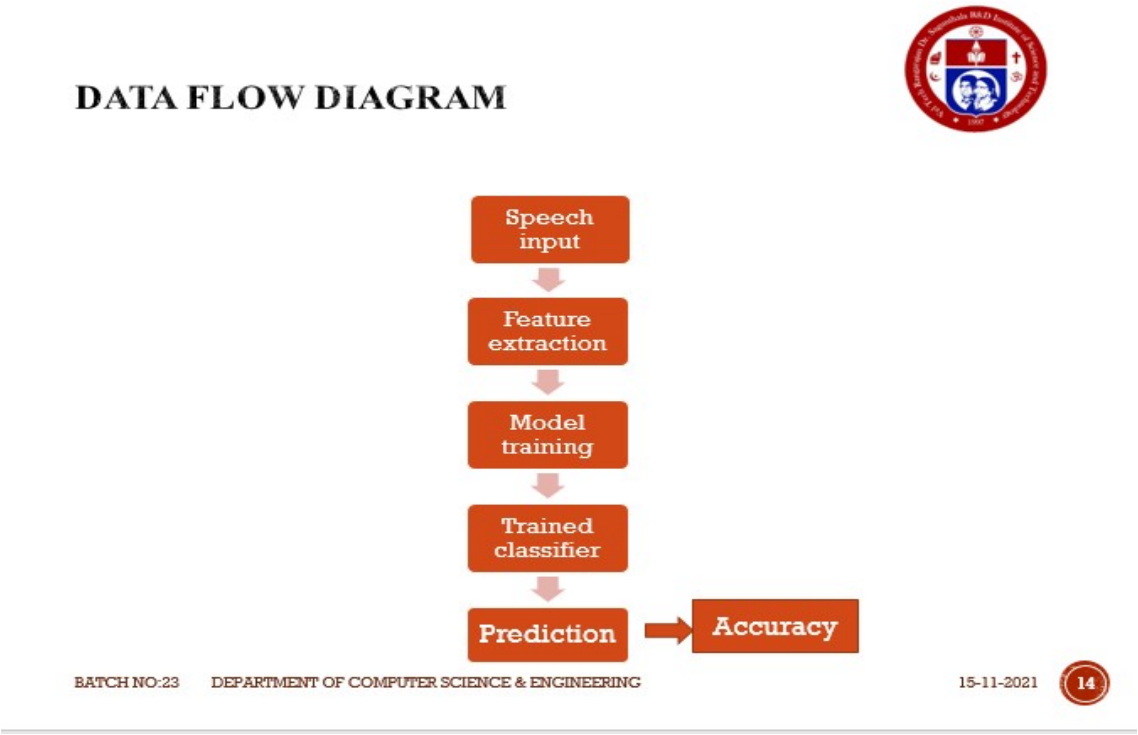


Figure 4.2: **Data Flow Diagram**

Description

4.3 Module Description

MODULE:1
Feature abstraction module training
The input to the model should be the features extracted along with the emotion category that it belongs to, stored correspondingly into respective arrays so that, classifier will be able to identify the patterns, correlations and then classify the data. This training helps the model to understand, which emotions have what range of the respective features. So, when an unseen data is given as an input, it will be able to correlate and predict the emotion.
MODULE 2:
TO GIVE TRAINED DATASET TO MLP CLASSIFIER
Subsequent work with multilayer perceptrons has shown that they are capable of approximating an XOR operator as well as many other non-linear functions. Multilayer perceptrons are often applied to supervised learning problems. They train on a set of input-output pairs and learn to model the correlation (or dependencies) between those inputs and outputs. The network being real values. If you

have categorical data, such as a sex attribute with the values “male” and “female”, emotion attributes such as “happy”, “sad”, “angry” etc. you can convert it to a real-valued representation which is called a one hot encoding.

Module 3:

To test the unknown audio dataset as input

Once a neural network has been trained it can be used to make various predictions. You can make predictions on test data in order to estimate the skill of the model on unseen data. You can also deploy it operationally and use it to make predictions continuously.

Chapter 5

IMPLEMENTATION AND TESTING

5.1 Input and Output

5.1.1 INPUT Design

```
RavdessData = "C:\\Users\\sujit yadav\\Desktop\\minor project\\voice sample\\audio_speech_actors_01-24\\"
ravdessDirectoryList = os.listdir(RavdessData)
fileEmotion = []
filePath = []
```

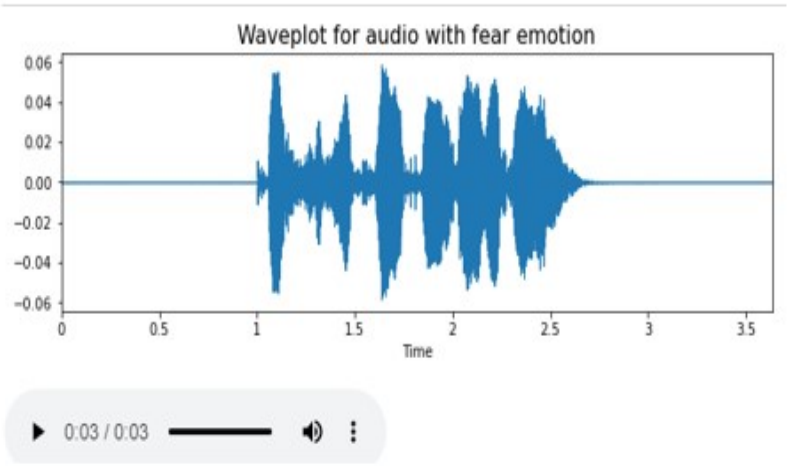


Figure 5.1: **Data Flow Diagram**

5.1.2 Output Design

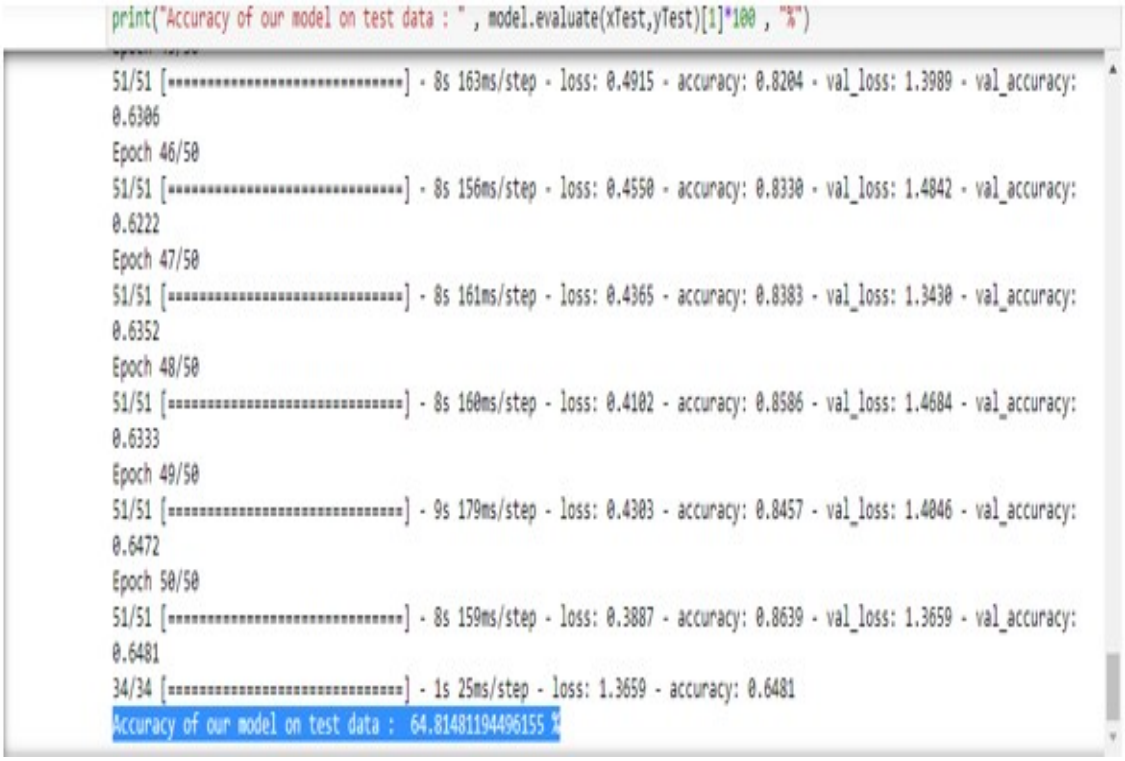


Figure 5.2: **Data Flow Diagram**

5.2 Testing

5.3 Types of Testing

UNIT TESTING
INTEGRATION TESTING
FUNCTIONAL TESTING

5.3.1 Unit testing

It concentrates on the tiniest a aspect of software development it is where we test a single unit or a collection of interconnected units. Using sample input and watching the accompanying outputs is a common method used by programmers.

5.3.2 Integration testing

The goal is to take unit –tested components and use them to create a program structure that is dictated by design. Integration testing occurs when a set of components is combined to achieve a result.

5.3.3 Functional testing

Every time a new module is added , the program is altered . This sort of testing ensures that the entire component functions properly even after it has been added to the entire application.

5.3.4 Test Result

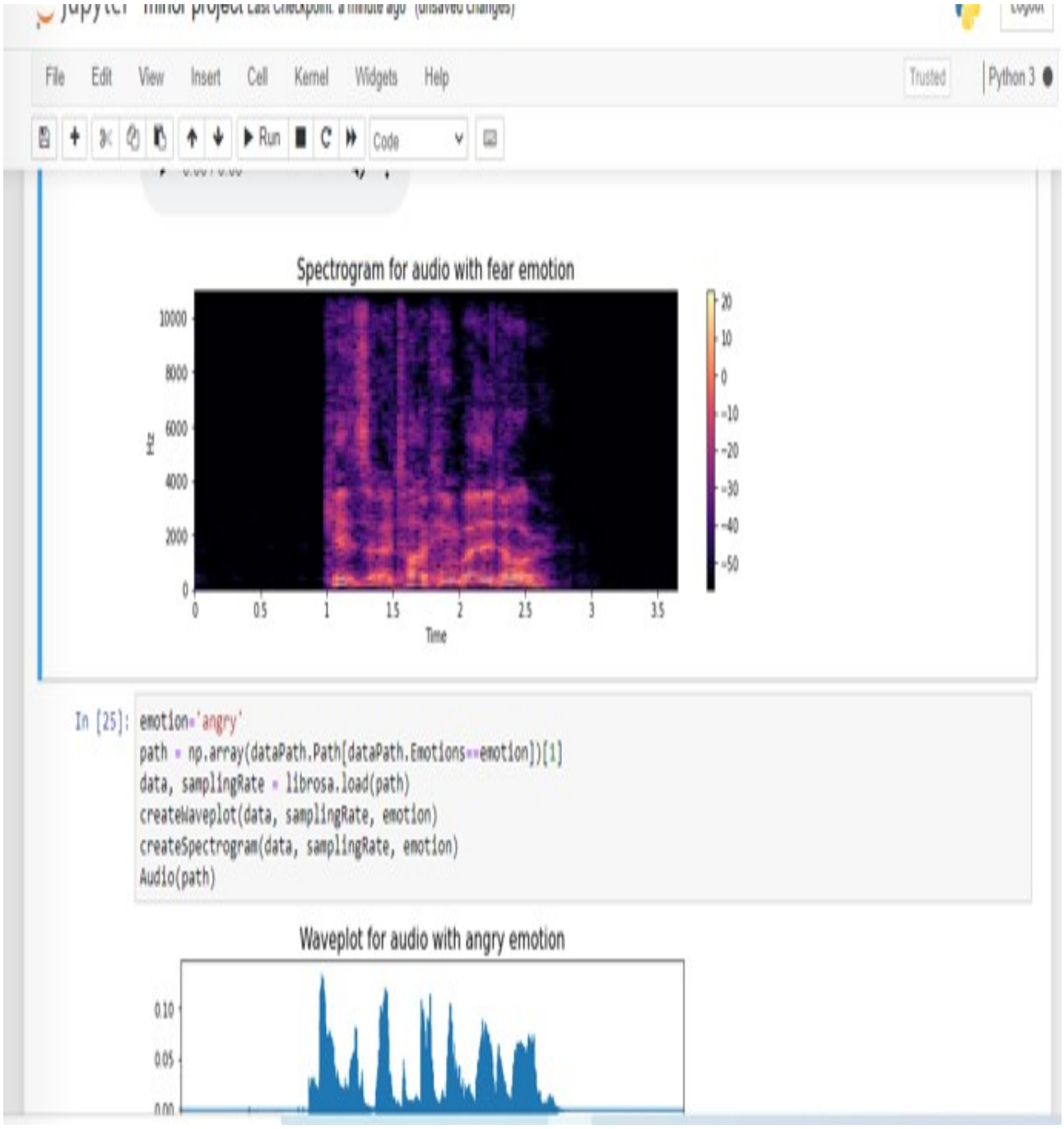


Figure 5.3: **Test Image**

Chapter 6

RESULTS AND DISCUSSIONS

6.1 Efficiency of the Proposed System

Efficiency of the proposed system is 65 percent.

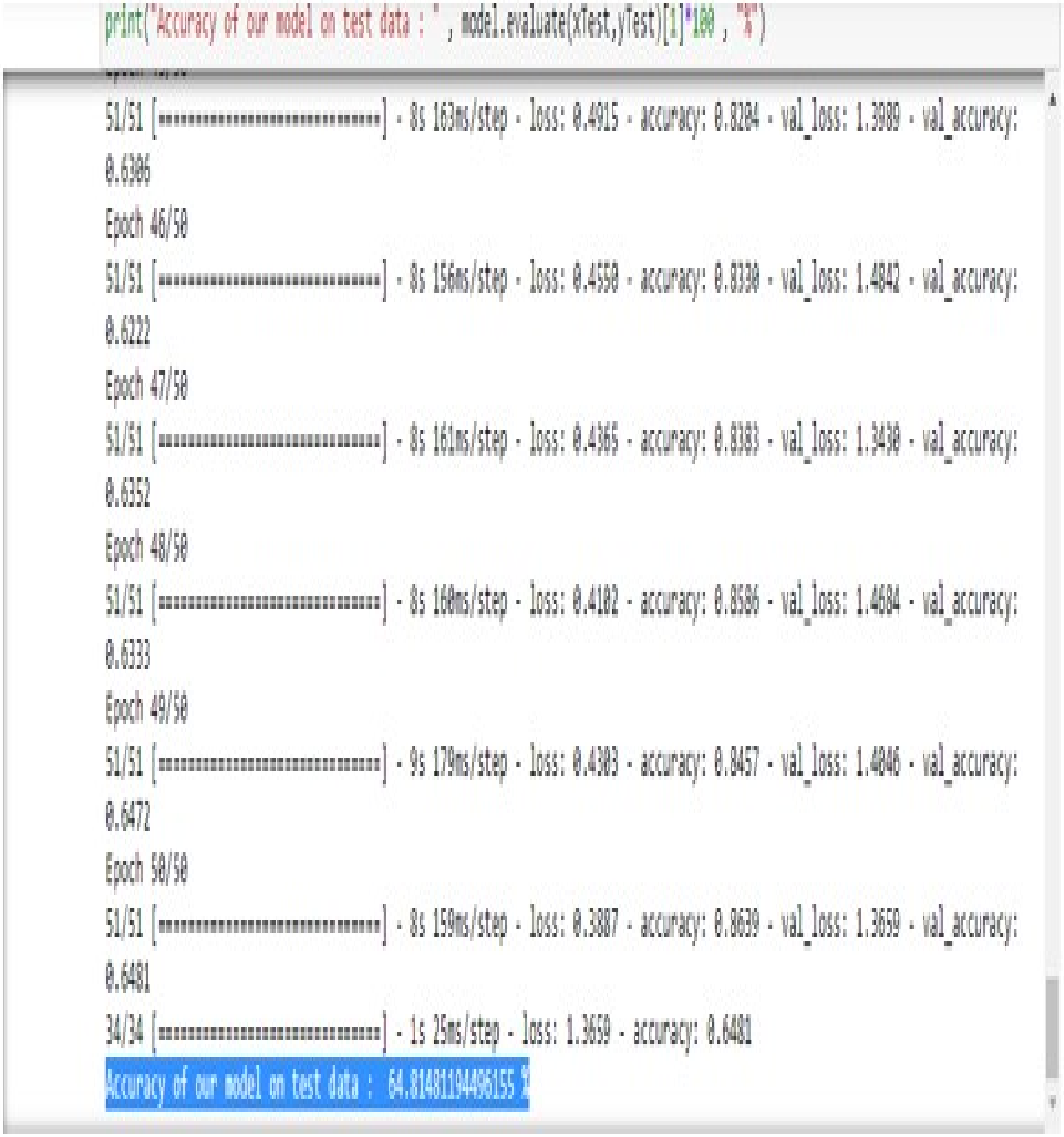


Figure 6.1: **Test Image**

dimensionality to improve the presentation.

6.2 Comparison of Existing and Proposed System

Better efficiency compare to Existing to proposed system.

6.3 Advantages of the Proposed System

Less number of parameters required

Higher performance compared to previous systems

Better classification of parameters is shown.

Can handle missing values, model complex relationships and support multiple inputs

6.4 Sample Code

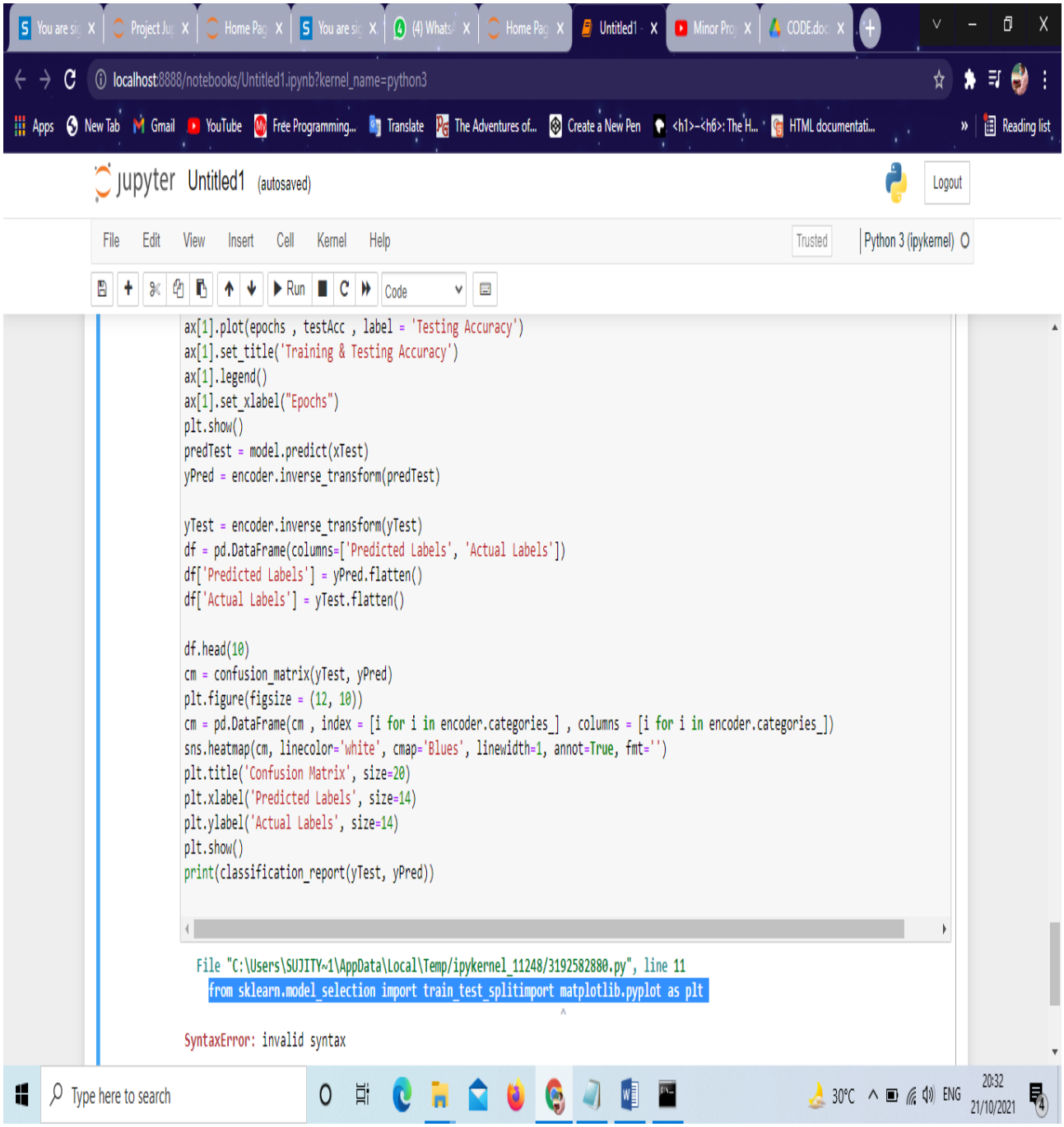


Figure 6.2: **Test Image**

```
write your code here
main code
```

Output

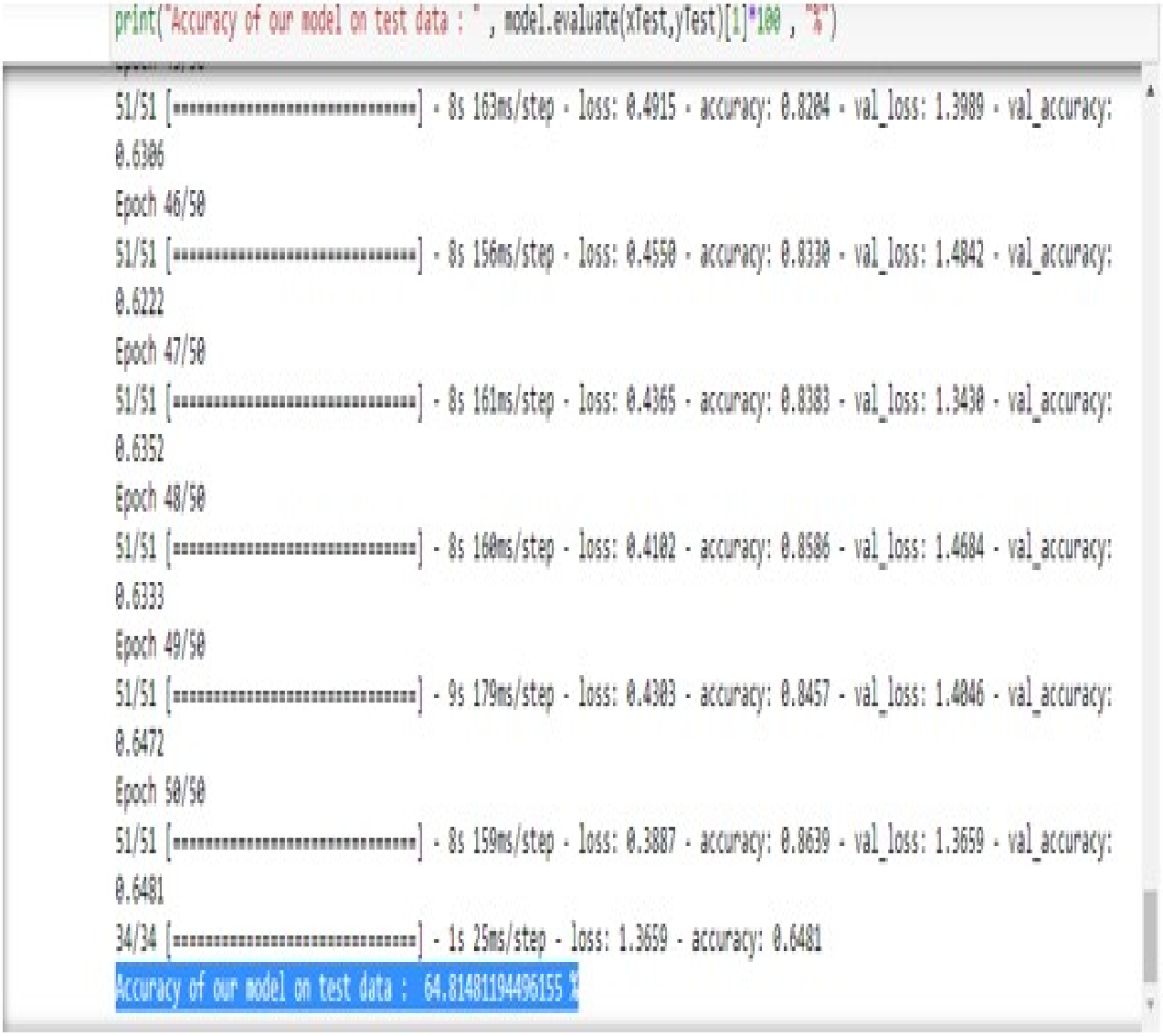
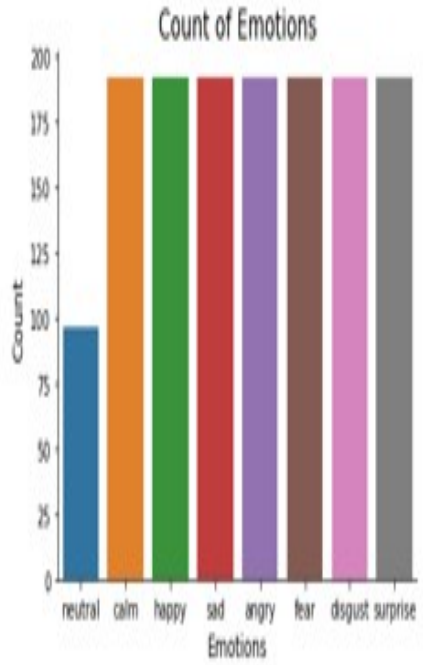


Figure 6.3: **Output 1**

```
dataPath.head()
plt.title('Count of Emotions', size=16)
sns.countplot(dataPath.Emotions)
plt.ylabel('Count', size=12)
plt.xlabel('Emotions', size=12)
sns.despine(top=True, right=True, left=False, bottom=False)
plt.show()
```




```
In [23]: def createWaveplot(data, sr, e):
             plt.figure(figsize=(10, 3))
             plt.title('Waveplot for audio with {} emotion'.format(e), size=15)
             librosa.display.waveplot(data, sr=sr)
```

Figure 6.4: **Output 2**

Chapter 7

CONCLUSION AND FUTURE ENHANCEMENTS

7.1 Conclusion

This project shows that MLPs are very powerful in classifying speech signals. Even with simplified models, a limited set of characters can be easily identified. We have obtained higher accuracies as compared to other approaches for individual emotions. The performance of a module is highly dependent on the quality of pre-processing. The results obtained in this study demonstrate that speech recognition is feasible, and that MLPs can be used for any task concerning recognizing of speech and demonstrating the accuracy of each emotion present in the speech.

7.2 Future Enhancements

To improved in result accuracy with neural networks technology.

Chapter 8

PLAGIARISM REPORT

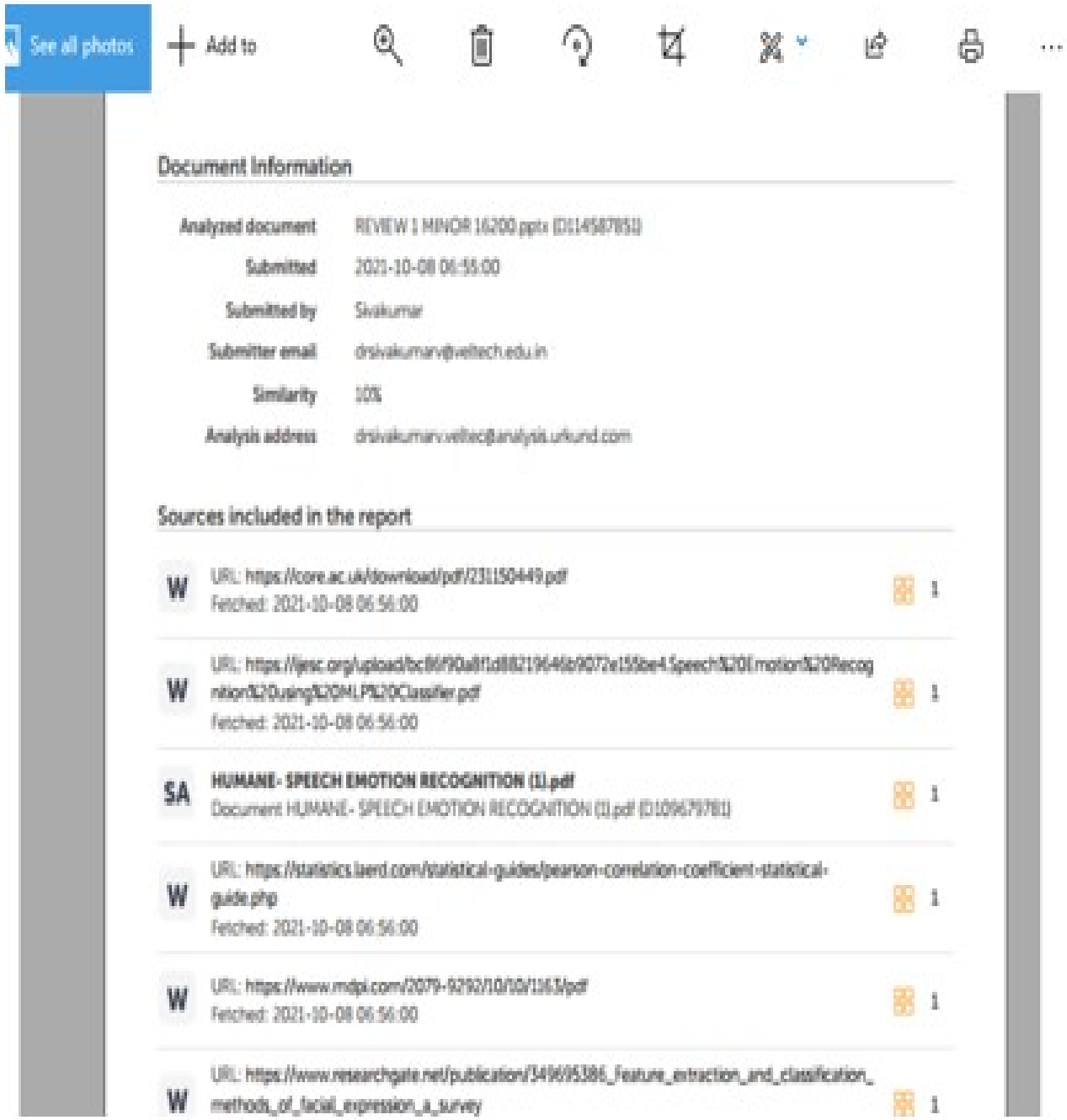


Figure 8.1: **Output 2**

Chapter 9

SOURCE CODE & POSTER PRESENTATION

9.1 Source code

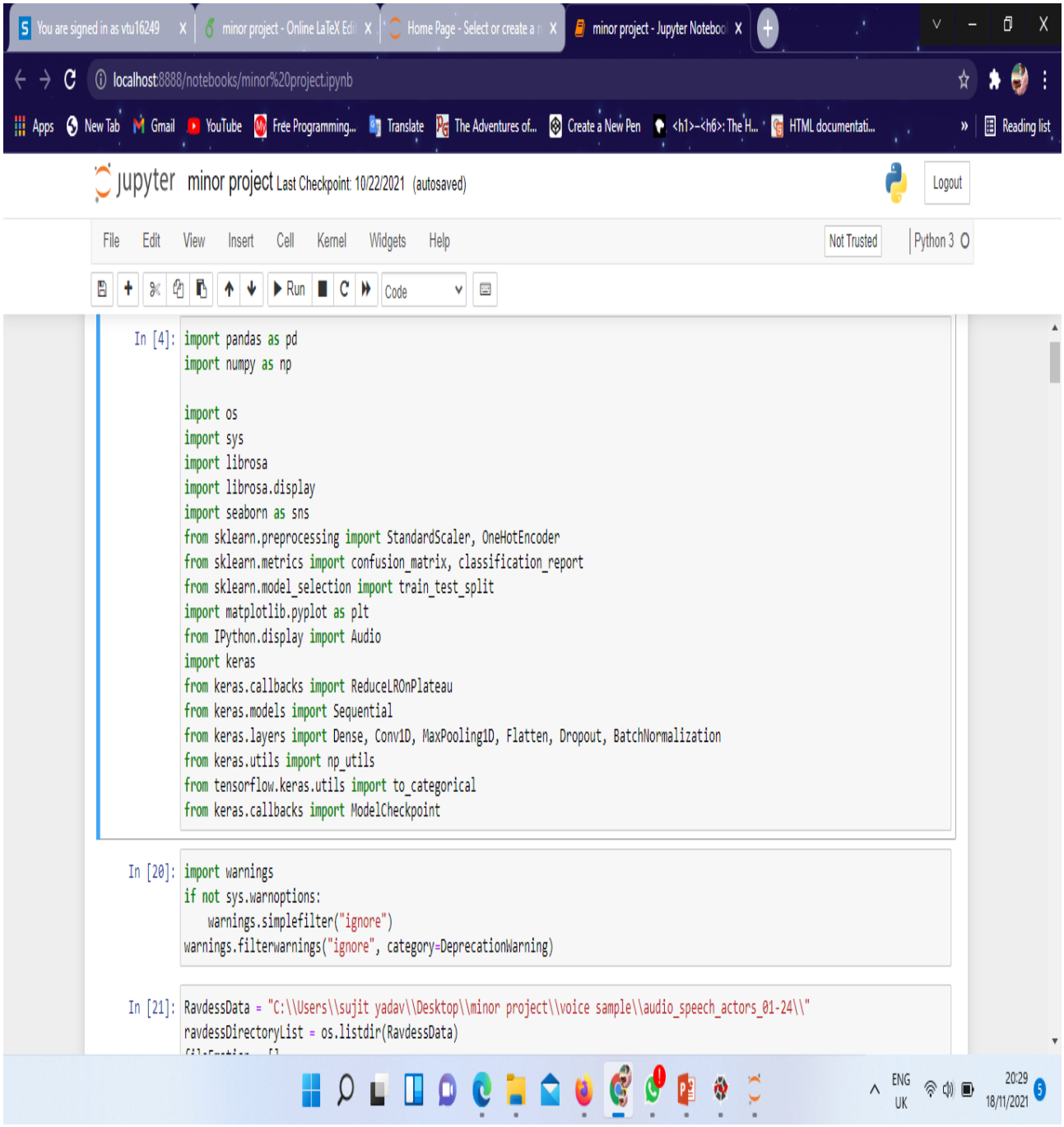


Figure 9.1: **Output 2**

9.2 Poster Presentation

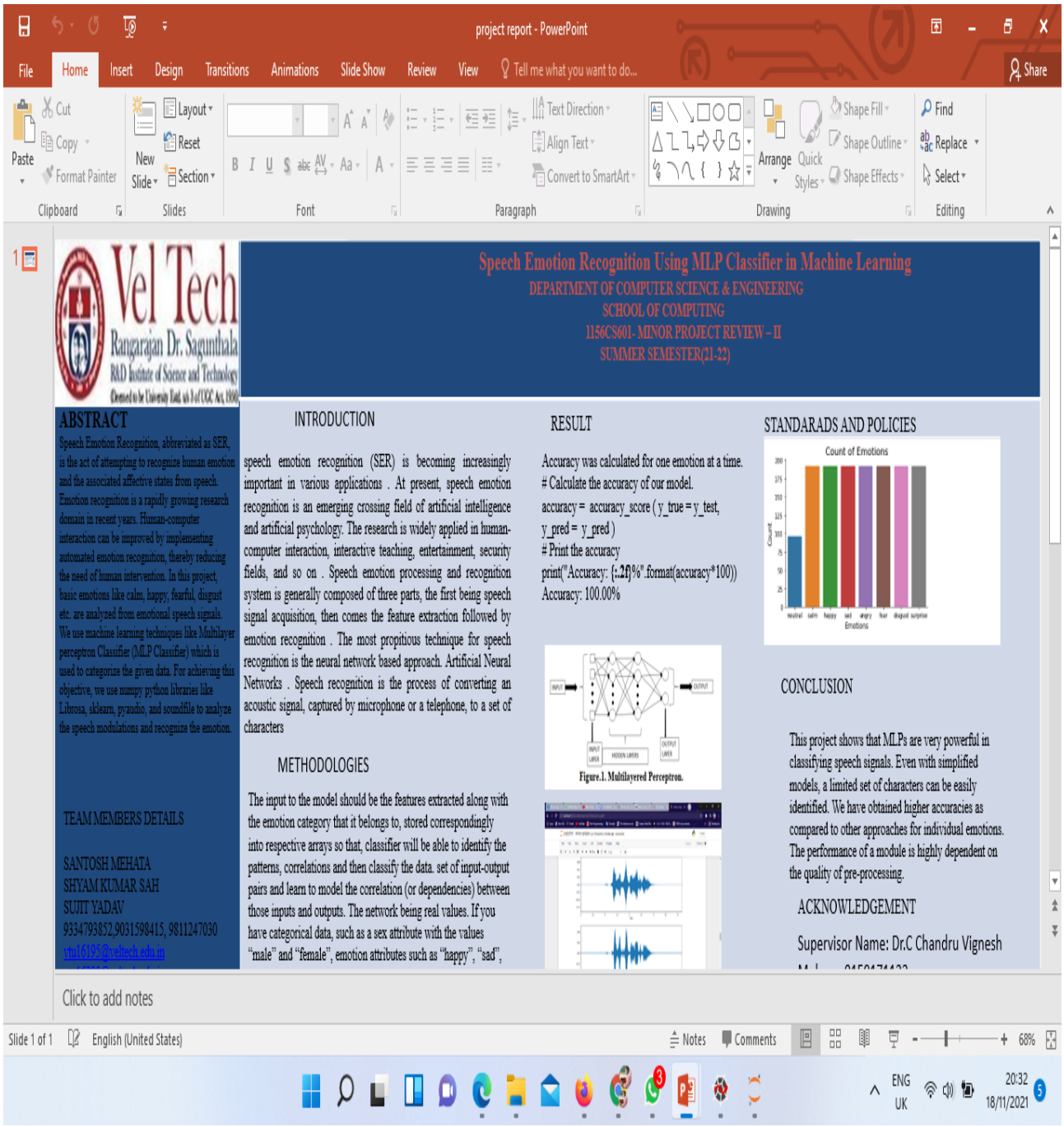


Figure 9.2: **Output 2**

References

[1] [1]Yang, Ningning, Dev Nilanjan, y Sherratt, R. Simon, Shi Fuqian. Recognize basic emotional statesin speech by machine learning techniques using melfrequency cepstral coefficient features. Journal of intelligent and fuzzy system, vol, 39 no, 2 pp.1925-1936, 2020. [2]Navya Damodar, Vani H Y, Anusuya M A. Voice Emotion Recognition using CNN and Decision Tree. InternationalJournal of Innovative Technology and Exploring Engineering(IJITEE), October 2019 [3] Sarma, M., Ghahremani, P., Povey, D., Goel, N. K., Sarma, K. K., Dehak, N. Emotion Identification from Raw Speech Signals Using DNNs. In Interspeech (pp. 3097-3101), 2018 [4] Brownlee, J. (2018). A Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning - Machine Learning Mastery.Machine Learning Mastery.[5].Jianfeng Zhao, Xia Mao, Lijiang Chen. Learning Deep features to Recognise Speech Emotion using Merged Deep CNN. IET Signal Process., 2018 [6].H.K. Palo, Mihir Narayana Mohanty and Mahesh Chandra. Use of different features for Emotion Recognition using MLP network. Springer India 2015, Computational Vision and Robotics, Advances in Intelligent Systems and Computing [7].Ayush Kumar Shah ,Mansi Kattel,Araju Nepal. Chroma Feature Extraction using Fourier Transform. Chroma$_{F}eature_{x}traction.January2019[8].Davis, S.Mermelstein, P.(1980)Comparisonof Para$
366.