

VISVESVARAYA TECHNOLOGICAL UNIVERSITY



BELAGAVI – 590018, Karnataka

INTERNSHIP REPORT

ON

“VOICE CLASSIFICATION USING ML”

Submitted in partial fulfilment for the award of degree(18ECI85)

**BACHELOR OF ENGINEERING IN
ELECTRONICS AND COMMUNICATION**

Submitted by:

Bharghavi S

1AT19EC020



Conducted at
COMPSOFT TECHNOLOGIES



ATRIA INSTITUTE OF TECHNOLOGY
Department of Electronics and Communication Engineering

Accredited by NBA

Adjacent Bangalore Baptist Hospital, Bangalore - 560024

ATRIA INSTITUTE OF TECHNOLOGY
Department of Electronics and Communication Engineering
Adjacent Bangalore Baptist Hospital, Bangalore- 560024



CERTIFICATE

This is to certify that the Internship titled “**Voice Classification using ML**” carried out by **Ms. Bharghavi S**, a bonafide student of Atria Institute of Technology, in partial fulfillment for the award of **Bachelor of Engineering, in Electronics and Communication engineering** under Visvesvaraya Technological University, Belagavi, during the year 2022-2023. It is certified that all corrections/suggestions indicated have been incorporated in the report.

The project report has been approved as it satisfies the academic requirements in respect of Internship prescribed for the course Internship / Professional Practice (18ECI85)

Signature of Guide

Signature of HOD

Signature of Principal

External Viva:

Name of the Examiner

Signature with Date

1) _____

2) _____

D E C L A R A T I O N

I, **Bharghavi S**, final year student of Electronics and Communication Engineering, Atria Institute of Technology-560024, declare that the Internship has been successfully completed, in **COMPSOFT TECHNOLOGIES**. This report is submitted in partial fulfillment of the requirements for award of Bachelor Degree in Electronics and Communication Engineering, during the academic year 2022-2023.

Date : _____ :

Place : Bangalore

USN : 1AT19EC020

NAME : Bharghavi S

OFFER LETTER



Date: 23rd August, 2022

Name: Bharghavi S

USN: 1AT19EC020

Dear Student,

We would like to congratulate you on being selected for the **Machine Learning With Python (Research Based)** Internship position with **Compsoft Technologies**, effective Start Date **23rd August, 2022**. All of us are excited about this opportunity provided to you!

This internship is viewed as being an educational opportunity for you, rather than a part-time job. As such, your internship will include training/orientation and focus primarily on learning and developing new skills and gaining a deeper understanding of concepts of **Machine Learning With Python (Research Based)** through hands-on application of the knowledge you learn while you train with the senior developers. You will be bound to follow the rules and regulations of the company during your internship duration.

Again, congratulations and we look forward to working with you!

Sincerely,

Nithin K. S

Project Manager

COMPSOFT TECHNOLOGIES

No. 363, 19th main road,

1st Block Rajajinagar

Bangalore - 560010

ACKNOWLEDGEMENT

This Internship is a result of accumulated guidance, direction and support of several important persons. We take this opportunity to express our gratitude to all who have helped us to complete the Internship.

We express our sincere thanks to our principal, for providing us adequate facilities to undertake this Internship.

We would like to thank our Head of Dept – branch code, for providing us an opportunity to carry out Internship and for his valuable guidance and support.

We would like to thank our (Lab assistant name) Software Services for guiding us during the period of internship.

We express our deep and profound gratitude to our guide, Guide name, Assistant/Associate Prof, for her keen interest and encouragement at every step in completing the Internship.

We would like to thank all the faculty members of our department for the support extended during the course of Internship.

We would like to thank the non-teaching members of our dept, for helping us during the Internship.

Last but not the least, we would like to thank our parents and friends without whose constant help, the completion of Internship would have not been possible.

NAME: Bharghavi S
USN:1AT19EC020

ABSTRACT

Voice classification using ML is a system where we determine emotions from input audio. Emotion recognition from speech signals is an important but challenging component of Human Computer Interaction. In voice classification using ML, many techniques have been utilized to extract emotions from signals, including many well-established speech analysis and classification techniques. Machine Learning techniques have been recently proposed as an alternative to traditional techniques in emotion recognition through speech. Different persons have different emotions and altogether a different way to express it. Speech emotion do have different energies, pitch variations are emphasized if considering different subjects. The voice classification for recognition of speech is based on the Artificial Neural Network (ANN) algorithm which uses different modules for the emotion recognition and the MLP classifier/CNN model/SVM model and LSTM-RNN model is used to differentiate emotions such as happiness, anger, neutral state, sadness etc. The dataset for the voice classification for speech emotion recognition system is the speech samples (RAVDESS and TESS/TORONTO) and the characteristics are extracted from these speech samples using LIBROSA package. The classification performance is based on extracted characteristics. Finally, we can determine the emotion of speech signal. The approach consists of three steps. First, numerical features are extracted from the sound database by using audio feature extractor. Then, feature selection method is used to select the most relevant features. Finally, a machine learning model is trained to recognize seven universal emotions: neutral, calm, happy, sad, angry, fearful, disgust and surprised.

Table of Contents

Sl no	Description	Page no
1	Company Profile	8-9
2	About the Company	10-13
3	Introduction	14-15
4	System Analysis	16-18
5	Requirement Analysis	19-21
6	Design Analysis	22-28
7	Implementation	29-32
8	Snapshots	33-35
9	Conclusion	36-38
10	References	39

CHAPTER 1

COMPANY PROFILE

1. COMPANY PROFILE

A Brief History of Compsoft Technologies

Compsoft Technologies, was incorporated with a goal” To provide high quality and optimal Technological Solutions to business requirements of our clients”. Every business is a different and has a unique business model and so are the technological requirements. They understand this and hence the solutions provided to these requirements are different as well. They focus on clients requirements and provide them with tailor made technological solutions. They also understand that Reach of their Product to its targeted market or the automation of the existing process into e-client and simple process are the key features that our clients desire from Technological Solution they are looking for and these are the features that we focus on while designing the solutions for their clients.

Sarvamoola Software Services. is a Technology Organization providing solutions for all web design and development, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and LINQ. Meeting the ever-increasing automation requirements, Sarvamoola Software Services. specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion and tailor-made software products, designing solutions best suiting client’s requirements.

Compsoft Technologies, strive to be the front runner in creativity and innovation in software development through their well-researched expertise and establish it as an out of the box software development company in Bangalore, India. As a software development company, they translate this software development expertise into value for their customers through their professional solutions.

They understand that the best desired output can be achieved only by understanding the clients demand better. Compsoft Technologies work with their clients and help them to define their exact solution requirement. Sometimes even they wonder that they have completely redefined their solution or new application requirement during the brainstorming session, and here they position themselves as an IT solutions consulting group comprising of high caliber consultants.

They believe that Technology when used properly can help any business to scale and achieve new heights of success. It helps Improve its efficiency, profitability, reliability; to put it in one sentence” Technology helps you to Delight your customers” and that is what we want to achieve.

CHAPTER 2

ABOUT THE COMPANY

2. ABOUT THE COMPANY



Compsoft Technologies is a Technology Organization providing solutions for all web design and development, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and LINQ. Meeting the ever-increasing automation requirements, Compsoft Technologies specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion and tailor-made software products, designing solutions best suiting client's requirements. The organization where they have a right mix of professionals as a stakeholder to help us serve our clients with best of our capability and with at par industry standards. They have young, enthusiastic, passionate and creative Professionals to develop technological innovations in the field of Mobile technologies, Web applications as well as Business and Enterprise solution. Motto of our organization is to "Collaborate with our clients to provide them with best Technological solution hence creating Good Present and Better Future for our client which will bring a cascading a positive effect in their business shape as well". Providing a Complete suite of technical solutions is not just our tag line, it is Our Vision for Our Clients and for us, we strive hard to achieve it.

Products of Compsoft Technologies.

Android Apps

It is the process by which new applications are created for devices running the Android operating system. Applications are usually developed in Java (and/or Kotlin; or other such option) programming language using the Android software development kit (SDK), but other development environments are also available, some such as Kotlin support the exact same Android APIs (and bytecode), while others such as Go have restricted API access.

The Android software development kit includes a comprehensive set of development tools. These include a debugger, libraries, a handset emulator based on QEMU, documentation, sample code, and tutorials. Currently supported development platforms include computers running Linux (any modern desktop Linux distribution), Mac OS X 10.5.8 or later, and Windows 7 or later. As of March 2015, the SDK is not available on Android itself, but software development is possible by using specialized Android applications.

Web Application

It is a client-server computer program in which the client (including the user interface and client-side logic) runs in a web browser. Common web applications include web mail, online

retail sales, online auctions, wikis, instant messaging services and many other functions. web applications use web documents written in a standard format such as HTML and JavaScript, which are supported by a variety of web browsers. Web applications can be considered as a specific variant of client-server software where the client software is downloaded to the client machine when visiting the relevant web page, using standard procedures such as HTTP. The Client web software updates may happen each time the web page is visited. During the session, the web browser interprets and displays the pages, and acts as the universal client for any web application. The use of web application frameworks can often reduce the number of errors in a program, both by making the code simpler, and by allowing one team to concentrate on the framework while another focuses on a specified use case. In applications which are exposed to constant hacking attempts on the Internet, security-related problems can be caused by errors in the program.

Frameworks can also promote the use of best practices such as GET after POST. There are some who view a web application as a two-tier architecture. This can be a “smart” client that performs all the work and queries a “dumb” server, or a “dumb” client that relies on a “smart” server. The client would handle the presentation tier, the server would have the database (storage tier), and the business logic (application tier) would be on one of them or on both. While this increases the scalability of the applications and separates the display and the database, it still doesn’t allow for true specialization of layers, so most applications will outgrow this model. An emerging strategy for application software companies is to provide web access to software previously distributed as local applications. Depending on the type of application, it may require the development of an entirely different browser-based interface, or merely adapting an existing application to use different presentation technology. These programs allow the user to pay a monthly or yearly fee for use of a software application without having to install it on a local hard drive. A company which follows this strategy is known as an application service provider (ASP), and ASPs are currently receiving much attention in the software industry.

Security breaches on these kinds of applications are a major concern because it can involve both enterprise information and private customer data. Protecting these assets is an important part of any web application and there are some key operational areas that must be included in the development process. This includes processes for authentication, authorization, asset handling, input, and logging and auditing. Building security into the applications from the beginning can be more effective and less disruptive in the long run.

Web design

It is encompassing many different skills and disciplines in the production and maintenance of websites. The different areas of web design include web graphic design; interface design; authoring, including standardized code and proprietary software; user experience design; and

search engine optimization. The term web design is normally used to describe the design process relating to the front-end (client side) design of a website including writing mark up. Web design partially overlaps web engineering in the broader scope of web development. Web designers are expected to have an awareness of usability and if their role involves creating markup then they are also expected to be up to date with web accessibility guidelines. Web design partially overlaps web engineering in the broader scope of web development.

Departments and services offered

Compsoft Technologies plays an essential role as an institute, the level of education, development of student's skills is based on their trainers. If you do not have a good mentor then you may lag in many things from others and that is why we at Compsoft Technologies gives you the facility of skilled employees so that you do not feel unsecured about the academics. Personality development and academic status are some of those things which lie on mentor's hands. If you are trained well then you can do well in your future and knowing its importance of Compsoft Technologies always tries to give you the best.

They have a great team of skilled mentors who are always ready to direct their trainees in the best possible way they can and to ensure the skills of mentors we held many skill development programs as well so that each and every mentor can develop their own skills with the demands of the companies so that they can prepare a complete packaged trainee.

Services provided by Compsoft Technologies.

- Core Java and Advanced Java
- Web services and development
- Dot Net Framework
- Python
- Selenium Testing
- Conference / Event Management Service
- Academic Project Guidance
- On The Job Training
- Software Training

CHAPTER 3

INTRODUCTION

3. INTRODUCTION

Introduction to ML

The field of study that gives computers the ability to learn without being explicitly programmed is Machine Learning.

- Machine learning is programming computers to optimize a performance criterion using example data or past experience. We have a model defined up to some parameters, and learning is the execution of a computer program to optimize the parameters of the model using the training data or past experience. The model may be predictive to make predictions in the future, or descriptive to gain knowledge from data.
- Machine learning implementations are classified into four major categories, depending on the nature of the learning “signal” or “response” available to a learning system.
- Supervised learning: Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs.
- Unsupervised learning: is a type of machine learning algorithm used to draw inferences from datasets consisting of input data without labeled responses.
- Reinforcement learning: is the problem of getting an agent to act in the world so as to maximize its rewards.
- Semi-supervised learning: is an approach to machine learning that combines small labeled data with a large amount of unlabeled data during training. Semi-supervised learning falls between unsupervised learning and supervised learning.

Categorizing based on required Output

- Classification: When inputs are divided into two or more classes, the learner must produce a model that assigns unseen inputs to one or more of these classes.
- Regression: Which is also a supervised problem, A case when the outputs are continuous rather than discrete.
- Clustering: When a set of inputs is to be divided into groups. (Unsupervised)

Problem Statement

Built a python application that analyses the sentiment behind the tone of the voice and predicts the Sentiment involved, you can use an open-source dataset for the same.

CHAPTER 4

SYSTEM ANALYSIS

4. SYSTEM ANALYSIS

1. Existing System

First, n -gram models use the previous n words as context to try to figure out a given word. So, for instance, if it looks at the previous two words, we call it a bi-gram system and $n=2$. While higher values for n lead to greater accuracy since the computer has more context to look at, it simply isn't practical to use a large number for n because the computational overhead is too much.

Other algorithm is the Hidden Markov Model (HMM), which basically just goes in the opposite direction. Instead of looking backwards, HMMs look forwards. Without including any knowledge of the previous state—in our case, the words that came before the word in question—a HMM algorithm uses probabilities and statistics to guess what comes next. The “hidden” part means that we can include information about the target word that's not obviously apparent, such as a part-of-speech tag (verb, noun, etc.).

2. Proposed System

This one solves a very specific problem with training speech recognition models. Remember that ML models learn from data; for instance, an image classifier can tell the difference between cats and dogs after we feed it pictures that we label as either “cat” or “dog.” For speech recognition, this amounts to feeding it hours upon hours of audio and the corresponding ground-truth transcripts that were written by a human transcriptionist.

Most neural networks “feed-forward,” meaning that nodes only send their output to nodes that are further down in the chain, the specific algorithms that we use for speech processing work a little differently. Dubbed the LSTM- Recurrent Neural Network (RNN)/CNN, these algorithms are ideal for sequential data like speech because they're able to “remember” what came before and use their previous output as input for their next move. Since words generally appear in the context of a sentence, knowing what came before and recycling that information into the next prediction goes a long way towards accurate speech recognition.

3. Objective of the System

The speech emotion recognition system is implemented as a Machine Learning (ML) model. The steps of implementation are comparable to any other ML project, with additional fine-tuning procedures to make the mode outperform previous state-of-the-art methods in assigning data to a minimum of one among 4 emotion categories (i.e., angry, happy, sad and neutral).

Choosing to follow the lexical features would require a transcript of the speech which might further require a further step of text extraction from speech if one wants to predict emotions from real-time audio. Similarly, going forward with analyzing visual features would require the surplus to the video of the conversations which could not be feasible in every case while the analysis on the acoustic features is often wiped out real-time while the conversation is happening as we'd just need the audio data for accomplishing our task. Hence, we elect to analyze the acoustic features during this work. The field of study is termed as Speech Processing and consists of three components: Speaker Identification, Speech Recognition, Speech Emotion Detection.

An emotion one out of a delegated set of emotions is identified with each unit of language (word or phrase or utterance) that was spoken, with the precise start of every such unit determined within the continual acoustic signal. A striking nature unique to humans is that the ability to change conversations supported the spirit of the speaker and also the listener

CHAPTER 5

REQUIREMENT ANALYSIS

5. REQUIREMENT ANALYSIS

Hardware Requirement Specification

- i5 8th gen processor
- 2 GB Graphic Card
- 8GB RAM

Software Requirement Specification

Python programming language is used. Python is an interpreter, high-level, and general- purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant indentation. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

CNN- Model

- Python 3.7(For Streamlit app,)/3.8
- Google Colab
- Librosa, os, glob,
- Matplotlib
- Numpy
- Joblib
- Sklearn
- Tensorflow-keras

LSTM-RNN Model

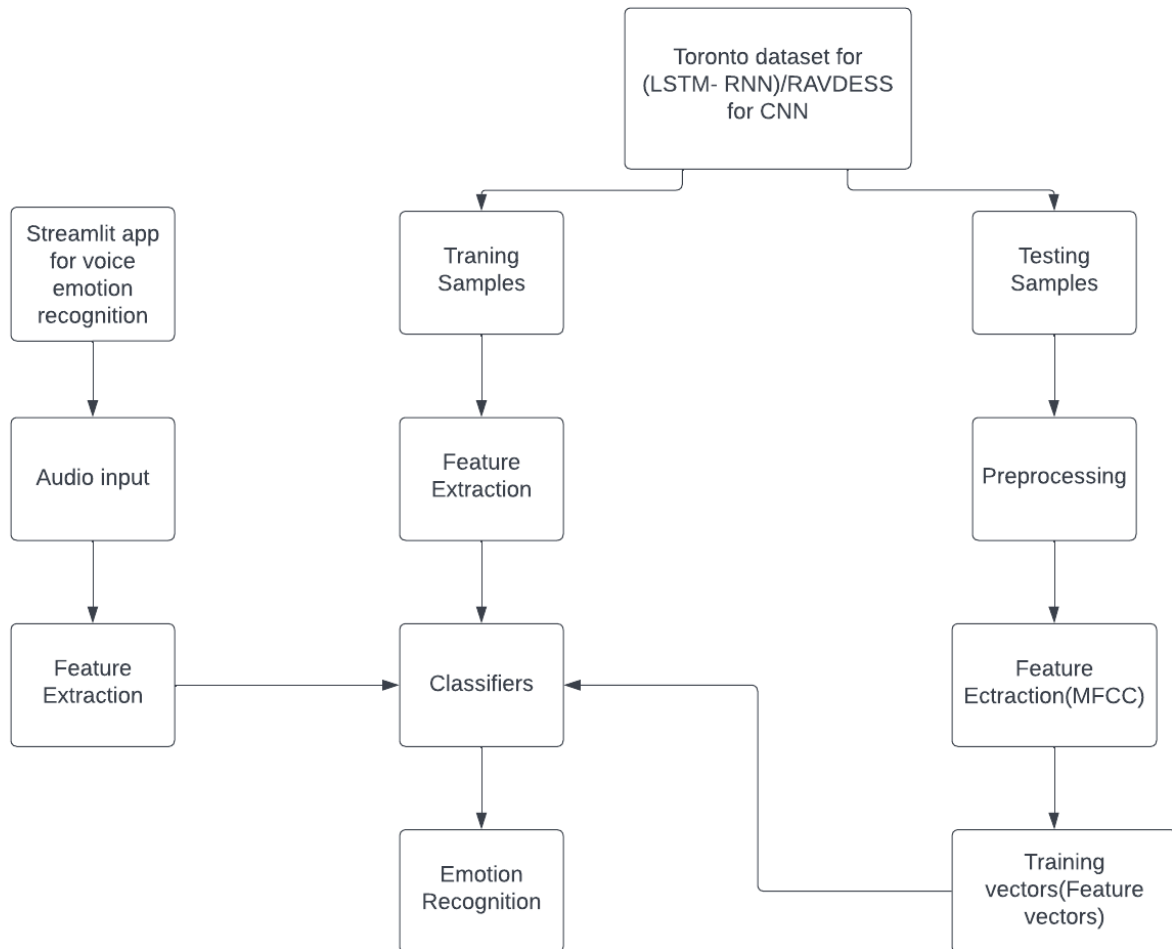
- Python 3.7(For Streamlit app,)/3.8
- Google Colab
- Librosa, os, glob, PyAudio
- Matplotlib
- Numpy
- Joblib
- Sklearn
- Tensorflow-keras
- Wave

- Python is a high-level, interpreted scripting language developed in the late 1980s by Guido van Rossum at the National Research Institute for Mathematics and Computer Science in the Netherlands. The initial version was published at in 1991, and version 1.0 was released in 1994. The Latest version of python is 3.11. Python has huge number of modules for covering every aspect of programming. These modules are easily available for use hence making Python an extensible language. Python is a scalable programming language because it provides an improved structure for supporting large programs than shell-scripts.
- Librosa is a Python package for music and audio analysis. Librosa is basically used when we work with audio data like in music generation (using LSTM's), Automatic Speech Recognition. It provides the building blocks necessary to create the music information retrieval systems.
- Scikit-learn is a free machine learning library for Python. It features various algorithms like support vector machine, random forests, and k-neighbors, and it also supports Python numerical and scientific libraries like NumPy and SciPy.
- NumPy, which stands for Numerical Python, is a library consisting of multidimensional array objects and a collection of routines for processing those arrays. Using NumPy, mathematical and logical operations on arrays can be performed. We used NumPy as it provides 50x faster an array object (called Nd array) for our sound file.
- Pyaudio provides Python bindings for Port Audio, the cross-platform audio I/O library. With Pyaudio, you can easily use Python to play and record audio on a variety of platforms". We are using Pyaudio to get the audio from the user.
- The OS module in Python provides functions for interacting with the operating system. OS comes under Python's standard utility modules. This module provides a portable way of using operating system dependent functionality. The `*os*` and `*os.path*` modules include many functions to interact with the file system
- TensorFlow is a free and open-source software library for machine learning and artificial intelligence. It can be used across a range of tasks but has a particular focus on training and inference of deep neural networks.

CHAPTER 6

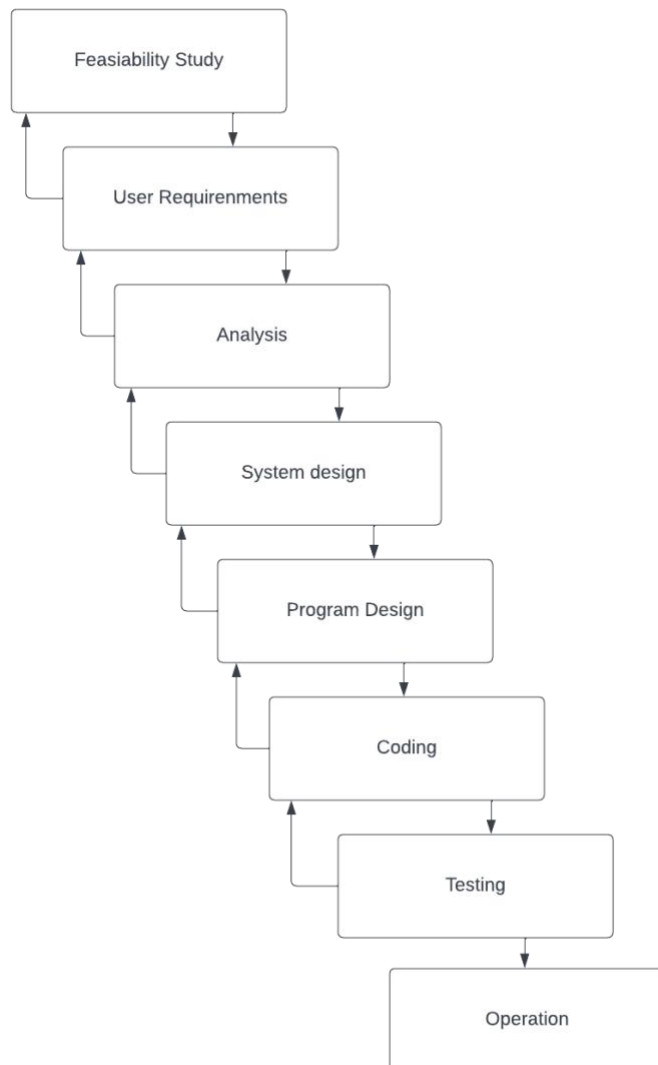
DESIGN ANALYSIS

6. DESIGN & ANALYSIS



The above flowchart shows the work flow of the project. RAVDESS dataset for CNN Model is used for training and testing of the project which is divided in 75% and 25% respectively. In testing, the dataset goes through pre-processing and then feature extraction is performed like MFCC. Then classifier MLP work on this extracted feature and give appropriate output. The same is also worked out for SVM, CNN and LSTM-RNN models, where the accuracy is high compared to when using an MLP classifier. In training, the features are extracted and then classifier uses it to determine the underlying emotion. The project works in such a way that we use audio testing using Streamlit app and then features are extracted. The classifiers for the respective models use the features to determine emotion.

DATA FLOW DIAGRAM MECHANISM



Waterfall approach was first SDLC Model to be used widely in Software Engineering to ensure success of the project. In "The Waterfall" approach, the whole process of software development is divided into separate phases. In this Waterfall model, typically, the outcome of one phase acts as the input for the next phase sequentially. The following illustration is a representation of the different phases of the Waterfall Model. The sequential phases in Waterfall model are :

- Requirement Gathering and analysis – All possible requirements of the system to be developed are captured in this phase and documented in a requirement specification document.
- System Design – The requirement specifications from first phase are studied in this phase and the system design is prepared. This system design helps in specifying hardware and system requirements and helps in defining the overall system architecture.

- Implementation – With inputs from the system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality, which is referred to as Unit Testing.
- Integration and Testing – All the units developed in the implementation phase are integrated into a system after testing of each unit. Post integration the entire system is tested for any faults and failures.
- Deployment of system – Once the functional and non-functional testing is done; the product is deployed in the customer environment or released into the market.
- Maintenance – There are some issues which come up in the client environment. To fix those issues, patches are released. Also, to enhance the product some better versions are released. Maintenance is done to deliver these changes in the customer environment. All these phases are cascaded to each other in which progress is seen as flowing steadily downwards (like a waterfall) through the phases. The next phase is started only after the defined set of goals are achieved for previous phase and it is signed off, so the name "Waterfall Model". In this model, phases do not overlap

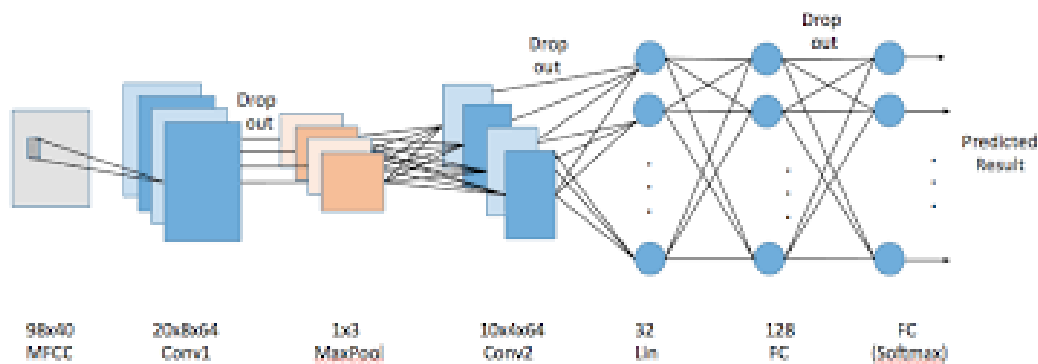
ALGORITHM

Multi-layer perceptron (MLP) is a supplement of feed forward neural network. It consists of three types of layers—the input layer, output layer and hidden layer, as shown in Fig. below. The input layer receives the input signal to be processed. The required task such as prediction and classification are performed by the output layer. An arbitrary number of hidden layers that are placed in between the input and output layer are the true computational engine of the MLP. Similar to a feed forward network in a MLP the data flows in the forward direction from input to output layer. The neurons in the MLP are trained with the back propagation learning algorithm. MLPs are designed to approximate any continuous function and can solve problems which are not linearly separable. The major use cases of MLP are pattern classification, recognition, prediction and approximation.

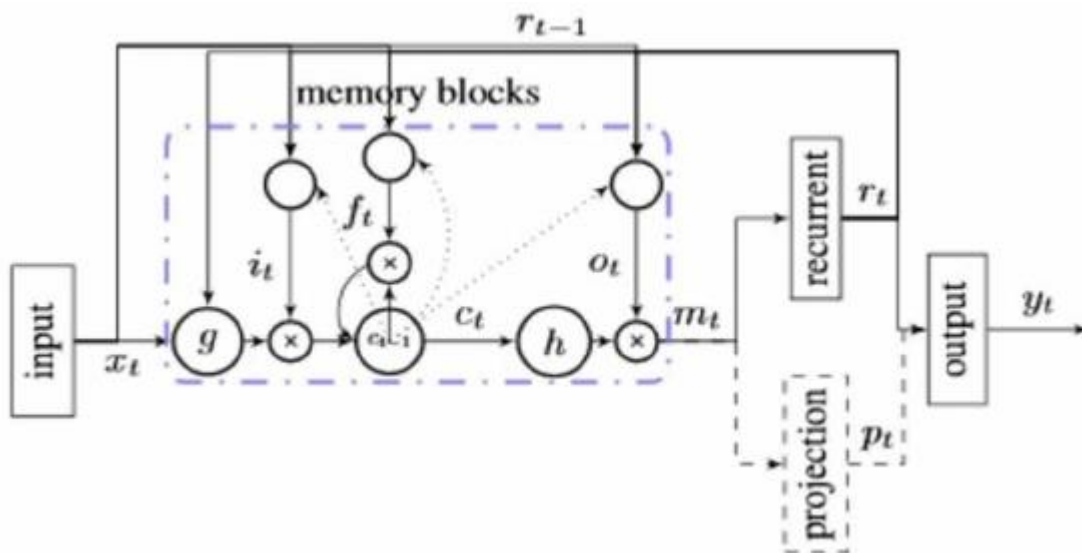
A multilayer perceptron (MLP) is a class of feedforward artificial neural network (ANN)-CNN/LSTM-RNN. The term MLP is used ambiguously, sometimes loosely to any feedforward ANN, sometimes strictly to refer to networks composed of multiple layers of perceptron's (with threshold activation). Multilayer perceptron's are sometimes colloquially referred to as "vanilla" neural networks, especially when they have a single hidden layer.

The multilayer perceptron is applied for supervised learning problems. The multi-layer perceptron is also issued for the purpose of classification. The MLP is made to train on the given dataset. The training phase enables the MLP to learn the correlation between the set of inputs and outputs. During training, the MLP adjusts model parameters such as weights and biases in order to minimize the error. The MLP uses Back propagation, to make weight and bias adjustments relative to the error. The model parameter has been set with hidden layer 300, iteration 500, which have found to be best by grid search.

CNN-MODEL



LSTM MODEL



FEATURE USED

Mel Frequency Cepstral Co-efficient (MFCC)

MFCC features represent phonemes (distinct units of sound) as the shape of the vocal tract (which is responsible for sound generation) is manifest in them [4]. The MFCC technique aims to develop the features from the audio signal which can be used for detecting the phones in the speech. MFCC is calculated using this equation: -

$$\hat{C}_n = \sum_{k=1}^K \left(\log \hat{S}_k \right) \cos \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right]$$

DATASET

The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) has been used. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) contains 7356 files (total size: 24.8 GB). The database 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. All conditions are available in three modality formats: Audioonly (16bit, 48kHz .wav), Audio-Video (720p H.264, AAC 48kHz, .mp4), and Video-only (no sound). The utterances of the speech are kept constant by speaking only 2 statements of equal lengths. Each of the RAVDESS files has a unique filename. The filename consists of a 7-part numerical identifier (e.g., 02-01-06-01-02-01-12.mp4). These identifiers define the stimulus characteristics. FILENAME IDENTIFIERS

- Modality (01 = full-AV, 02 = video-only, 03 = audio-only).
- Vocal channel (01 = speech, 02 = song).
- Emotion (01 = neutral, 02 = calm, 03 = happy, 04 = sad, 05 = angry, 06 = fearful, 07 = disgust, 08 = surprised).
- Emotional intensity (01 = normal, 02 = strong). NOTE: There is no strong intensity for the 'neutral' emotion.
- Statement (01 = "Kids are talking by the door", 02 = "Dogs are sitting by the door").
- Repetition (01 = 1st repetition, 02 = 2nd repetition).

- Actor (01 to 24. Odd numbered actors are male, even numbered actors are female).

FILENAME EXAMPLE: 02-01-06-01-02-01-12.mp4

- Video-only (02)
- Speech (01)
- Fearful (06)
- Normal intensity (01)
- Statement “dogs” (02)
- 1st Repetition (01)
- 12th Actor (12)
- Female, as the actor ID number is even

CHAPTER 7

IMPLEMENTATION

7. IMPLEMENTATION

Implementation is the stage where the theoretical design is turned into a working system. The most crucial stage in achieving a new successful system and in giving confidence on the new system for the users that it will work efficiently and effectively.

The system can be implemented only after thorough testing is done and if it is found to work according to the specification. It involves careful planning, investigation of the current system and its constraints on implementation, design of methods to achieve the changeover and an evaluation of change over methods as a part from planning.

Two major tasks of preparing the implementation are education and training of the users and testing of the system. The more complex the system being implemented, the more involved will be the system analysis and design effort required just for implementation.

The implementation phase comprises of several activities. The required hardware and software acquisition is carried out. The system may require some software to be developed. For this, programs are written and tested. The user then changes over to his new fully tested system and the old system is discontinued.

TESTING

The testing phase is an important part of software development. It is the Information system will help in automate process of finding errors and missing operations and also a complete verification to determine whether the objectives are met and the user requirements are satisfied. Software testing is carried out in three steps:

1. The first includes unit testing, where in each module is tested to provide its correctness, validity and also determine any missing operations and to verify whether the objectives have been met. Errors are noted down and corrected immediately.
2. Unit testing is the important and major part of the project. So errors are rectified easily in particular module and program clarity is increased. In this project entire system is divided into several modules and is developed individually. So unit testing is conducted to individual modules.
3. The second step includes Integration testing. It need not be the case, the software whose modules when run individually and showing perfect results, will also show perfect results when run as a whole.

The speech emotion recognition system is implemented as a Machine Learning (ML) model. The steps of implementation are comparable to any other ML project, with additional fine-tuning procedures to make the model function better.

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.

The second step, called feature engineering, is a collection of several machine learning tasks that are executed over the collected data. These procedures address the several data representation and data quality issues. The third step is often considered the core of an ML project where an algorithmic based model is developed. This model uses an ML algorithm to learn about the data and train itself to respond to any new data it is exposed to.

The final step is to evaluate the functioning of the built model. Very often, developers repeat the steps of developing a model and evaluating it to compare the performance of different algorithms. Comparison results help to choose the appropriate ML algorithm most relevant to the problem. This project is completely based on machine learning and deep learning where we train the models with RAVDESS Dataset/TORONTO which consists of audio files which are labeled with basic emotions.

After the feature extraction, we need to make the system knows about the feature for instance, we are using only “angry”, “sad”, “neutral”, “happy” emotions in our system. Then the step is Defining the file which is what emotion then training the machine using each feature extracted to the known emotion and testing, where we are using 75 % of data for training and 25 % for the testing by splitting.

At first loading the data from a folder which can be done using python library glob and getting base name using os library as we know RAVDEES dataset is made such a way that emotion on 2nd base so declaring X for feature and y for emotion.

X is obtained from “extract feature” and y obtained using “base name” after splitting the base name and we know the “base name” refers to what emotion. For testing the input audio, “app.py” has been created where we are using “pyaudio” module to take voice and we add some noise to make the feature extraction better and it is then passed to extract sound feature and where MLP classifier predicts the emotion and We have added the extended feature that give translation of voice to text using “speech Recognition” module. which need to installed and imported.

After testing a number of input audios, the compiler gives warning of maximum iteration overhead. With this information, we increase the number of iterations. As well as during testing you can change the splitting model of training and testing. As the number of iteration increase, the time for running this program also increases exponentially. The testing of live audio can be emphasized by varying pitch and energy. For testing, the internet of the device should be available. As the system uses Google Streamlit app, it will show result when the system has internet access. The neurons in ML uses random weights, therefore there is possibility of accuracy of audio to vary.

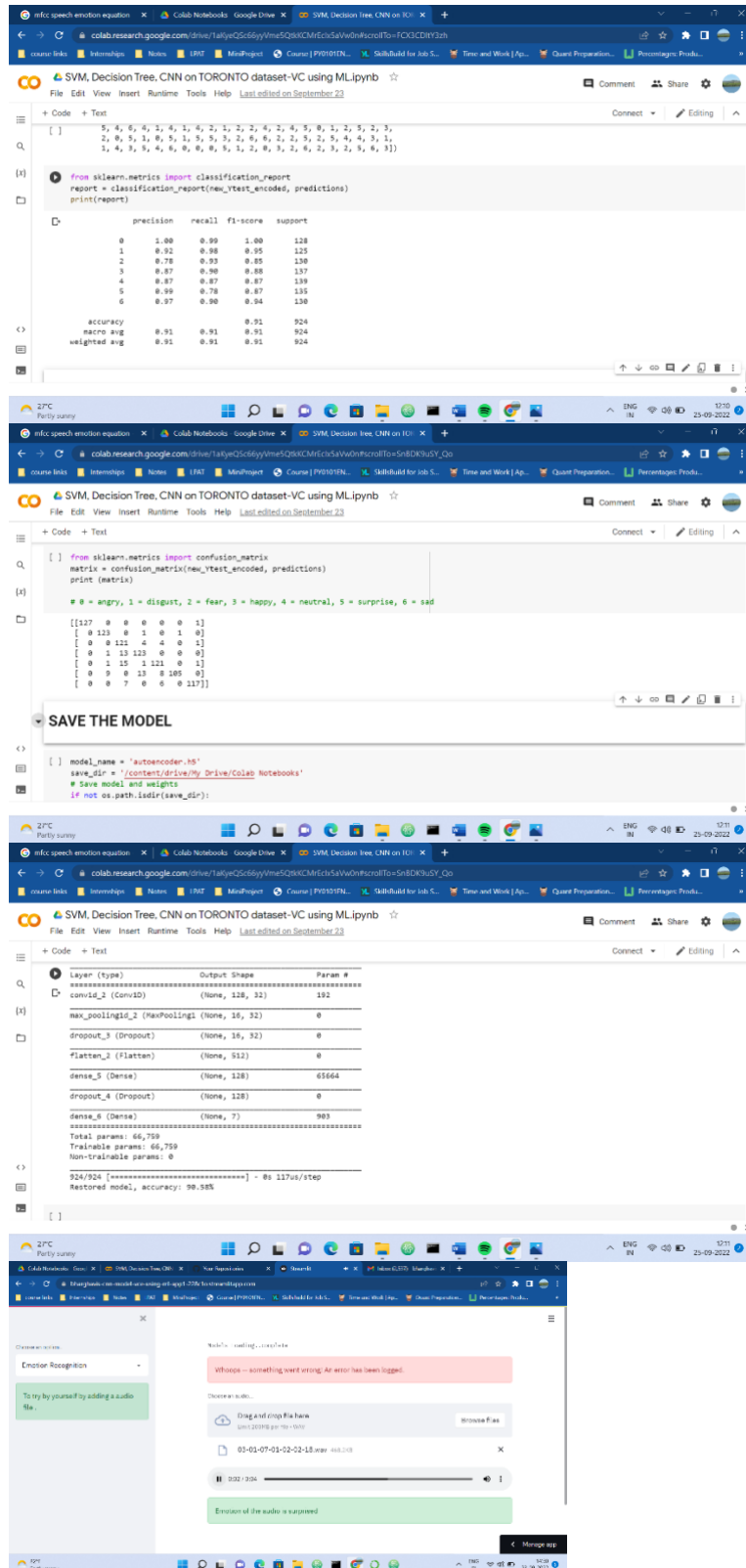
The evaluation of the speech emotion recognition system is based on the level of naturalness of the database which is used as an input to the speech emotion recognition system. The project speech emotion recognition was able to understand and detect the underlying emotion of speaker. The emotion is categorized in four types – anger, happy, neutral and sad. We used 75 % data for training and 25% for testing. The result had frequently showed anger, happy emotion than sad and neutral emotions. We had used grid confusion matrix with MLP classifier. After adding more data set of sad and neutral audios, the system was able to detect the emotion to a satisfactory level.

CHAPTER 8

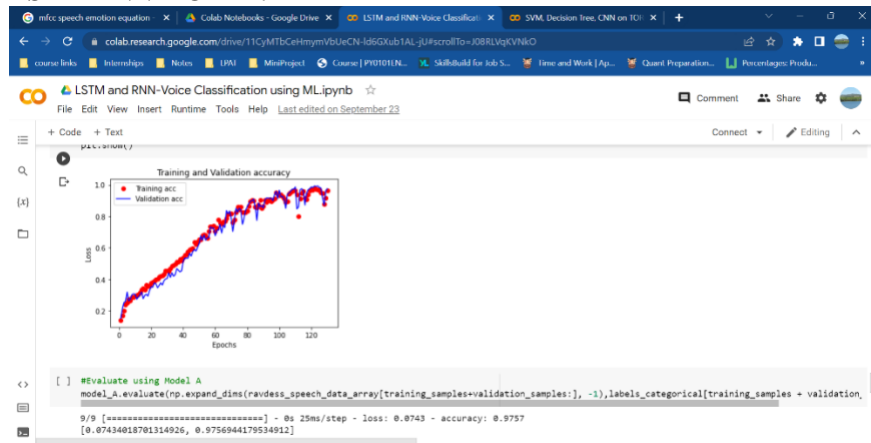
SNAPSHOTS

8. SNAPSHOTS

CNN MODEL:



LSTM-RNN MODEL:



```
[ ] test_point=extract_mfcc(wav_filepath)
test_point=np.reshape(test_point, newshape=(1,40,1))
predictions=model_A.predict(test_point)
print(emotions[np.argmax(predictions[0])+1])

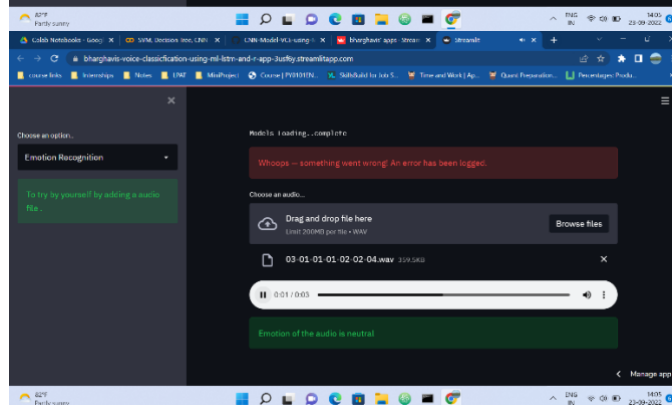
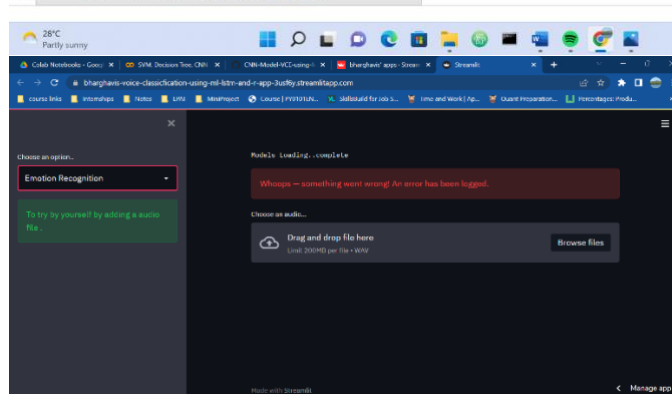
[ ] predict('/content/drive/MyDrive/Colab Notebooks/archive (1)/Actor_03-01-01-01-01-03.wav')
neutral

[ ] predict('/content/drive/MyDrive/Colab Notebooks/archive (1)/Actor_16-03-01-03-01-01-02-16.wav')
happy

[ ] predict('/content/drive/MyDrive/Colab Notebooks/archive (1)/Actor_21-03-01-02-01-02-02-21.wav')
calm

[ ] model_A.save('mymodel.h5')

[ ] model=tf.keras.models.load_model('mymodel.h5')
```



CHAPTER 9

CONCLUSION

9. CONCLUSION

The package was designed in such a way that future modifications can be done easily. The following conclusions can be deduced from the development of the project:

- ❖ Automation of the entire system improves the efficiency
- ❖ It provides a friendly graphical user interface which proves to be better when compared to the existing system.
- ❖ It gives appropriate access to the authorized users depending on their permissions.
- ❖ It effectively overcomes the delay in communications.
- ❖ Updating of information becomes so easier
- ❖ System security, data security and reliability are the striking features.
- ❖ The System has adequate scope for modification in future if it is necessary.

Through this project, we showed how we can leverage Machine learning to obtain the underlying emotion from speech audio data and some insights on the human expression of emotion through voice. This system can be employed in a variety of setups like Call Centre for complaints or marketing, in voice-based virtual assistants or chatbots, in linguistic research, etc. [14]. A few possible steps that can be implemented to make the models more robust and accurate are the following: -

- An accurate implementation of the pace of the speaking can be explored to check if it can resolve some of the deficiencies of the model.
- Figuring out a way to clear random silence from the audio clip.
- Exploring other acoustic features of sound data to check their applicability in the domain of speech emotion recognition. These features could simply be some proposed extensions of MFCC like RAS-MFCC or they could be other features entirely like LPCC, PLP or Harmonic cepstrum.
- Following lexical features-based approach towards SER and using an ensemble of the lexical and acoustic models. This will improve the accuracy of the system because in some cases the expression of emotion is contextual rather than vocal.
- Adding more data volume either by other augmentation techniques like time-shifting or speeding up/slowing down the audio or simply finding more annotated audio clips.

10. REFERENCE

1. Umair Ayub. Speech emotion recognition using machine learning < <https://medium.com/analyticsvidhya/speech-emotion-recognition-using-machine-learning-df31f6fa8404>>.
2. RAVDESS DATASET (Ryerson Audio-Visual Database of Emotional Speech and Song) < <https://paperswithcode.com/dataset/ravdess>>
3. Jouni Pohjalainen and Paavo Alku. Multi-scale modulation filtering in automatic detection of emotions in telephone speech. International Conference on Acoustic, Speech and Signal Processing, 980-984, 2014
4. Mirlab Audio Signal Processing tutorials, “Speech feature MFCC Calculation guide”
5. Mandar Gilke , Pramod Kachare , Rohit Kothalikar , Varun Pius Rodrigues and Madhavi Pednekar. MFCCbased vocal emotion recognition using ANN, International Conference on Electronics Engineering and Informatics, 150-154, 2012.
6. Jana Tuckova and Martin Sramka. Emotional speech analysis using Artificial Neural Networks. Proceedings of the International Multi conference on Computer Science and Information Technology, 141-147, 2010.
7. Tin Lay New, Say Wei Foo and Liyanage C. De Silva. Speech emotion recognition using Hidden Markov Models. Speech Communications 41, 603-623, 2003.
8. Kamran Soltani and Raja Noor Ainon. Speech emotion detection based on neural networks. In 9th International Symposium on Signal Processing and its Applications, IEEE, 2007.
9. Sirko Molau, Michael Pitz, Ralf Schlöter, and Hermann Ney. Computing mel-frequency cepstral coefficients on the power spectrum. IEEE Transaction, 2011.
10. Website: < <https://www.python.org>>
11. Ian Mccloughlin, Applied Speech and Audio Processing with MATLAB Examples. Cambridge University Press 2009. ISBN-13 978-0-521-51954-0/2.
12. <https://www.sciencedirect.com/topics/computer-science/multilayer-perceptron>
13. Website: < https://www.tutorialspoint.com/sdlc/sdlc_waterfall_model>
14. Al-Talabani, A., Sellahewa, H., & Jassim, S. A. (2015). Emotion recognition from speech: tools and challenges. Mobile Multimedia/Image Processing, Security, and Applications 2015, 9497(May 2020), 94970N. 30