A PROJECT REPORT ON

“Speech Based Emotion Recognition”

Submitted

In the partial fulfilment of the requirements for

The award of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE & ENGINEERING

By

S.Deepak (171FA04428)

V.Shanmukhi Reddy(171FA04436)

Under the esteemed guidance of

Mrs. P. Jhansi Lakshmi, Assistant Professor.

VIGNAN’S
Foundation for Science, Technology & Research
(Deemed to be University)
-Estd. u/s 3 of UGC Act 1956

NAAC 'A' Accredited

NBA NATIONAL BOARD OF ACCREDITATION

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

VIGNAN'S FOUNDATION FOR SCIENCE, TECHNOLOGY AND RESEARCH

Deemed to be UNIVERSITY

Vadlamudi, Guntur.

VIGNAN'S FOUNDATION FOR SCIENCE, TECHNOLOGY AND RESEARCH Deemed to be UNIVERSITY

VADLAMUDI, GUNTUR DIST, ANDHRA PRADESH, INDIA, PIN-522 213.

VIGNAN’S
Foundation for Science, Technology & Research
(Deemed to be University)
-Estd. u/s 3 of UGC Act 1956

NAAC 'A' Accredited
NBA NATIONAL BOARD OF ACCREDITATION

CERTIFICATE

This is to certify that the Project Report entitled **“Speech Based Emotion Recognition”** that is being submitted by **S.Deepak (171FA04428) and V.Shanmukhi Reddy(171FA04436)** in partial fulfilment for the award of B.Tech degree in Computer Science and Engineering to the Vignan’s Foundation for Science, Technology and Research, Deemed to be University, is a record of bonafide work carried out by them under my supervision.

Mrs. P. Jhansi Lakshmi	**External Examiner**	**Dr. Dondeti Venkatesulu**
Assistant Professor.		**Professor, HOD.**

DECLARATION

I hereby declare that the project entitled **“Speech Based Emotion Recognition”**submitted for the **DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**. This dissertation is our original work and the project has not formed the basis for the award of any degree, associate-ship and fellowship or any other similar titles and no part of it has been published or sent for publication at the time of submission.

By

S.Deepak(171FA04428)

V.Shanmukhi Reddy(171FA04436)

Date:

ACKNOWLEDGEMENT

Semester long project is a golden opportunity for learning and self-development. Consider myself very lucky and honored to have so many wonderful people lead me through in the completion of this project.

We feel it is our responsibility to thank **Mrs. P. Jhansi Lakshmi** under whose valuable guidance that the project came out successfully after each stage, and also it is our responsibility to extend our thanks to **Mr. D. Yakobu, Mr. P. Vijaya Babu, Dr. Balaji Vijayan, Department Project Coordinators,** for extending their support towards the completion of project in **Department of CSE**, in **VFSTR**.

It is a great pleasure for us to express our sincere thanks to **Dr. Venkatesulu Dondeti, HOD, CSE** of VFSTR Deemed to be University**,** for providing me an opportunity to do our project at Department of CSE.

It is our pleasure to extend our sincere thanks to **Vice-Chancellor Dr. M.Y.S. Prasad** and we are very grateful to our beloved **Chairman Dr. Lavu. Rathaiah**, and **Vice Chairman Mr. Lavu. Krishna Devarayalu**, for their love and care.

We extend our whole hearted gratitude to all our **faculty members** of Department of Computer Science and Engineering who helped us in our academics throughout course.

Finally we wish to express thanks to our family members for the love and affection overseas and forbearance and cheerful depositions, which are vital for sustaining effort, required for completing this work.

With Sincere regards,
S.Deepak(171FA04428).
V.Shanmukhi Reddy(171FA04436).

Date:

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATION

ANN	Artificial Neural Networks
MFCC	Mel Frequency Cepstral Coefficients
MLP	Multi-layer Perceptron
STFT	Short-time Fourier transform
SVM	Support vector machines
KNN	K-Nearest Neighbors
HMM	Hidden Markov Model
GMVAR	Gaussian Mixture Vector Autoregressive
GMM	Gaussian Mixture Model

ABSTRACT

Speech is the most usual way of conveying our intention as humans. So**,** to expand these services to computer applications, here we use Speech based Emotion recognition System. We can define Speech based emotion recognition (SER) systems as a collection of methods that processes and classifies speech signals and detects the embedded emotions from a given audio. Human emotion recognition plays a vital role in the interpersonal relations. So, understanding and extraction of emotion is of high importance for the interaction and communication between human and machine.This project critically examined the current accessible approaches in speech based emotion recognition methods based on the three considering parameters (feature selection, classification of features, accuracy). Moreover, it highlights the ongoing hopeful direction for improvement of speech emotion recognition systems.

CHAPTER - 1

INTRODUCTION

1.1: Introduction

Machine Learning is very useful for a variety of real-life problems. It is commonly used for tasks such as classification, recognition, detection and predictions. Moreover, it is very efficient to automate processes that use data. The basic idea is to use data to produce a model capable of returning an output. This output may give a right answer with a new input or produce predictions towards the known data.

The goal of this project is to train a Machine Learning algorithm capable of classifying emotions over different audio files based on some features like Tone, Pitch, Frequency. This particular problem can be useful for recognition of emotion based on speech., For example. The method we'll be using here is machine learning model with the help of Artificial Neural Networks based on Sklearn and Librosa modules.

Artificial Neural Networks is part of a broader family of Machine learning methods. It is based on the use of layers that process the input data, extracting features from them and producing a mathematical model. In this specific project, we'll be aiming to classify different audio files based on emotions, which means that the computer will have to "learn" the features of each emotion and classify them correctly. For example, if it is given an audio which is of happy emotion, then the output of the model needs to classify the respective audio into a happy category.

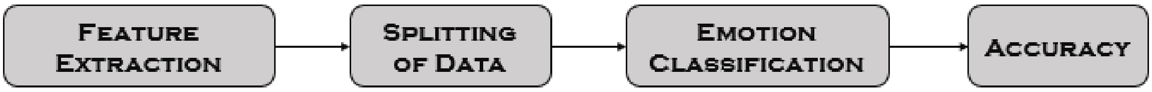


Figure 1-1:Design

1.2: Applications:

Speech based Emotion recognition system has been applied for different applications on different domains as an emotion classifier. For example, Call centres, Investigation centre, Psychology, Medical systems etc. Overview of some Speech based Emotion Recognition areas are listed below.

1.2.1: Call centers:

Speech based Emotion Recognition systems could also be used in noticing company’s associations with customers through call centers. Currently to examine the emotions in such a consultation, a human specialist with limited capabilities has to be included.

But, if one engages machines to do this task it provides a probability to respond to customers as per the detected emotional state or to pass the control over then the task will be much cheaper and output will be more accurate.

1.2.2: Investigation centers:

Investigation could also be benefitted from such an approach. Namely, it is feasible to analyze the embedded emotions in the voices, i.e. speeches of leaders. Such knowledge could be of high value and treasurable for the society, as frame of mind and uprightness of politicians could be investigated.

1.2.3: Psychology:

Human Emotion perception either through Speech or Face became a relatively preliminary research area. Speech based Emotion Recognition covers the undertaking and receiving a speaker's feelings from their dialogue chronicles. Receiving feelings from dialogue can go far in concluding one’s physical and mental condition of prosperity. These emotions will also be used for additional assessment of patient's position for better diagnosis.

1.2.4: Medical Assessment:

Emotion recognition can also be employed for medical and health assessment. Thus, machine learning algorithm will be a helpful tool for the classification of emotions. While many models have been established in this sector, there is a lack of practical representations for classification of emotions for therapy. Here, we introduce a tool which facilitates users to take audio samples and recognize a range of emotions (happy, sad, angry, surprised, disgust, and fear) from audio elements through a machine learning model. This project is designed based on local therapists needs for innate representations of data in order to gain informative analyses and intuition of their sessions with their patients.

CHAPTER - 2

LITERATURE SURVEY

2.1: Literature survey

During the past years, an intense investigation has been accomplished to acknowledge emotions by using speech. Cao et al. [1] introduced a ranking SVM model for harmonizing information regarding emotion recognition to resolve the binary classification problem. This method, directs SVM algorithm for some specific emotions, considering data from each and every speaker as a different query then mixing all prognosis from rankers to cover multi-class prediction. Ranking approach reaches considerable gain with regard to accuracy when measured to conventional SVM in two public datasets of acted emotional speech, Berlin and LDC. In both of those acted data, spontaneous data, which comprises neutral extreme emotional speeches, ranking-based SVM reached better accuracy in identifying emotional speeches than the conventional SVM methods. Accuracy achieved is 44.4% [1].

Narayanan [2] introduced a domain-specific emotion recognition by making use of speech signals from a call center application. In this research, recognizing positive(happy) and negative(anger) emotions are of the focus of attention. There are various types of information that is utilized for emotion prediction and it includes acoustic, lexical, and discourse. Moreover, Info Theoretic contents of emotional importance is introduced to acquire data at emotion information at the basic level. Both classifiers k-NN and linear discriminant are used to work with various kinds of features. The result of the experiment certifies that when we use the combination of acoustic and language data the finest results are obtained. The accuracy of the classification improves by 40.7% for males and 36.4% for females. When Compared to previous work the percentage increase in accuracy is from 0.75% to 3.96% for female and 1.4% to 6.75% for male.

Albornoz et al. [3] introduced a new spectral feature so as to achieve the classification of emotions and to distinguish groups. In this research, emotions are classified based on acoustic features and a novel hierarchical classifier. Various classifiers such as HMM, GMM and MLP have been analyzed with different configuration and input properties to design a novel hierarchical techniques for emotion classification. The discovery of the suggested method is of two things, first the selection of major executing features and second is engaging of major class-wise grouping performance of total properties are identical as the classifier. The result in Berlin dataset signifies that the hierarchical approach reaches the greater performance when contrasted to best standard classifier, with decuple cross-validation. For example, HMM method achieved 68.57% and the hierarchical model achieved 71.75% [3].

Lee et al. [4] introduced hierarchical structure for binary decision tree to classify the recognized emotions. This method focuses on detection of the basic classification obstacle at higher level of tree to decrease the accumulation of error. This method also depicts the input speech data into one of the emotion classes via followed by the binary classification layer. The output achieves 70.1% for two-class and 65.1 for a five-class problem respectively. As a replacement solution instead of out- putting hard labels at every step, computing of possibility as a soft label can makes the framework useful for modeling. Bayesian Logistic Regression and SVM were worked as a binary classifier.

El Ayadi et al. [5] introduced a Gaussian mixture vector autoregressive (GMVAR) approach. It is a mixture of GMM with vector autoregressive for classifying speech emotion recognition problem. The main idea of GMVAR is its ability to multi-modality in their distribution and to design the dependency between the feature set of speech. Berlin emotional dataset is used for estimating of GMVAR. The result depicts that the accuracy of classification achieves 76% when for HMM reached 71%, for k-NN 67% and 55% for feed-forward neural networks.

Ref	Types of classifiers	Types of features	Recognition Rate	Type of Dataset	Methods
[1]	SVM	Prosodic and spectral features	44.4%	Berlin & LDC & FAU Aibo dataset	Ranking SVM
[2]	K-NN & linear discriminate	Fundamental frequency (F0), energy, duration, and the first and formant	40.7% for males & 36.4% for females	Private speech database from call center	Domain-specific emotion recognition by k-NN and linear discriminate classifier
[3]	HMM, GMM, MLP and hierarchical model	Mean of the log-spectrum (MLS), MFCCs and prosodic features	HMM 68.57, Hierarchical model 71.75	Berlin dataset	Spectral characteristics of signals are used in order to group emotions based on acoustic rather than psychological considerations
[4]	Bayesian Logistic Regression, SVM	Large-margin feature	70.1% & 65.1% for two and five class	AIBO dataset	Hierarchical structure for binary decision tree
[5]	G GMVAR	Mel-frequency spectrum coefficient (MFCC)	76%	Berlin emotional speech database	Gaussian mixture vector autoregressive (GMVAR) is a mixture of GMM with vector autoregressive for classification.

Table 1-1:Literature Survey Overview

CHAPTER - 3

SOFTWARE REQUIREMENTS SPECIFICATION

3.1: Software Requirements

The software interface is the operating system, and there are several other requirements for the development of emotion recognition model.

Operating System : Windows

Technologies Used : Python, Artificial Neural Networks(ANN)

Platform : Geany, Sublime Text

3.2: Packages Required

3.2.1: Numpy

- NumPy is a Python package. It stands for “Numerical Python” It is a library consisting of multidimensional array objects and a collection of routines for processing of array.
- Installing NumPy and getting started
- Standard Python distribution doesn’t come bundled with NumPy module. A lightweight alternative is to install NumPy using the popular Python package installer, pip.
- pip install numpy
- “Import numpy as np”

3.2.2: Seaborn

Seaborn is an open source, BSD-licensed Python library providing high level API for visualizing the data using the Python programming language. In the world of Analytics, the best way to get insights is by visualizing the data. Data can be visualized by representing it as plots which are easy to understand, explore and grasp. Such data helps in drawing the attention of key elements.

To analyse a set of data using Python, we make use of Matplotlib, a widely implemented 2D plotting library. Likewise, Seaborn is a visualization library in Python. It is built on top of Matplotlib.

3.2.3: Matplotlib

Matplotlib is one of the most popular Python packages used for data visualization. It is a crossplatform library for making 2D plots from data in arrays. It can be used in Python and IPython shells, Jupyter notebook and web application servers also.

3.2.4: Sklearn

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistent interface in Python. This library, which is largely written in Python, is built upon NumPy, SciPy and Matplotlib.

3.2.5: Librosa

Librosa is powerful Python library built to work with audio and perform analysis on it. It is the starting point towards working with audio data at scale for a wide range of applications such as detecting voice from a person to finding personal characteristics from an audio.

3.2.6: Sound file

SoundFile can read and write sound files. File reading/writing is supported through libsndfile, which is a free, cross-platform, open-source (LGPL) library for reading and writing many different sampled sound file formats that runs on many platforms including Windows, OS X, and Unix. SoundFile represents audio data as NumPy arrays.

3.2.7: Glob

Glob is a general term used to define techniques to match specified patterns according to rules related to Unix shell. Linux and Unix systems and shells also support glob and also provide function glob() in system libraries.

3.2.8: OS

OS module provides functions which is very helpful in interacting with the operating system.

CHAPTER - 4

PROPOSED METHODOLOGY

The method we are using for emotion recognition is machine learning with the help of Artificial Neural Networks(ANN) based on Librosa and Sklearn. In this method how the emotion is recognised from the audio is explained clearly.

4.1: The Overview of the Model

Emotion recognition from audio requires feature extraction, in this phase totally 180 features are extracted from each audio file. The features like pitch, energy, frequency, intensity etc.., are extracted by using MFFC, CHROMA, MEL methods which are crucial to recognize a particular emotion accurately. The types of emotions those are recognised by this model are Happy, Angry, Sad, Fearful, Disgust, Surprised. After the extraction of each category of features from audio, the mean of each category is calculated and data is stored. The training and testing data is separated. By using MLP (Multi-Layer Perceptron) Classifier the emotion is classified. The accuracy of the model is calculated by comparing classified data and testing data.

4.2: Model:

|

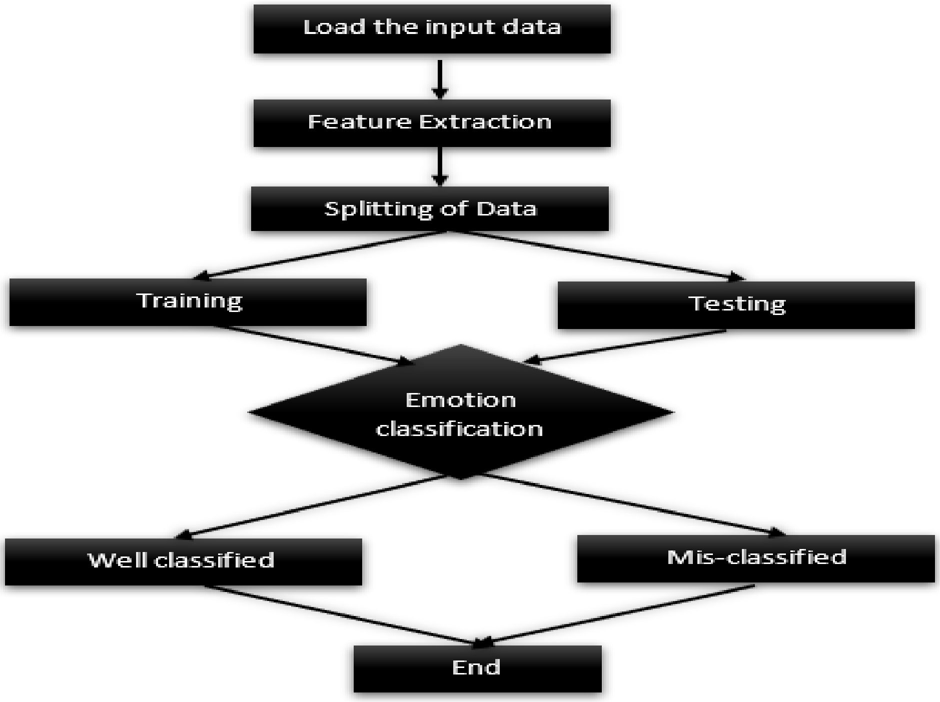


Fig 4-1:Flow Chart

4.3: Loading the Input

The audio files can be loaded to the model, after specifying the location of the data set in the program i.e., by using glob function. Before we begin the process of feature extraction, we find number of frames in each audio file by using sound_file.read(dtype="float32") this function reads the number of frames in the given format. Here in this model we are using float32 bit format. It returns the output as an array. We also find the sample rate of each audio file by the function sound_file.samplerate. It returns an integer specifying how many times per second a sound is sampled i.e., the frequency of samples used in recording.

Data set

In this project the dataset consists of 1200 audio files varies of 6 different emotions they are Happy, Angry, Sad, Fearful, Disgust, Surprised. The data set is named as ravdess-data.

4.4: Feature extraction

In this phase by using librosa module we are extracting three categories of features by MFCC, CHROMA and MEL functions.

4.4.1: MFCC(Mel Frequency Cepstral Coefficient) Features

By using librosa.feature.mfcc() function, we can extract 39 features, in that 12 values corresponds to the cepstral coefficients and energy terms and other 27 values corresponds to delta and double delta values. The parameters that are required for mfcc features are sample rate of audio file, number of frames and n_mfcc, here this parameter is used for specifying how many features that it should extract.

4.4.2: CHROMA Features

By using librosa.feature.chroma_stft() function, we can extract 120 features. These corresponds to twelve different pitch classes and some are related to the amplitude. The parameters that are required for chroma features are sample rate of audio file, number of frames and for extracting Chroma features, we should also calculate stft value by using librosa.stft(number of frames). The STFT represents a signal in the time-frequency domain by computing discrete Fourier transforms (DFT) over short overlapping windows.

4.4.3: MEL Features

By using librosa.feature.melspectrogram() function, we can extract 21 features related to magnitude and frequency of the spectrogram. The parameters that are required for Mel spectrogram are sample rate of audio file and number of frames.

After extracting each category of features from each file, the mean is calculated for each category of features by using NumPy module based on np.mean() and the output is stored in horizontal stack for further use.

4.5: Splitting of Data

In this phase, by using sklearn module, we are splitting training and the testing data. This process is done by train_test_split() function. This function splits the training and testing data based on the values given for the parameters' test_size and train_size, for example if test_size=0.1 then 10% of files from the dataset is taken for testing and other are taken for training. The parameters those are also to be passed with this function are mean of the features, emotions and random state which will generate random number and shuffle the audio files every time when they are splitting and decides splitting of data into training or testing.

4.6: Emotion Classification

In this phase, by using sklearn module, we are classifying six types of emotions Happy, Angry, Sad, Fearful, Disgust, Surprised with the help of a classifier in Artificial Neural Network (ANN) i.e., MLP classifier.

4.6.1: MLP (Multi-layer Perceptron) Classifier

A MLP is a feedforward artificial neural network that generates a set of outputs from a set of inputs, it is characterized by several layers of input nodes connected as a direct graph between the input and output layers. It uses backpropagation as a supervised learning technique. In this, each node apart from input nodes has a non-linear activation function. By using MLPClassifier(), we can classify six types of emotions by giving training data as input. The parameters that are to be passed to MLPClassifier() are alpha value which is a L2 penalty parameter. Batch size is a parameter which specifies how many inputs can be taken at a time, hidden layer sizes represents number of neurons in the ith hidden layer. Learning rate is used for weight updates. There are three types values can be given for it those are constant, adaptive, invscaling in our model. We are using adaptive learning rate which keeps learning rate constant as long as training loss keeps decreasing. Activation, it specifies the algorithm for weight optimization across the nodes. There are four algorithms which are considered as its values. They are relu, identity, logistic, tsnh. Here, in our model, we are using logistic sigmoid function. Max_iter, by using this, we can specify the maximum number of iterations to the solver. The solver iterates until the convergence of this number of iterations. The model.fit() function used for training model using dataset. It consists of two parameters x and y where x is taken as input and y is the target. The model.predict() function enables us to predict the labels i.e., emotions of the data values on the basis of trained model.

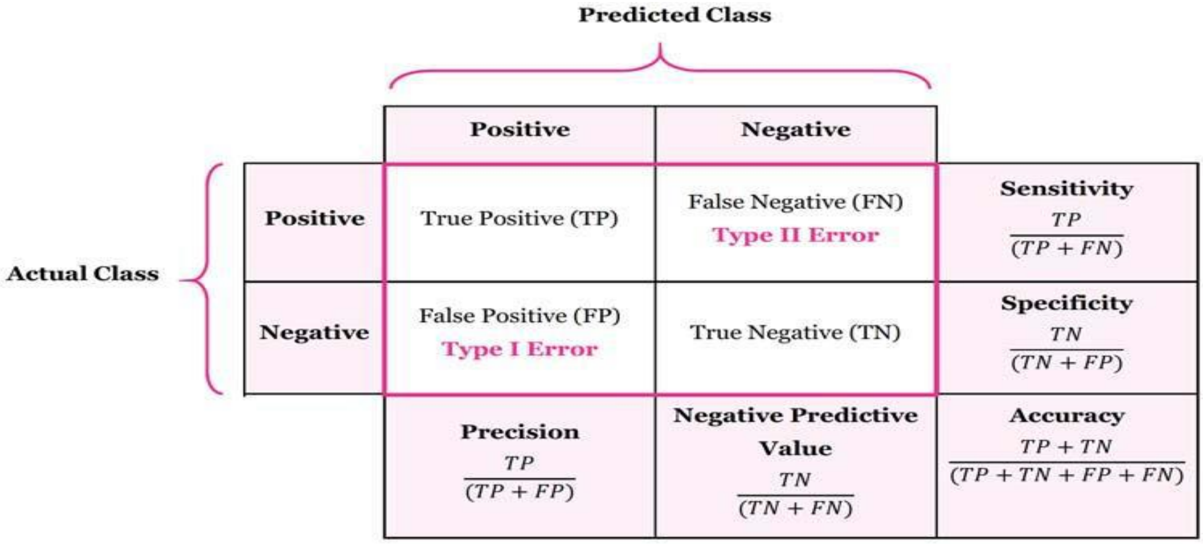
4.7: Accuracy

The accuracy of model is nothing but, how well the model is classifying the emotions correctly. The accuracy of model is calculated by using accuracy_score(), it returns the accuracy of the model by comparing predicted data with the testing data.

4.8: Performance evaluation metrics

4.8.1: Confusion matrix

A confusion matrix is an N X N matrix, where N is the number of classes being predicted. The Confusion matrix is one of the most intuitive and easiest metrics used for finding the correctness and accuracy of the model. It contains number of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).



Predicted Class

		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error	Sensitivity $\frac{TP}{(TP+FN)}$
	Negative	False Positive (FP) Type I Error	True Negative (TN)	Specificity $\frac{TN}{(TN+FP)}$
		Precision $\frac{TP}{(TP+FP)}$	Negative Predictive Value $\frac{TN}{(TN+FN)}$	Accuracy $\frac{TP+TN}{(TP+TN+FP+FN)}$

Fig 4-2:Confusion Matrix

Accuracy:

The proportion of the total number of predictions that were correct.

The diagonal values in the confusion matrix generated by our model are correctly classified data and other then those values are mis-classified data.

CHAPTER - 5

IMPLEMENTATION

5.1: Workflow

This process of developing this model is divided into four different phases

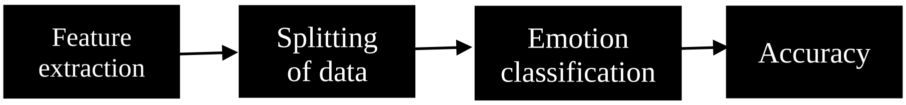


Fig 5-1:Workflow

5.1.1: Feature extraction

In this phase, by using librosa module, we are extracting three categories of features by MFCC, CHROMA and MEL functions. The means of each category of features are stored in a horizontal stack. Totally 180 features are extracted from each audio file.

5.1.2: Splitting of data

In this phase, by using sklearn module, we are splitting training and the testing data. This process is done by train_test_split() function.

5.1.3: Emotion classification

In this phase, by using sklearn module, we are classifying six types of emotions Happy, Angry, Sad, Fearful, Disgust, Surprised with the help of a classifier in Artificial Neural Network (ANN) i.e., MLP classifier.

5.1.4: Accuracy

The accuracy of model is calculated by using accuracy_score(), it returns the accuracy of the model by comparing predicted data with the testing data.

5.2: Sample Code

```
import librosa
import soundfile
import glob,os
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.neural_network import MLPClassifier
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns

def extract_feature(file_name, mfcc, chroma, mel):
    with soundfile.SoundFile(file_name) as sound_file:
        X = sound_file.read(dtype="float32")
        sample_rate=sound_file.samplerate
        if chroma:
            stft=np.abs(librosa.stft(X))
        result=np.array([])
        if mfcc:
            mfccs=np.mean(librosa.feature.mfcc(y=X, sr=sample_rate, n_mfcc=40),axis=1)
            result=np.hstack((result, mfccs))
        if chroma:
            chroma=np.mean(librosa.feature.chroma_stft(S=stft, sr=sample_rate),axis=1)
            result=np.hstack((result, chroma))
        if mel:
            mel=np.mean(librosa.feature.melspectrogram(X, sr=sample_rate),axis=1)
            result=np.hstack((result, mel))
    return result
```

```
emotions={
  '01':'neutral',
  '02':'calm',
  '03':'happy',
  '04':'sad',
  '05':'angry',
  '06':'fearful',
  '07':'disgust',
  '08':'surprised'
}
observed_emotions=['angry', 'happy', 'fearful', 'disgust','sad','surprised']

def load_data(test_size):
    x,y=[],[]
    count=0
    for file in glob.glob(r"C:\Users\Deepak\Desktop\speech-emotion-recognition-ravdess-data\Actor_*\*.wav"):
        file_name=os.path.basename(file)
        emotion=emotions[file_name.split("-")[2]]
        if emotion not in observed_emotions:
            continue
        if emotion in observed_emotions:
            count+=1
            print("file: ",count,"    ","emotion: ",emotion)
        feature=extract_feature(file, mfcc=True, chroma=True, mel=True)
        x.append(feature)
        y.append(emotion)
    return train_test_split(np.array(x), y, test_size=test_size, random_state=9)

x_train,x_test,y_train,y_test=load_data(test_size=0.1)
print((x_train.shape[0], x_test.shape[0]))
print(f'Features extracted: {x_train.shape[1]}')
model=MLPClassifier(alpha=0.01,batch_size=10,hidden_layer_sizes=(1500,),
activation='logistic',learning_rate='adaptive',max_iter=15000)
```

```
model.fit(x_train,y_train)
y_pred=model.predict(x_test)
print(x_train)
print(x_test)
print(y_test)
print(y_pred)

accuracy=accuracy_score(y_pred,y_test)
print("Accuracy: {:.2f}%".format(accuracy*100))
cm=confusion_matrix(y_test,y_pred)

co=[]
wr=[]
k=0
u=0
for i in range(6):
  for j in range(6):
    if i==j:
      co.append(cm[i][j])
      u=cm[i][j]
    else:
      k=sum(cm[i])-u
  wr.append(k)

fig = plt.figure(figsize = (10, 5))
plt.bar(observed_emotions,co, color ='red',width = 0.4)
plt.xlabel("emotions")
plt.ylabel("correctly classified data")
plt.title("correctly classified emotions")

ind = np.arange(6)
width = 0.4
fig = plt.subplots(figsize =(10, 7))
p1 = plt.bar(ind, co, width)
```

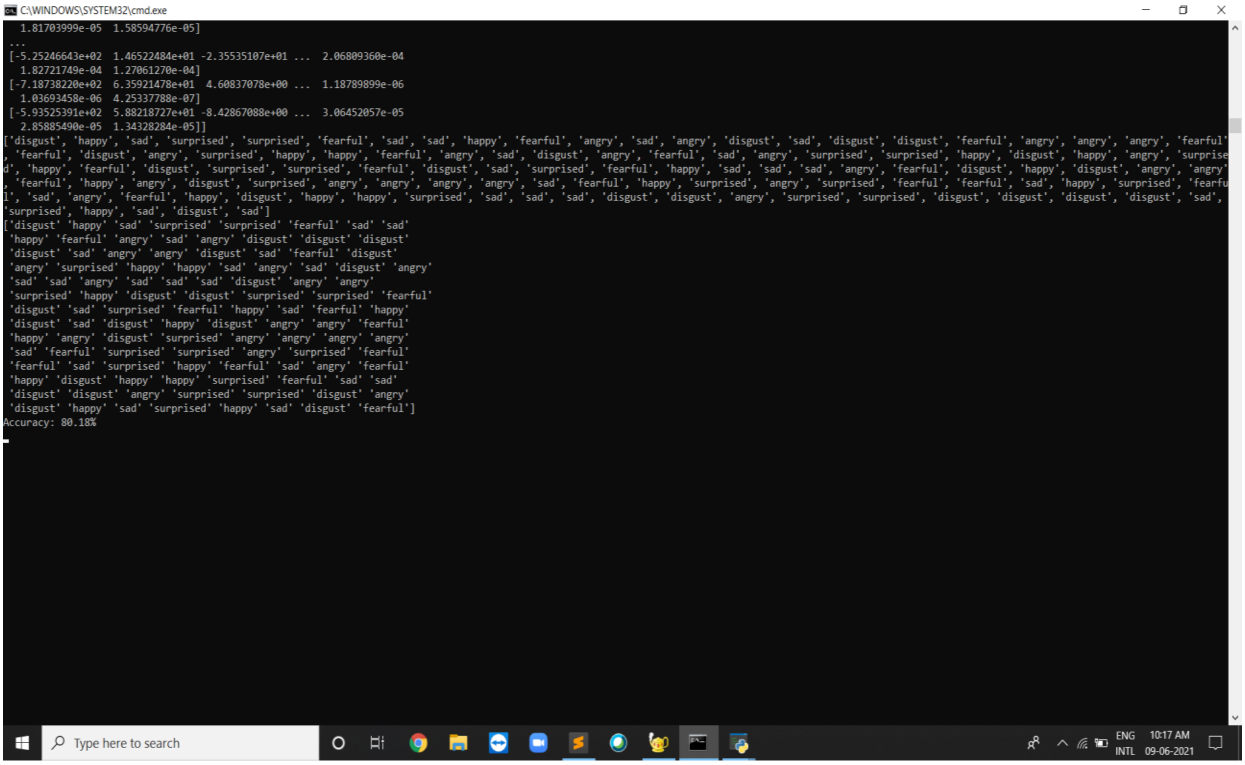
```
p2 = plt.bar(ind, wr, width,bottom = co)
plt.xlabel("emotions")
plt.ylabel("correctly classified data")
plt.title("emotion classification")
plt.xticks(ind, ('angry', 'happy', 'fearful', 'disgust','sad','surprised'))
plt.yticks(np.arange(0, 30, 7))
plt.legend((p1[0], p2[0]), ('Correctly classified data','Miss classified data'))
plt.show()

dig = np.arange(6)
plt.title("emotion classification metrics")
sns.heatmap(cm,annot=True,fmt="d",cmap='coolwarm',linewidth=.9)
plt.xticks(dig,('angry', 'happy', 'fearful', 'disgust','sad','surprised'))
plt.yticks(dig,('angry', 'happy', 'fearful', 'disgust','sad','surprised'))
plt.xlabel("emotions")
plt.ylabel("correctly classified data")
plt.show()
```

CHAPTER – 6

RESULT & ANALYSIS

6.1: Accuracy

For classifying the emotions from audio requires feature extraction. Here, we are extracting 180 features from each audio file. The features like pitch, energy, frequency, intensity etc.., are extracted by using MFFC, CHROMA, MEL methods which are crucial to recognize a particular emotion accurately. The types of emotions those are recognized by this model are Happy, Angry, Sad, Fearful, Disgust, Surprised. After the extraction of each category of features from audio, the mean of each category is calculated and data is stored. The training and testing data is separated. By using, MLP (Multi-Layer Perceptron) Classifier the emotion is classified. The accuracy of the model is calculated by comparing classified data and testing data. By using, this model we got an accuracy of 80.18% for these 112 sample files where the emotions for 86 audio files are classified correctly and 26 files are mis-classified. The best accuracy for this model is 80.18 %.



```
  1.81703999e-05  1.58594776e-05]
 ...
 [-5.25246643e+02  1.46522484e+01 -2.35535107e+01 ...  2.06809360e-04
  1.82721749e-04  1.27061270e-04]
 [-7.18738220e+02  6.35921478e+01  4.60837078e+00 ...  1.18789899e-06
  1.03693458e-06  4.25337788e-07]
 [-5.93525391e+02  5.88218727e+01 -8.42867088e+00 ...  3.06452057e-05
  2.85885490e-05  1.34328284e-05]]
['disgust', 'happy', 'sad', 'surprised', 'surprised', 'fearful', 'sad', 'sad', 'happy', 'fearful', 'angry', 'sad', 'angry', 'disgust', 'sad', 'disgust', 'disgust', 'fearful', 'angry', 'angry', 'angry', 'fearful'
, 'fearful', 'disgust', 'angry', 'surprised', 'happy', 'happy', 'fearful', 'angry', 'sad', 'disgust', 'angry', 'fearful', 'sad', 'angry', 'surprised', 'surprised', 'happy', 'disgust', 'happy', 'angry', 'surprise
d', 'happy', 'fearful', 'disgust', 'surprised', 'surprised', 'fearful', 'disgust', 'sad', 'surprised', 'fearful', 'happy', 'sad', 'sad', 'sad', 'angry', 'fearful', 'disgust', 'happy', 'disgust', 'angry', 'angry'
, 'fearful', 'happy', 'angry', 'disgust', 'surprised', 'angry', 'angry', 'angry', 'angry', 'sad', 'fearful', 'happy', 'surprised', 'angry', 'surprised', 'fearful', 'fearful', 'sad', 'happy', 'surprised', 'fearfu
l', 'sad', 'angry', 'fearful', 'happy', 'disgust', 'happy', 'happy', 'surprised', 'sad', 'sad', 'sad', 'disgust', 'disgust', 'angry', 'surprised', 'surprised', 'disgust', 'disgust', 'disgust', 'disgust', 'sad',
'surprised', 'happy', 'sad', 'disgust', 'sad']
['disgust' 'happy' 'sad' 'surprised' 'surprised' 'fearful' 'sad' 'sad'
 'happy' 'fearful' 'angry' 'sad' 'angry' 'disgust' 'disgust' 'disgust'
 'disgust' 'sad' 'angry' 'angry' 'disgust' 'sad' 'fearful' 'disgust'
 'angry' 'surprised' 'happy' 'happy' 'sad' 'angry' 'sad' 'disgust' 'angry'
 'sad' 'sad' 'angry' 'sad' 'sad' 'sad' 'disgust' 'angry' 'angry'
 'surprised' 'happy' 'disgust' 'disgust' 'surprised' 'surprised' 'fearful'
 'disgust' 'sad' 'surprised' 'fearful' 'happy' 'sad' 'fearful' 'happy'
 'disgust' 'sad' 'disgust' 'happy' 'disgust' 'angry' 'angry' 'fearful'
 'happy' 'angry' 'disgust' 'surprised' 'angry' 'angry' 'angry' 'angry'
 'sad' 'fearful' 'surprised' 'surprised' 'angry' 'surprised' 'fearful'
 'fearful' 'sad' 'surprised' 'happy' 'fearful' 'sad' 'angry' 'fearful'
 'happy' 'disgust' 'happy' 'happy' 'surprised' 'fearful' 'sad' 'sad'
 'disgust' 'disgust' 'angry' 'surprised' 'surprised' 'disgust' 'angry'
 'disgust' 'happy' 'sad' 'surprised' 'happy' 'sad' 'disgust' 'fearful']
Accuracy: 80.18%
```

Figure 6-1: Accuracy

6.2: Graphical representation of well and mis-classified data

This is the graphical representation of the output of our model. Here, we can see that the correctly classified data in each emotion are highlighted in blue colour and the mis-classified data is represented with yellow colour.

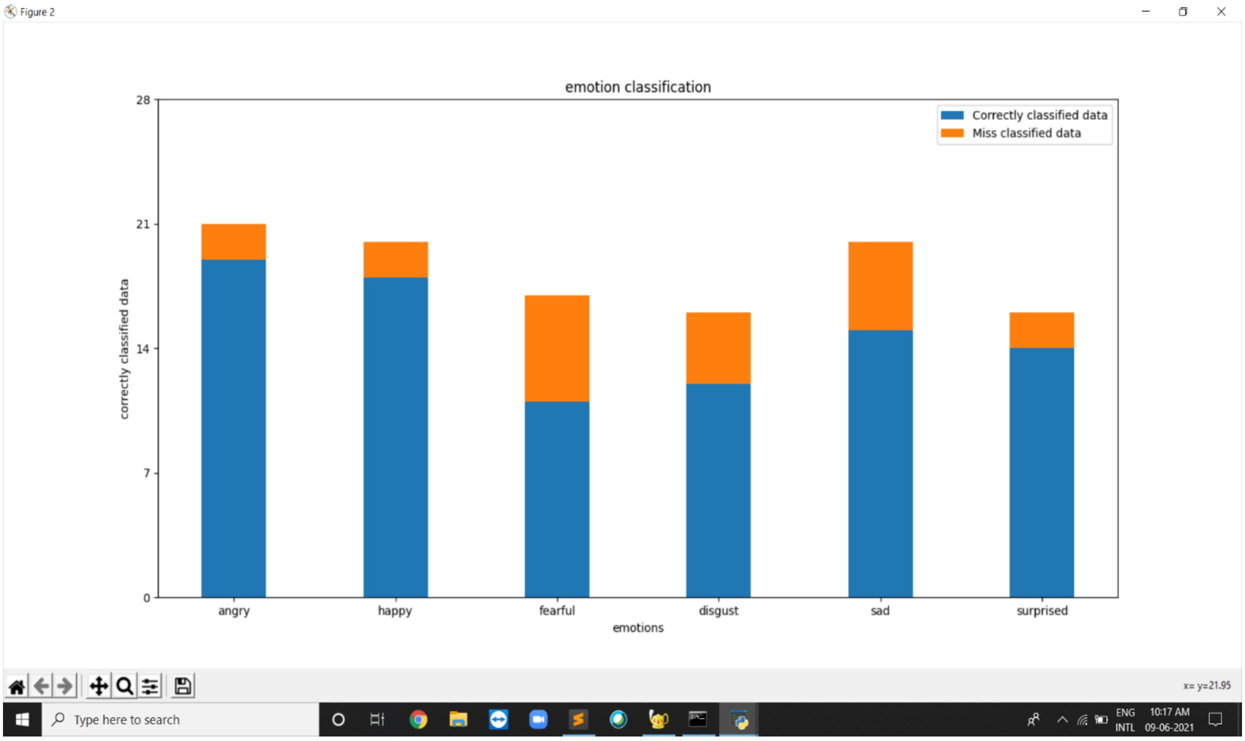
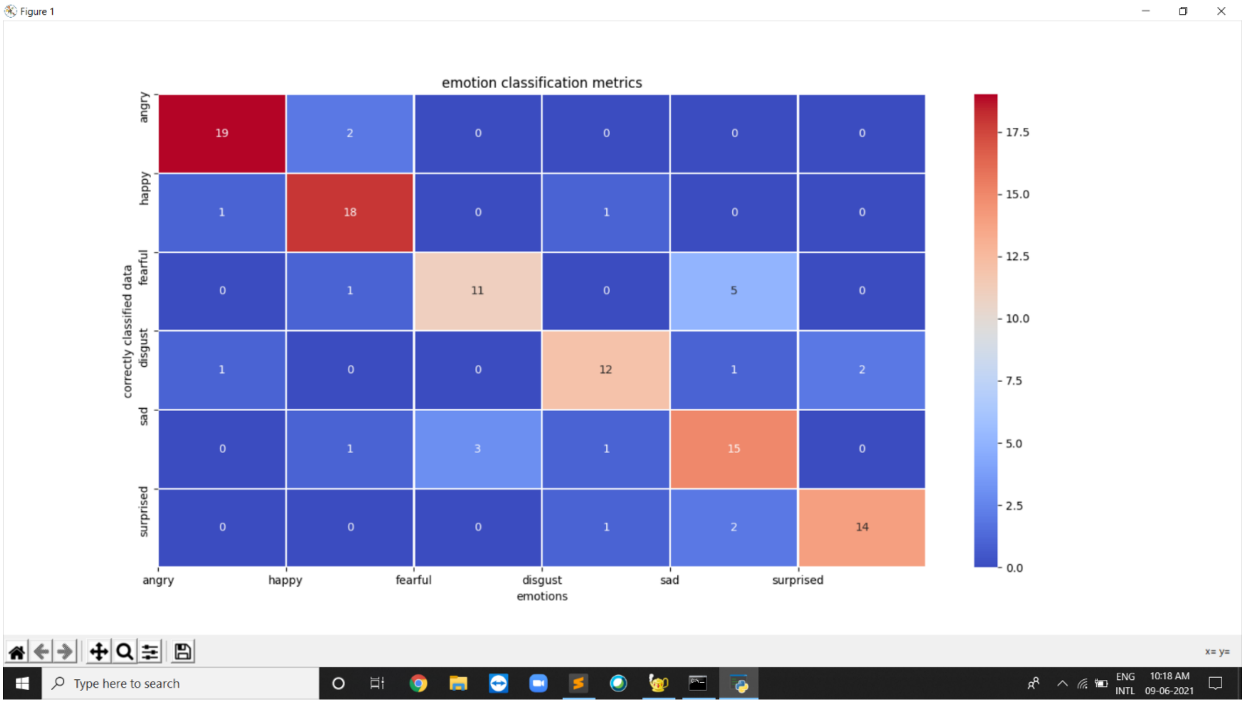


Figure 6-2: Graphical Representation

6.3: Confusion matrix

This is the visual representation of the confusion matrix. Here, each row and column represent an emotion. Correctly classified emotions can be seen at diagonals and the mis-classified emotions can be seen at a position other than diagonal. In the above matrix we can see that 19 audio files are correctly classified as angry out of 21 audio files and the remaining 2 audio files are mis-classified.



	angry	happy	fearful	disgust	sad	surprised
angry	19	2	0	0	0	0
happy	1	18	0	1	0	0
fearful	0	1	11	0	5	0
disgust	1	0	0	12	1	2
sad	0	1	3	1	15	0
surprised	0	0	0	1	2	14

Figure 6-3:Emotion Classification Matrix

CHAPTER - 7

CONCLUSION

7.1: Conclusion of project

By this project, we presented how we can use Machine learning to classify the hidden emotion from speech data and some understanding of emotion through voice. In this project, we learned how to categorize emotions through speech. We used an MLP Classifier for this and the librosa module to extract features from it. The Accuracy of our delivered model is 79.36%. This system has a wide range of applications in a variety of setups like Call Centre for complaints, Psychology, Investigation etc..

CHAPTER -8

SCOPE OF FUTURE WORKS

This project can be further improved in way where the accuracy of the model will be increased. We can achieve this by using different types of classifiers and features. This model can be used in music applications like Ganaa, Spotify etc., which plays songs according to the users emotions. This model can also be used in places like call centres, during investigation and for medical purposes, psychology etc...

CHAPTER - 9

REFERENCES

1. H. Cao, R. Verma, and A. Nenkova, “Speaker-sensitive emotion recognition via ranking: Studies on acted and spontaneous speech,” Comput. Speech Lang., vol. 28, no. 1, pp. 186–202, Jan. 2015.

2. S. S. Narayanan, “Toward detecting emotions in spoken dialogs,” IEEE Trans. Speech Audio Process., vol. 13, no. 2, pp. 293–303, Mar. 2005.

3. E. M. Albornoz, D. H. Milone, and H. L. Rufiner, “Spoken emotion recognition using hierarchical classifiers,” Comput. Speech Lang., vol. 25, no. 3, pp. 556–570, Jul. 2011.

4. C.-C. Lee, E. Mower, C. Busso, S. Lee, and S. Narayanan, “Emotion recognition using a hierarchical binary decision tree approach,” Interspeech, vol. 53, pp. 320–323, 2009.

5. M. M. H. El Ayadi, M. S. Kamel, and F. Karray, “Speech Emotion Recognition using Gaussian Mixture Vector Autoregressive Models,” in 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP ’07, 2007, vol. 4, pp. IV–957–IV–960.

6. 2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)

7. https://towardsdatascience.com/how-i-understood-what-features-to-consider-while-training-audio-files-eedfb6e9002b

8. https://scikitlearn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html

9. https://scikit-learn.org/stable/modules/neural_networks_supervised.html

10. https://www.hindawi.com/journals/mpe/2014/749604