VIRTUAL ASSISTANT USING NATURAL LANGUAGE PROCESSING

Project report submitted
in partial fulfillment of the requirement for award of the degree of

Bachelor of Technology
in
Computer Science and Engineering
By

B.MANEESH KUMAR (17UECS0079)
B.S.MANEESH (17UEAG0009)
K.UMA SUDAN (17UECS0773)

Under the guidance of
N.MANJUNATHAN, B.Tech, M.Tech,
ASSISTANT PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE AND 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)
CHENNAI 600 062, TAMILNADU, INDIA

April, 2020

CERTIFICATE

It is certified that the work contained in the project report titled ”Virtual Assistant using natural language processing” by ”B.MANEESH KUMAR (17UECS0079), B.S.MANEESH (17UEAG0009), K.UMA SUDAN (17UECS0773)” has been carried out under my/our supervision and that this work has not been submitted elsewhere for a degree.

Signature of Supervisor

N.Manjunadhan

Computer Science & Engineering

School of Computing

Vel Tech Rangarajan Dr.Sagunthala R&D

Institute of Science and Technology

April, 2020

Signature of Head of the Department

Dr.V.Srinivasa Rao

Computer Science & Engineering

School of Computing

Vel Tech Rangarajan Dr.Sagunthala R&D

Institute of Science and Technology

April, 2020

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 I 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)

(B.MANEESH KUMAR)

Date: / /

(Signature)

(B.S.MANEESH)

Date: / /

(Signature)

(K.UMA SUDAN)

Date: / /

APPROVAL SHEET

This project report entitled (Virtual Assistant using natural language processing) by B.MANEESH KUMAR (17UECS0079), B.S.MANEESH (17UEAG0009), K.UMA SUDAN (17UECS0773) is approved for the degree of B.Tech in Computer Science & Engineering.

Examiners

Supervisor

Mr.N.Manjunadhan,.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. V.S.S. KUMAR, Ph.D.,** for providing us with an environment to complete our project successfully.

We obligated to our beloved **Registrar Smt. N.S PREMA,** for providing immense support in all our endeavors.

We thankful to our esteemed **Director of Academics Dr. ANNE KOTESWARA RAO, Ph.D.,** providing a wonderful environment to complete our project successfully.

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

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

We also take this opportunity to express a deep sense of gratitude to Our Internal Supervisor **Mr.Manjunadhan,M.tech,** for her cordial support, valuable information and guidance, he helped us in completing this project through various stages.

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

B.MANEESH KUMAR (8806)(17UECS0079)
B.S.MANEESH (9992)(17UEAG0009)
K.UMA SUDAN (8789)(17UECS0773)

ABSTRACT

In today's world every aspect of life strongly depends on technology. In the past two centuries the world has seen development in Science and Technology in rates unprecedented in human history.This development of technology has given rise to an vibrant economy opening up new markets depending upon fields previously unknown.This boom in Science and Technology has made it difficult for people to keep up with it.Many people of old and middle ages find difficulty in adopting and taking proper advantage of their newly acquired technology. This inability to keep up with the current growth in technology and the need for specialization has given an pathway to the rise of Automation.Automation enables the users to use technology more efficiently by abstracting basic and mostly used functions and enabling the person to concentrate on the required and more important tasks.Automation is solely responsible for increasing production efficiency while not using up unnecessary raw manpower. As tasks which required to be automated increase new techniques are developed to increase productivity and efficiency. Machine learning and Artificial intelligence are two such vibrant fields. Usage of artificial neural networks has enabled machines to do many jobs which were previously considered to be only possible by humans.Our project aims to create a virtual assistant to be an aid to its users.

LIST OF FIGURES

LIST OF ACRONYM AND ABBREVIATION

SMPT	Simple Mail Transfer Protocol
GUI	Graphical User Interface
NLP	Natural Language Processing

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

1.1 Introduction

In today's world every aspect of life strongly depends on technology. In the past two centuries the world has seen development in Science and Technology in rates unprecedented in human history.This development of technology has given rise to an vibrant economy opening up new markets depending upon fields previously unknown.This boom in Science and Technology has made it difficult for people to keep up with it.Many people of old and middle ages find difficulty in adopting and taking proper advantage of their newly acquired technology.our project aims to enable old/middle aged people who are not good with using an computer and enables them to use some common functionalities by automating them.

1.2 Aim of the project

To create a virtual assistant using natural language processing and fundamentals of neural networks.

1.3 Project Domain

Artificial intelligence and machine learning

1.4 Scope of the Project

The project aims to provide assistance to old age people to operate the computer through voice input and provide the reasonable output as required ,in computer-bot voice.

1.5 Methodology

The base GUI is gives an way the computer can understand new input from speech recognition or by common text Forms the core of the project connecting all other modules together to be easily accessed PYQT5 is used to make the GUI in Python.Inputs are taken via Speech recognition,Text or Button events.The graphical user interface is a form of user interface that allows users to interact with electronic devices through graphical icons and audio indicator such as primary notation, instead of text-based user interfaces, typed command labels or text navigation. Python is an proficient tool in creating graphical user interface supporting many librarys to do so one such library is PyQt5.PyQt is a Python binding of the cross-platform GUI toolkit Qt, implemented as a Python plug-in. PyQt is free software developed by the British firm Riverbank Computing. It is available under similar terms to Qt versions older than 4.5; this means a variety of licenses including GNU General Public License (GPL) and commercial license, but not the GNU Lesser General Public License (LGPL).PyQt supports Microsoft Windows as well as various flavours of UNIX, including Linux and MacOS (or Darwin).this can be used to take in event/text inputs and give the required output.

Chapter 2

LITERATURE REVIEW

Trends in Convolutional Neural Networkin Aarya Agarwal - Westwood school of computing (2018) International Journal on Natural LanguageComputing(IJNLC) Nadine Kuhnert and Andreas Maier University, Erlangen-Nuremberg(2018) chatbot using knowledge in database.Bayu sethiaji (2016) The impact of fake news: Politics. Lexology.Kowalewski, J. (2017)Natural Language Processing: A Computer Grammar of English and its Applications.Naomi Sagar - New York University (2013).

Chapter 3

PROJECT DESCRIPTION

3.1 Existing System

The old age people find difficulty in operating regular computer due to high complexity in user interface; so the lack ability of accessing the computer's advantage becomes hard for such people

3.2 Proposed System

The proposed system provides an insight of the historical customer data over a period of time to exploit the need of computers knowledge to such people concludingly the availability to simple and intelligent user interface software becomes essential for such people to access the computer.

ADVANTAGES

The software acts as a ultimate intermediate for old age people and the computer,and will becomes easy for such people to access computer with no hesitation in order to becomes more user friendly.

3.3 Feasibility Study

The feasibility of the proposed model can be determined by the accuracy score of the machine learning model employed.We try to employ atleast three models and pick the best among them to be provided as a result.A proposed model is said to be feasible if and only if all of its metric measures are above 0.65 or 65 percent.

Three key consideration in the feasibility analysis are:

1.Economic Feasibility.

2.Technical Feasibility.

3.Social Feasibility.

3.3.1 Economic Feasibility

By Economically feasible,we mean the proposed model is within the budget limits of the individual person providers.Although the solutions provided take into account the number of customers boarding at a point,thus making it optimal to execute practically.

3.3.2 Technical Feasibility

The proposed model provides only the insights of the customer data over a period of time and suggests a ML algorithm for efficient Natural language processing.There isn't a question of technical feasibility here.

3.3.3 Social Feasibility.

The proposed model is socially feasible as it improves customer experience during any hours of the day.

3.4 System Specification

3.4.1 Hardware Specification

- OS(64 bit)-Microsoft Windows
- RAM-Minimum of 4GB for Jupyter Notebook
- Space-Minimum of 16GB

3.4.2 Software Specification

- Internet
- Windows Desktop

Chapter 4

MODULE DESCRIPTION

4.1 General Architecture

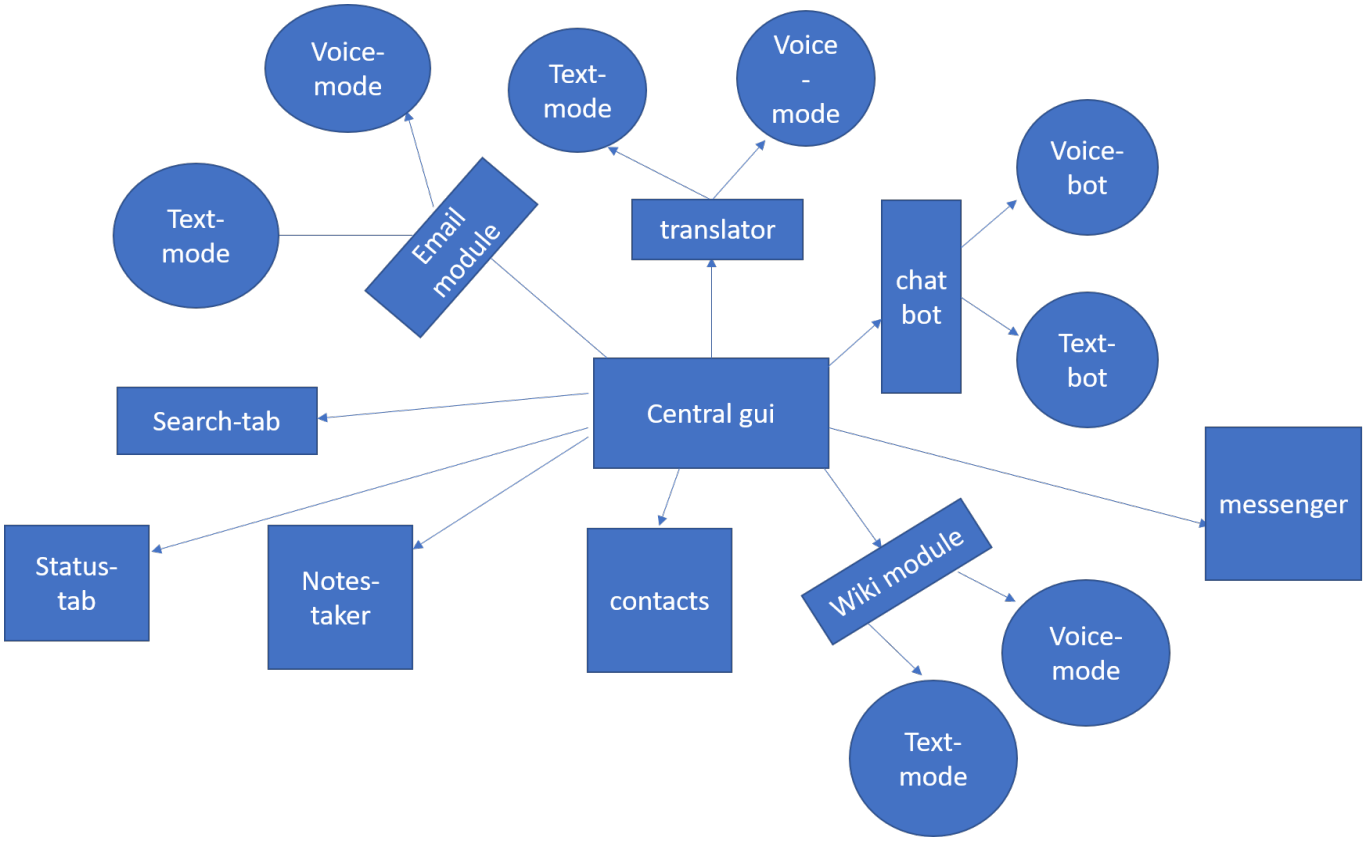

Figure 4.1: **General Architecture Diagram**

4.1.1 Detailed Architecture Diagram

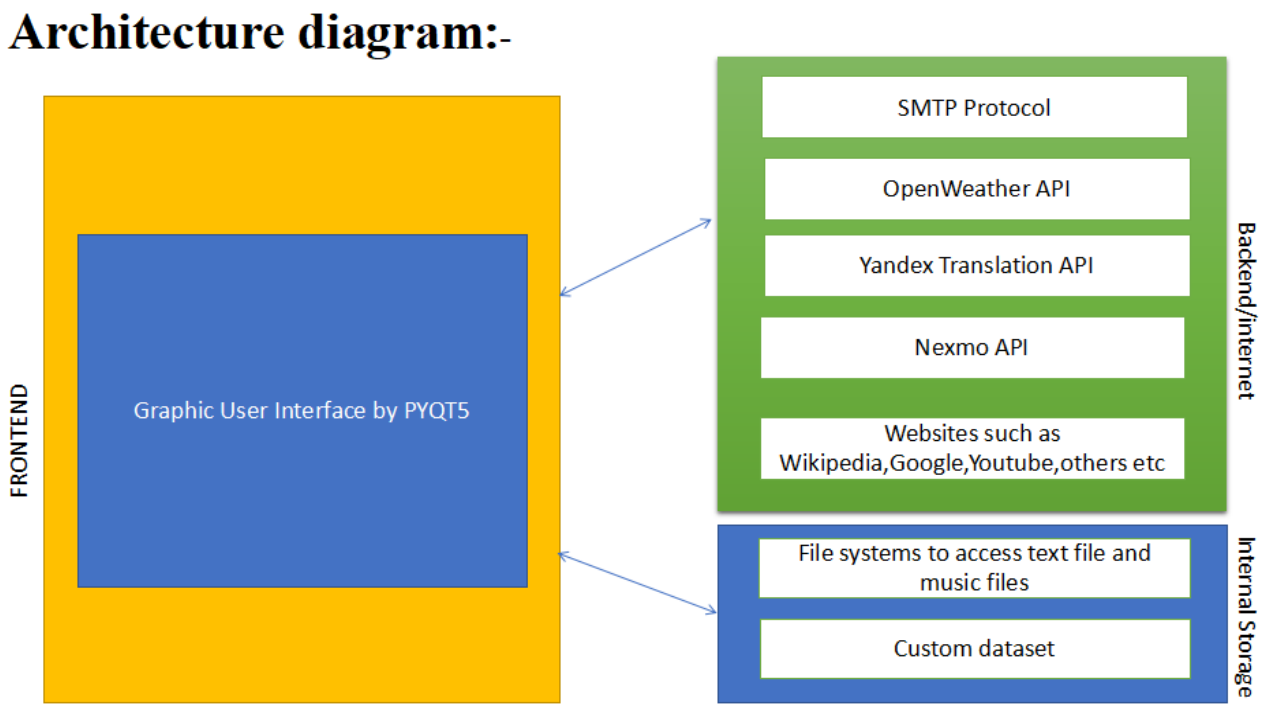

Figure 4.2: **Detailed Architecture Diagram**

.

4.2 Design Phase

4.2.1 Data Flow Diagram

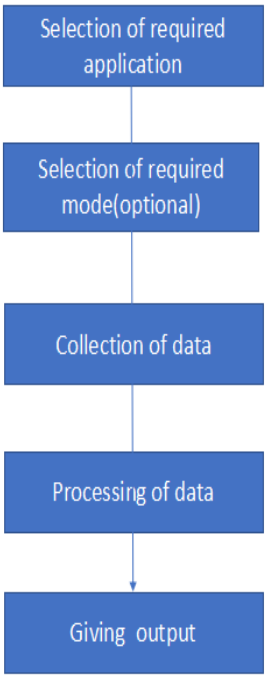

Figure 4.3: **Data Flow Diagram**

Description of Data Flow Diagram

4.2.2 Sequence Diagram

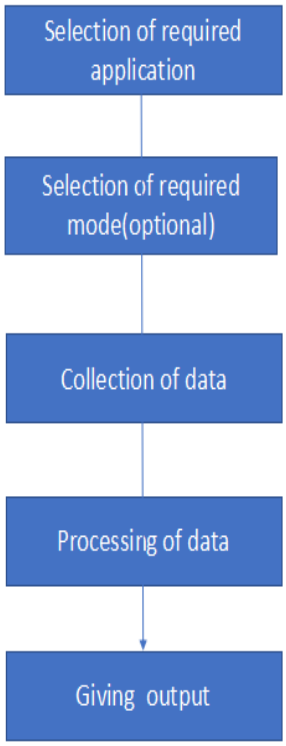

Figure 4.4: **Sequence Diagram**

Description of Sequence Diagram

4.2.3 ER Diagram

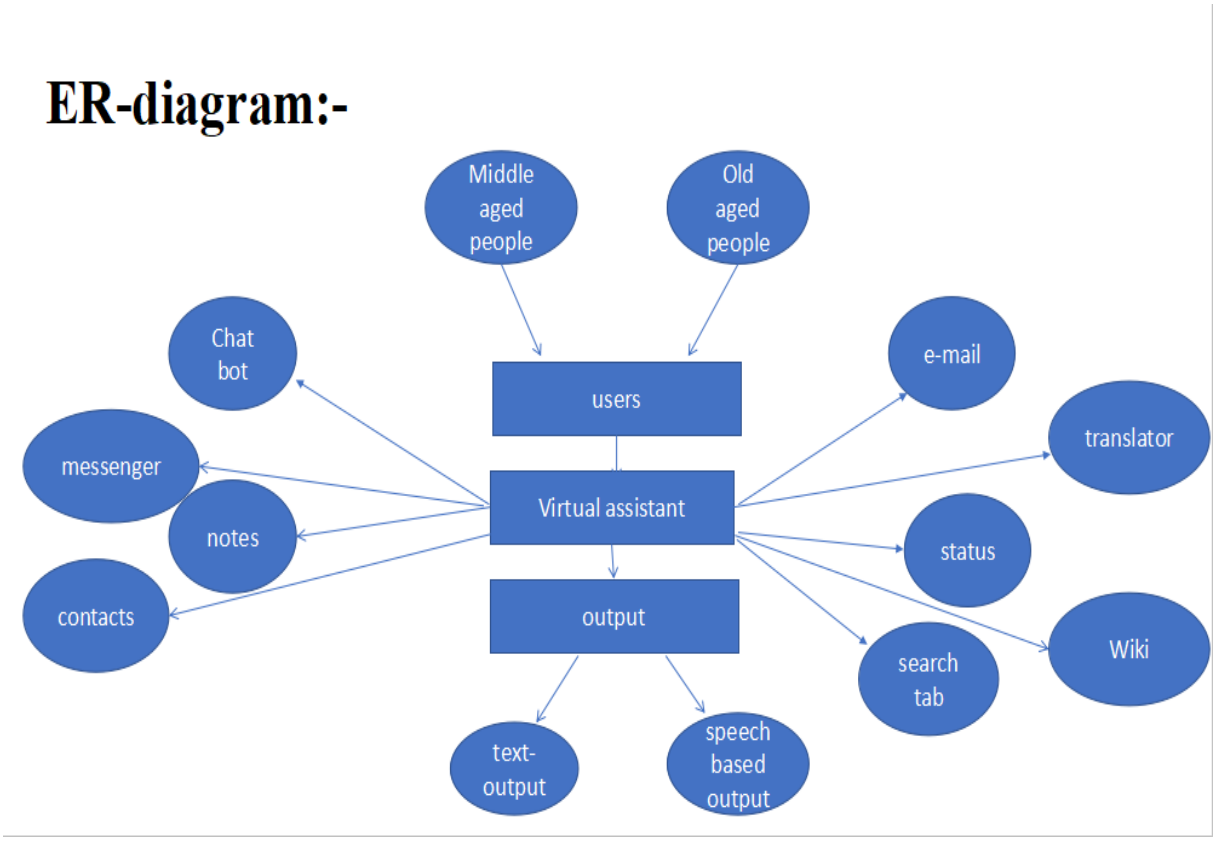

Figure 4.5: **ER Diagram**

Description of ER Diagram

4.2.4 Collaboration Diagram

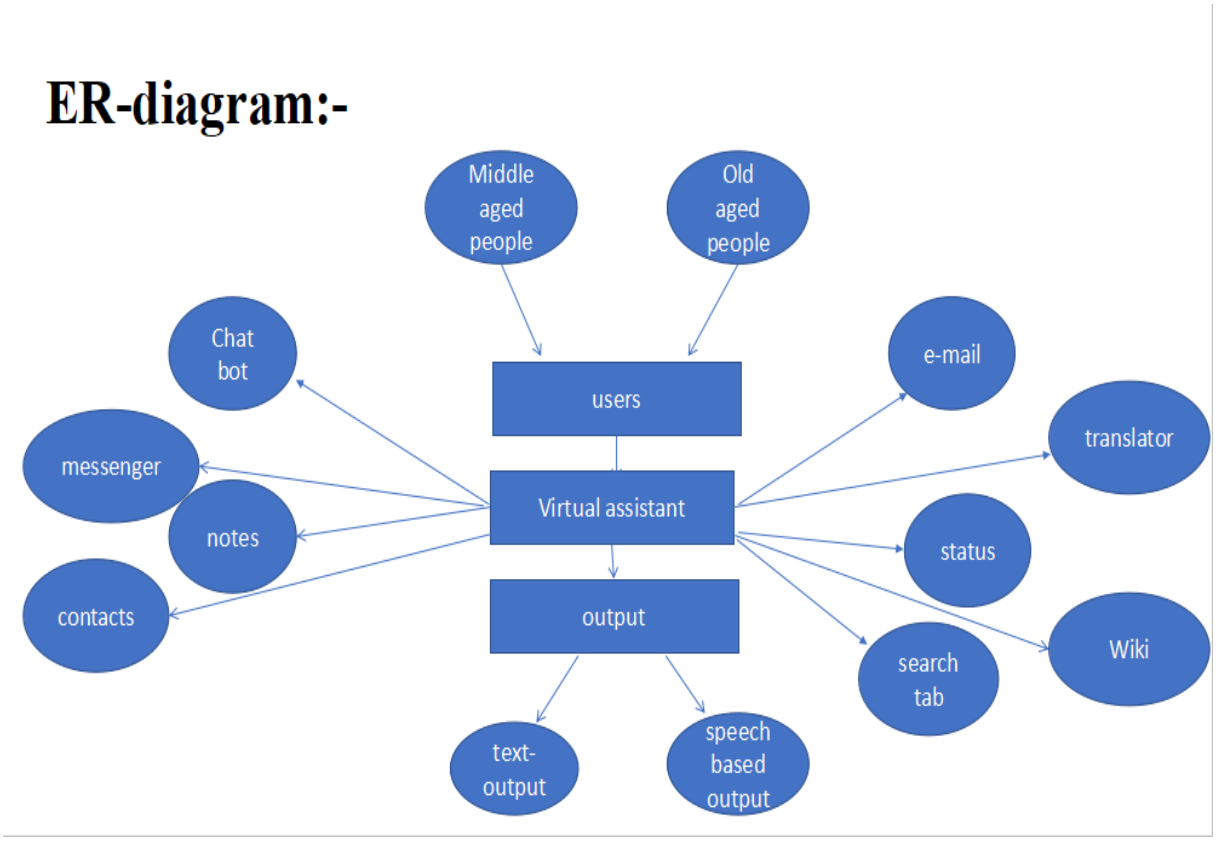

Figure 4.6: **Collaboration Diagram**

Description of Collaboration Diagram

4.3 Module Description

4.3.1 Data Owner

In simple terms, speech recognition is simply the ability of a software to recognize speech. Anything that a person says, must be recognized by the software. Speech recognition technology can be used to perform an action based on the instructions defined by the human. The human needs to train the speech recognition system by storing speech patterns and vocabulary into the system. By doing so, they can essentially train the system to understand them when they speak.

Figure 4.7: **Data Owner**

4.3.2 End User

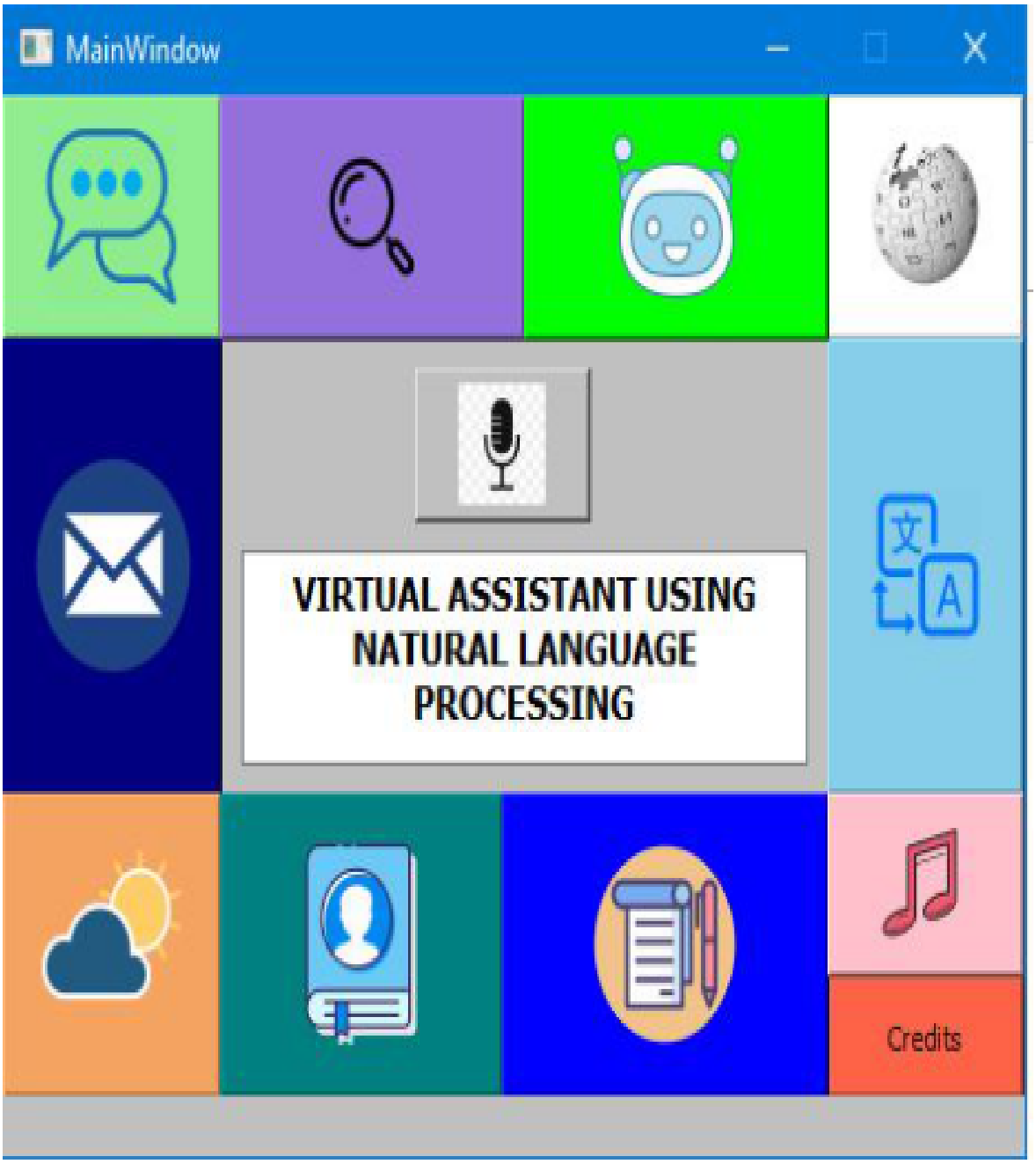

Figure 4.8: **End User**

Chapter 5

IMPLEMENTATION TESTING

Implementation is the area of the enterprise when the problematic plan is converted out into a working framework. With those sentences it is pretty good may be sited as a very general method in satisfying a delightful new design architecture and in providing the client, certainty that the new framework will be executed and will be much more effective.The implementation arrange includes cautious arranging, examination of the present design architecture and its necessary goods on pursue, procedure of methods to achieve a conversion and estimation of conversion approach.

5.1 Input Output

5.1.1 Input Design

Input Design assumes a crucial job in the existence pattern of programming improvement, it requires cautious consideration of engineers. The input design is to take care of information to the application as precise as could be expected under the circumstances. So inputs should be designed adequately with the goal that the mistakes happening while at the same time taking care of are limited. As indicated by Software Engineering Concepts, the input structures or screens are designed to give

to have an approval command over as far as possible, go and other related approvals.

This framework has input screens in practically all the modules. Blunder messages are created to alarm the client at whatever point he submits a few slip-ups and manages him in the correct manner with the goal that invalid sections are not made. Let us see profoundly about this under module structure.

Input design is the way toward changing over the client made input into a PC based organization. The objective of the input design is to make the information passage sensible and liberated from mistakes. The mistake is in the input are constrained by the input design. The application has been created in easy to use way. The structures have been designed in such a manner during the handling the cursor is set in the position where must be entered. The client is likewise furnished with in a choice to choose a proper input from different options identified with the field in certain cases.Validations are required for every datum entered. At whatever point a client enters an incorrect information, blunder message is shown and the client can proceed onward to the resulting pages in the wake of finishing all the sections in the present page.

5.1.2 Output Design

The task includes preparing a Machine Learning model to order the sort of movement an individual is performing including sitting, standing, laying, strolling, strolling upstairs and strolling first floor utilizing information gathered from smartphones.In this venture, I take the action acknowledgment dataset. The dataset incorporates sensor readings of 30 unique people and the kind of movement they were recorded for. Here, I'll utilize the dataset from Kaggle to group different exercises.

5.1.3 Sample Screenshots

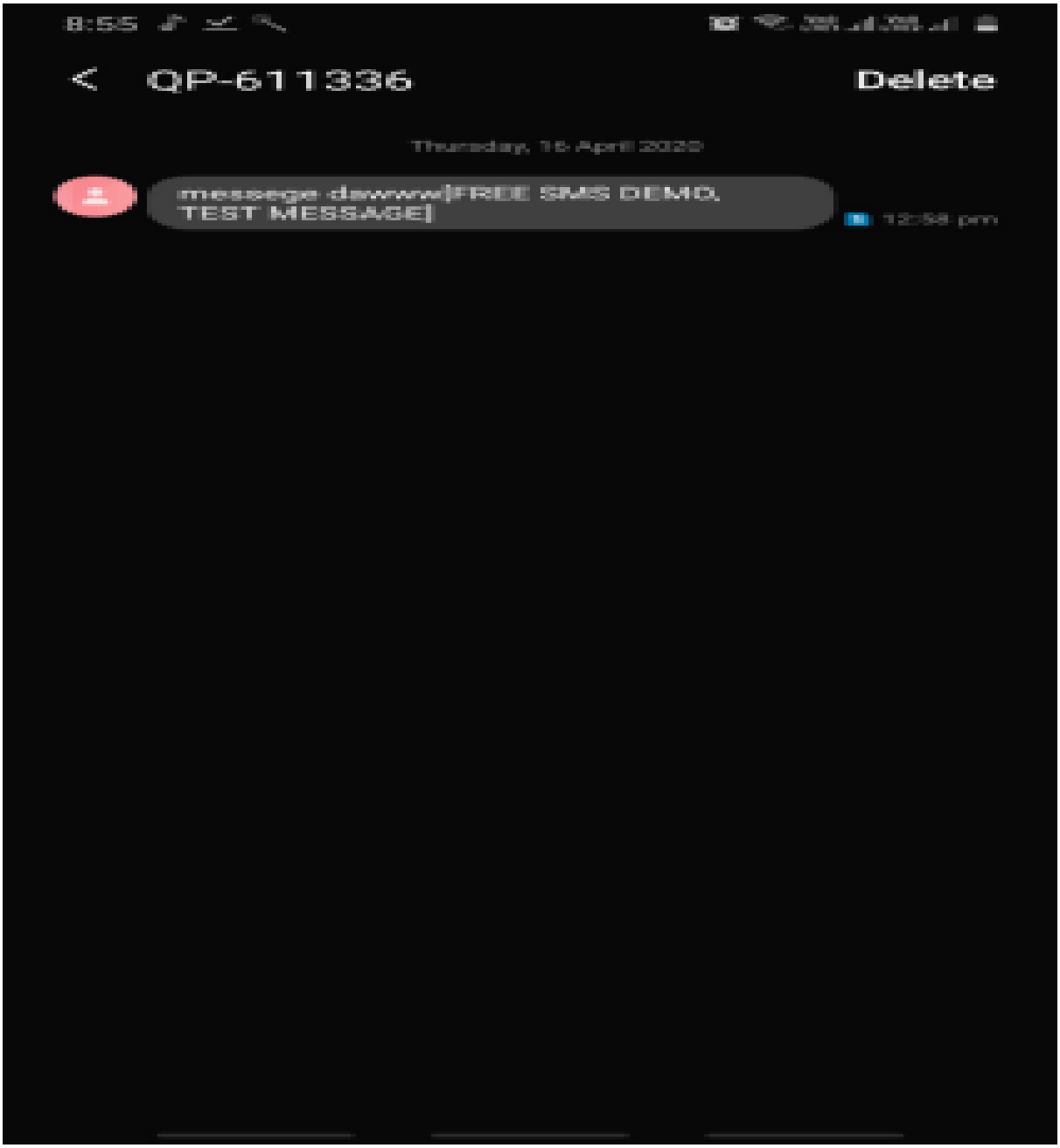

Figure 5.1: **Sms Output**

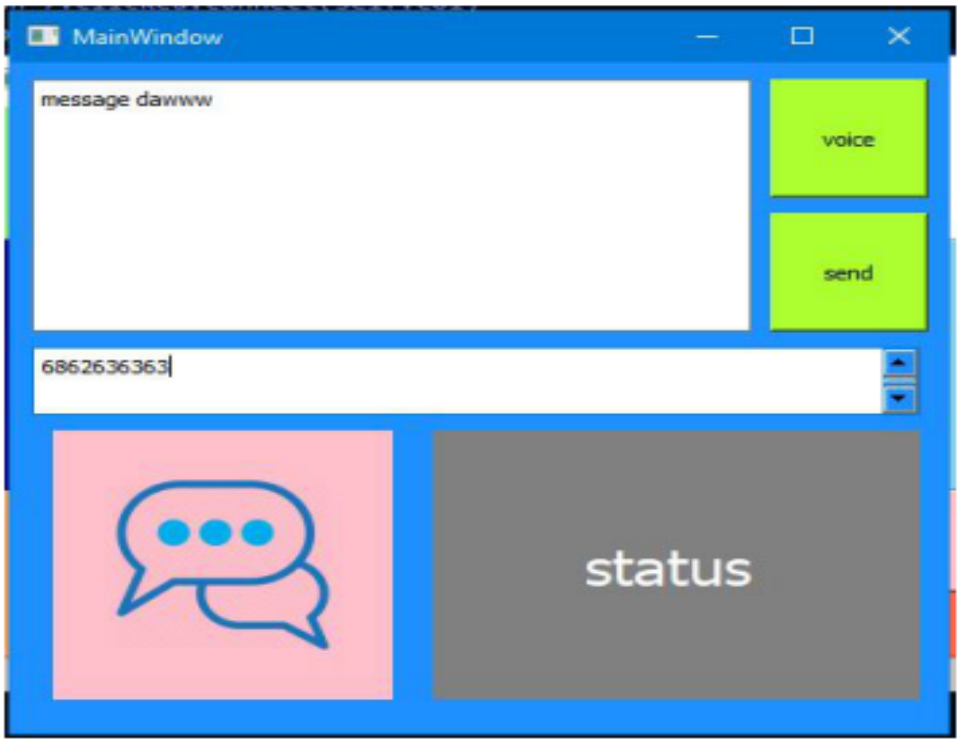

Figure 5.2: **messaging interface**

Figure 5.3: **web Scraping output**

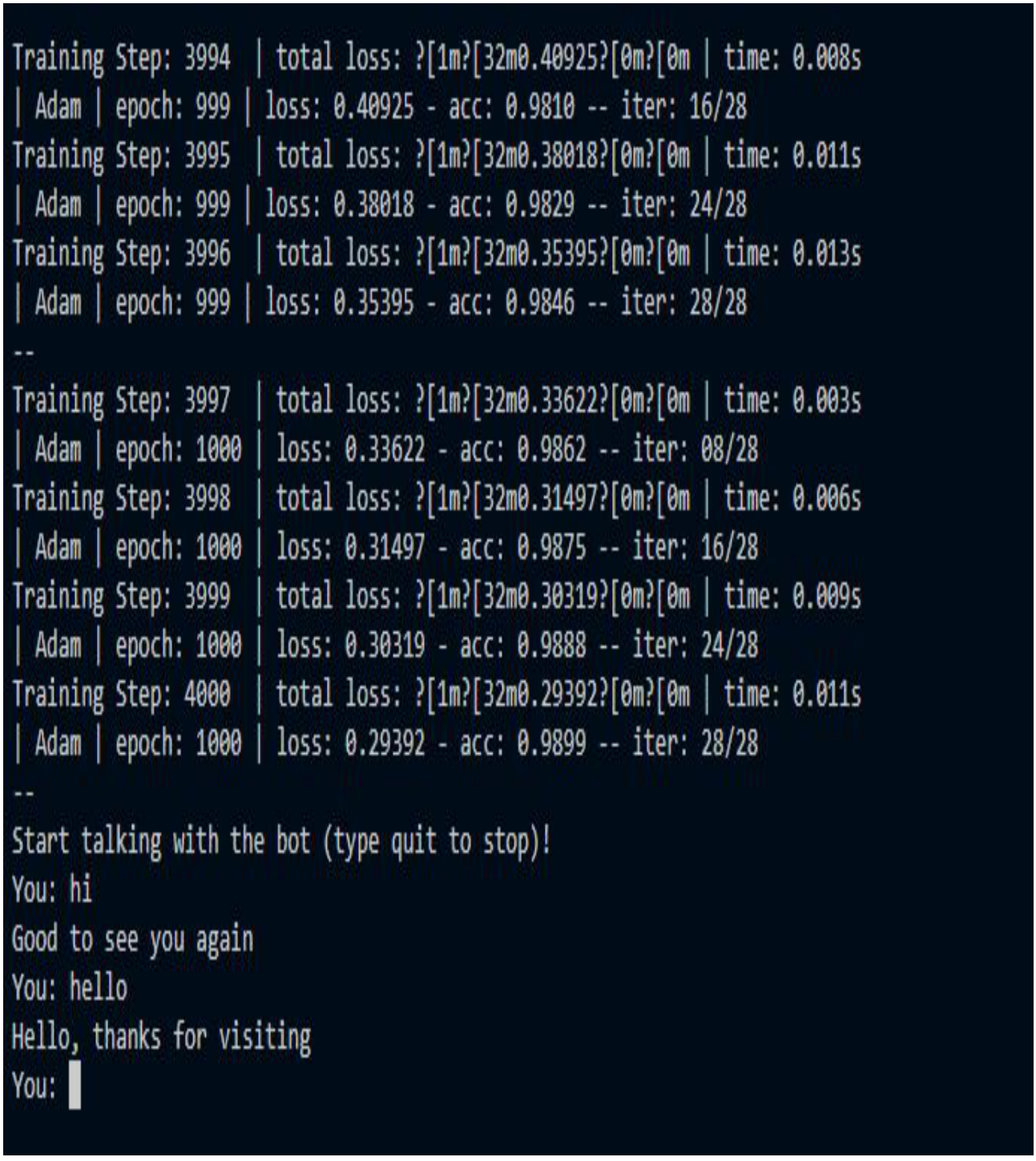

Figure 5.4: **interior setup**

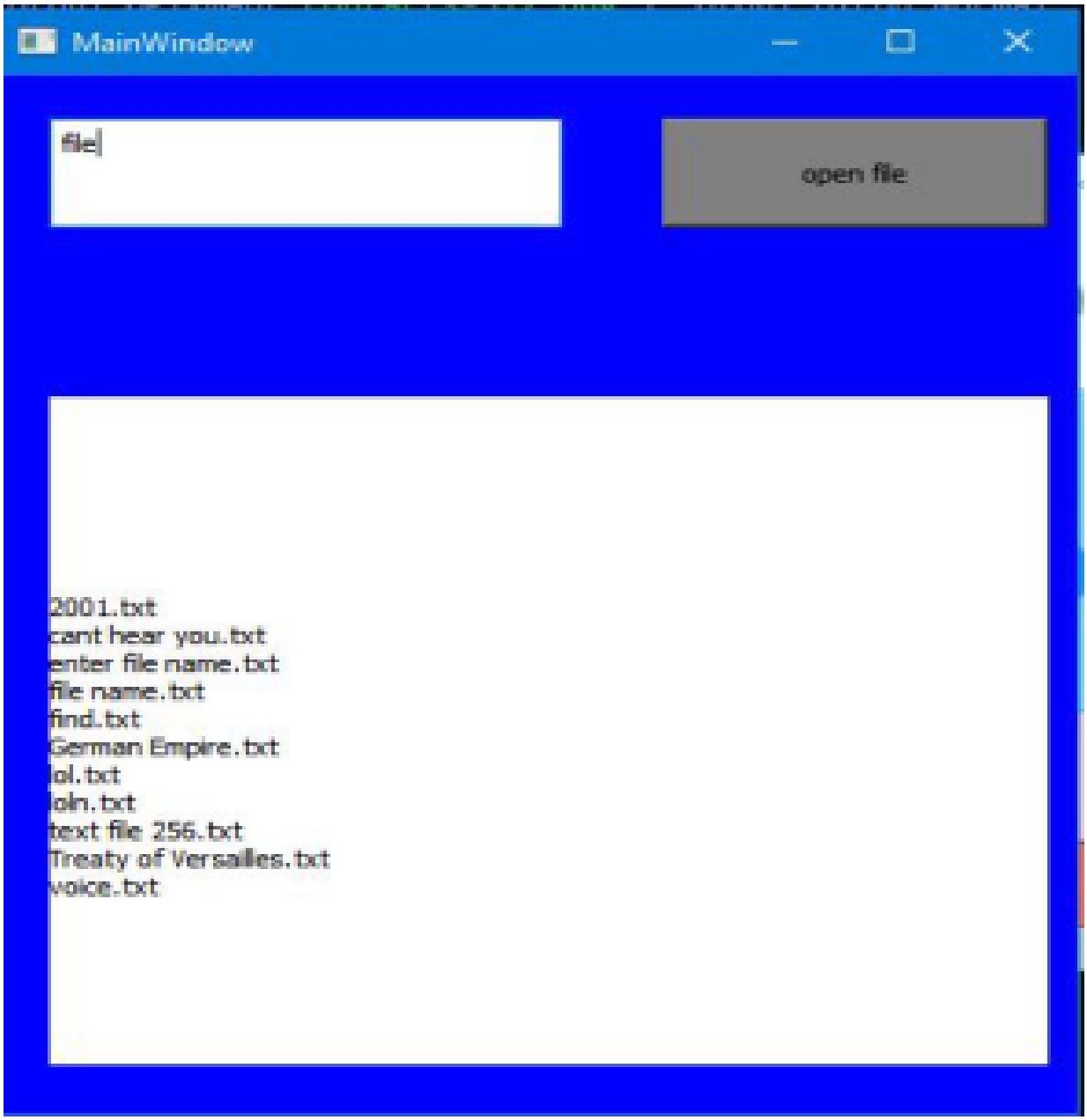

Figure 5.5: **voice to text**

5.2 Testing

The reason for testing is to find mistakes. Testing is the way toward attempting to find each possible shortcoming or shortcoming in a work item. It gives an approach to check the usefulness of segments, sub-gatherings, congregations as well as a completed item It is the way toward practicing programming with the goal of guaranteeing that the Software framework lives up to its necessities and client desires and doesn't flop in an inadmissible way.

5.3 Types of Testing

5.3.1 Unit testing

Input

```
import os
AWS_KEY_ID = os.getenv('AWS_ACCESS_KEY_ID')
AWS_SECRET_KEY = os.getenv('AWS_SECRET_ACCESS_KEY')
con = snowflake.connector.connect(
    user='XXXXXXXXXXX',
    password='XXXXXXXXXXX',
    account='XXXXXXXXXXX',
    session_parameters={
        'QUERY_TAG': 'EndOfMonthFinancials',
    }
)
con.cursor().execute("ALTER SESSION SET QUERY_TAG = 'Taxi_data'")
```

Test result

It is checked whether the information is properly flowing in to the program unit and properly happen out of it or not using module interface testing.

5.3.2 Integration testing

Input

```
USER = '<login_name>'
PASSWORD = '<password>'
con = snowflake.connector.connect(
    user=<USER>,
    password=<PASSWORD>,
    account=<SNOWFLAKE_ACCOUNT>,
    auth='https://<account_name>.com/',
    warehouse=<WAREHOUSE>,
    database=<DATABASE>,
    schema=<SCHEMA>
)
```

Test result

From the datasets we are seeing 5 information's and checks where singular units are joined and tried as a gathering.

5.3.3 Functional testing

Input

```
Create or replace storage integration s2_obj type= external_stage
    storage_provider=s3 enabled=true storage_aws_role_arn="arn:aws:iam
    ::840369218167:role/snowflake" storage_allowed_locations=("s3://
    ameexsnowflake/");

Create or replace file format demo_file_format type=csv field_delimeter="," skip
     header=1 nullid=('NULL'.'null') emptyFieldAsNull=true;

Create or replace stage snow_stage storage_integration=s3_obj url="s3//
    ameexsnowflake" file_format=demo_file_format;

```

```
Create or replace pipe newpipe auto_ingest=true as copy into "attempt" from
    @stage_name;
Desc pipe newpipe;
```

```
def load_into_df(warehouse,db,schema,table):
Conn = self.conn
iter = conn.cursor()
iter.execute("USE WAREHOUSE" + warehouse)
iter.execute("USW DATABASE" + db)
try:
    result = conn.cursor().execute("select * from "+schema+'.'+table)
    result_list = result.fetchall()
    cols = iter.execute("select COLUMN_NAME from INFORMATION>SCHEMAS>COLUMNS
        where TABLE_NAME+{} and TABLE_SCHEMA = {}".format("'"+table.upper()+"'","
        '"+
schema.upper()+"'")); cols = cols.fetchall(); columns = [col for x in cols
    for col in x;
    data = pd.DataFrame(result_list,columns=columns);
    return data;
finally:
    conn.cursor().close();
conn.cursor().close();
```

Test Result

It checks whether the information is partitioned and showed up as visualized way.

5.3.4 White Box Testing

```
import os
AWS_KEY_ID = os.getenv('AWS_ACCESS_KEY_ID')
AWS_SECRET_KEY = os.getenv('AWS_SECRET_ACCESS_KEY')
con = snowflake.connector.connect(
    user='XXXXXXXXXXX',
    password='XXXXXXXXXXX',
    account='XXXXXXXXXXX',
    session_parameters={
        'QUERY_TAG': 'EndOfMonthFinancials',
    }
)
con.cursor().execute("ALTER SESSION SET QUERY_TAG = 'Taxi_data'")

USER = '<login_name>'
PASSWORD = '<password>'
con = snowflake.connector.connect(
  user=<USER>,
  password=<PASSWORD>,
  account=<SNOWFLAKE_ACCOUNT>,
  auth='https://<account_name>.com/',
  warehouse=<WAREHOUSE>,
  database=<DATABASE>,
  schema=<SCHEMA>
)
def load_into_df(warehouse,db,schema,table):
Conn = self.conn
iter = conn.cursor()
iter.execute("USE WAREHOUSE" + warehouse)
iter.execute("USW DATABASE" + db)
try:
    result = conn.cursor().execute("select * from "+schema+'.'+table)
    result_list = result.fetchall()
    cols = iter.execute("select COLUMN_NAME from INFORMATION>SCHEMAS>COLUMNS
        where TABLE_NAME+{} and TABLE_SCHEMA = {}".format("'"+table.upper()+"'","
        '"+
schema.upper()+"'")); cols = cols.fetchall(); columns = [col for x in cols
    for col in x;
    data = pd.DataFrame(result_list,columns=columns);
```

```
    return data;
finally:
    conn.cursor().close();
conn.cursor().close();
```

5.3.5 Black Box Testing

```
Test case:Checking connectivity of python notebook with snowflake
Test Input:
sfaccount:'ameex_snowflake.southeast-asia',
sfuser:'intern_practice_account',
sfdatabase:'taxis_latest.csv',
sfschema:'taxis_schema'


Test Output:connected and able to retrieve data
Expected Output:proper connection and able to manipulate/view data
```

5.3.6 Test Result

Figure 5.6: **Passenger Count by Location**

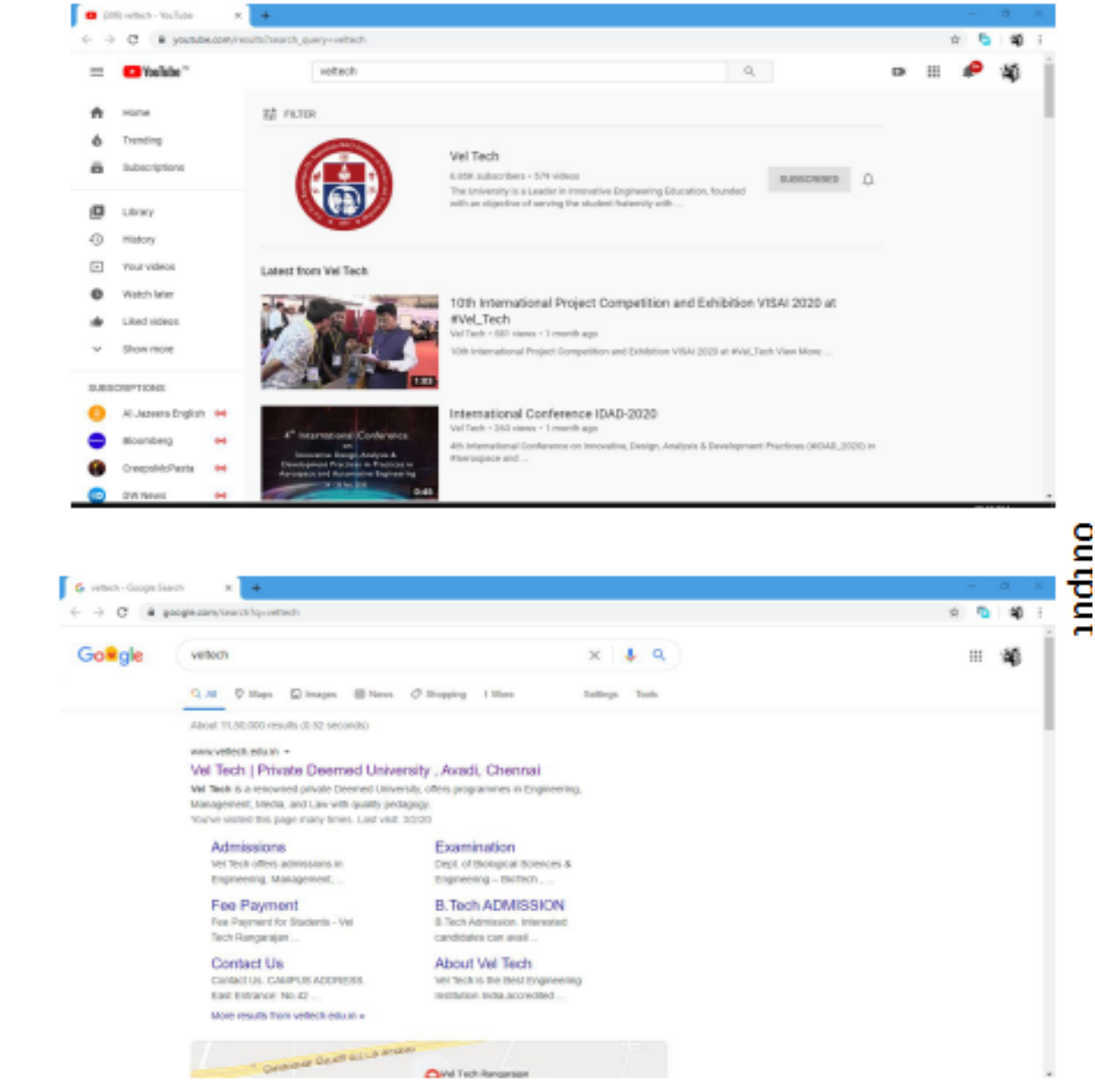

Figure 5.7: **Web searched results**

Figure 5.8: **Language Translation**

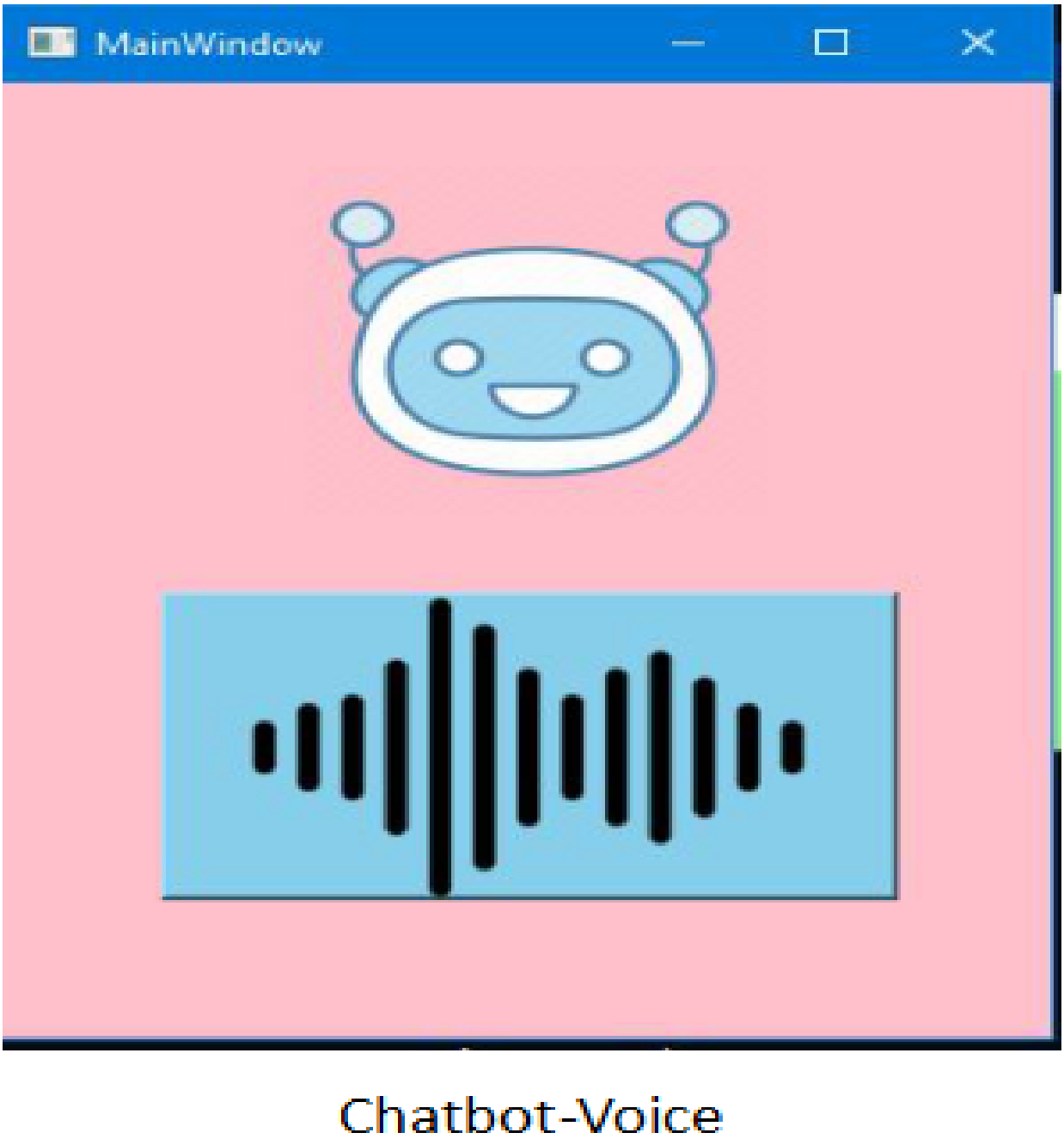

Figure 5.9: **Chat Bot Output**

5.4 Testing Strategy

A methodology for framework testing incorporates framework experiments and structure systems into an all around arranged arrangement of steps that outcomes in the effective development of graphical portrayal. The testing procedure must co-work test arranging, experiment configuration, test execution, and the resultant information assortment and assessment .A technique for programming testing must suit average-order assessments which is important to check that a small main program have precisely actualized just as elevated level tests that approve significant framework capacities against client prerequisites. Testing speaks to an intriguing peculiarity for the examination framework. Consequently, a progression of testing are performed for the proposed framework before the framework is prepared for client acknowledgment testing.

Chapter 6

RESULTS AND DISCUSSIONS

6.1 Efficiency of the Proposed System

SMTP stands for Simple Mail Transfer Protocol. SMTP is a set of communication guidelines that allow software to transmit an electronic mail over the internet is called Simple Mail Transfer Protocol. It is a program used for sending messages to other computer users based on e-mail addresses. It provides a mail exchange between users on the same or different computers, and it also supports: It can send a single message to one or more recipients. Sending message can include text, voice, video or graphics. It can also send the messages on networks outside the internet. The main purpose of SMTP is used to set up communication rules between servers. The servers have a way of identifying themselves and announcing what kind of communication they are trying to perform. They also have a way of handling the errors such as incorrect email address. For example, if the recipient address is wrong, then receiving server reply with an error message of some kind..

The existing system comprises of a blind fold technique where the cabs were stationed at random areas in the town/city without having a clear outlook of where there is chance of getting higher customers hence decreasing the efficiency of pick up and

drop off , also making to wait the customer and increasing the fuel consumption.In the proposed system based on the model that is created from the historical customer data obtained from the cab service provider certain visualizations can be made with a predictive model to show which area has higher traffic at what time based on which the cabs can be moved on.Hence this proves advantageous to the cab service provider when compared to the existing system.

6.2 Advantages of the Proposed System

- Old people can access the computer
- knowledge of computer and internet to reach everyone

Chapter 7

CONCLUSION AND FUTURE ENHANCEMENTS

7.1 Conclusion

Our project is especially useful for middle aged or old aged people who are not good and have not yet adopted to fully utilize the many features offered by an computer. It abstracts many features which are commonly used in computers in order to help people who are not familiar with technology. It is able to perform many essential and some niche tasks which normal people use computers .

7.2 Future Enhancements

Automation is solely responsible for increasing production efficiency while not using up unnecessary raw manpower. As tasks which required to be automated increase new techniques are developed to increase productivity and efficiency. Machine learning and Artificial intelligence are two such vibrant fields. Usage of artificial neural networks has enabled machines to do many jobs which were previously considered to be only possible by humans.Our project aims to create a virtual assistant to be an aid to its users.

References

[1] *Joe Kraynak and David Baum, 2019, "Cloud Warehousing for Dummies.*; Wiley, vol. 2, pp. 12-35

[2] *Susan Dean and Barbara llowsky, 2009, "Principles of Business Statistics," in: EUROCRYPT*; Connexions, vol. 76, pp. 354-389

[3] *M. M. Vazifeh1*, P. Santi1,2, G. Resta2, S. H. Strogatz3 C. Ratti, 2010, "Addressing the minimum fleet problem in on-demand urban mobility"*; Connexions, pp. 111-115

[4] *Hinton GE, Osindero S, 2008 , IEEE "The Y W.A fast learning algorithm for deep belief nets[J]" , pp:1527-1554.*

[5] *Arel I, Rose D C,Karnowski T P. 2010 , IEEE , "Deep machine learning-A new in artificial intelligence ", pp: 13-18.*

[6] *Liqi Xu, Richard L. Cole, Daniel Ting, 2019, IEEE , "Learning to optimize Federated Queries", pp: 45-60.*

[7] *Dipesh Bhawnani,2015, IEEE, "Big Data Analytics on cab Conpany's customer Dataset" vol1 - pp: 2395-3470.*

[8] *Geeta Kesavaraj,2013, IEEE, "Customer Relationship Management" vol1 - - pp:2347-3000.*

[9] *Kumar, Kishore, Ramesh, 2016, IEEE, "Factors Influencing the Customers in Selection of Cab Services", vol4 - pp:557-561.*

[10] *Rexi A, 2016, IEEE, "Customer Satisfaction towards Cab Availability" vol1 - pp:58-60.*

[11] *Andrew McNutt, Gordon Kindlmann,Michael Correll, 2020, IEEE,"Surfacing Visualization Mirages", pp:25-46.*

[12] *Andrew McNutt, Anamaria Crisan, Michael Correll, 2020, IEEE,"Divining Insights: Visual Analytics Through Cartomancy",pp 123-145.*

[13] *Chan YH, 2010, IEEE, "Biostatistics 201:Linear Regression Analysis", pp:211-222.*

[14] *Carroll Mariana, Merwe Alta van der, 2011, IEEE, "Secure Cloud Security-Trends and Research Directions", pp:1023-1049.*