MINOR PROJECT 2

Mid-Term Report

ON

Sensei - A career assistant

Submitted By

Name	Sap ID	Roll No.	Branch
Krishan Rupani	500067535	R171218057	Btech-CSE DevOps
Animesh Jain	500069453	R171218022	Btech-CSE DevOps
Aaryan Bajaj	500071247	R164218100	Btech-CSE - IoT
Varnit Kaushik	500069026	R142218070	Btech-CSE - G & G

Under the guidance of

Dr. Rohit Srivastava
Assistance Professor
School of Computer Science

UPES
UNIVERSITY WITH A PURPOSE

Department of Cybernetics,
School of Computer Science
UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
Dehradun-248007

School of Computer Science
University of Petroleum & Energy Studies, Dehradun

Project Proposal Approval Form (2019-2020)

Minor | II |

Project Title: Sensei - A career assistant

Abstract: (Half to 2/3 of a page) (Mention Keyword at the Bottom)

Sensei is an AI-ML based Chabot which helps high school graduates to choose a perfect and apt college as per their desire. Our Chabot acts like an actual mentor and analysis the requirement of student and gives a personalized answer to every question asked from it. The name 'Sensei' hold a special meaning in itself, Sensei in Japanese means 'Teacher' or 'Mentor'. Sensei will be AI based which means it will simulate the intelligent behavior to its best possible extent. Sensei will guide students down the right path.

Keywords: Career Assistant, Mentor, Chabot, AI-ML, Cognitive Computing, Supervised learning, Narrow AI, python, OOPS

Table of Content

S.No	Topic	Page Number
1.	Introduction	4
2.	Problem Statement	5
3.	Objective	6
4.	Literature Review	6
5.	Methodology	7
6.	System Requirement	8
7.	Execution	9-14
8.	Output	15-16
9.	Pert Chart	17
10.	UML Diagram	18
11.	Data Flow Diagram	19
11.	References	20

Introduction:

Sensei in Japanese means 'Teacher' or 'Mentor'. Our Sensei will be mentor to all those who are facing difficulty in choosing a career path after High School. Many of my batch mates were confused on which college to choose after 12th and in which stream to go, many were forced upon the path decided by their parents, some had their interests and they followed that but a large pool was clueless, they didn't know what were their best options with the grade they got in 12th or with their entrance exam result. I saw many of my friends paying thousands to career counselors to hear what they thought was best. It was a total chaos. That's why we thought of making an intelligent system which will help students and also guide them impartially.

What we are focusing on is building an AI based Chabot which uses concepts of Machine Learning and Deep learning for giving personalized answers with real time response using conversation Strategies. The Sensei will be using Supervised learning i.e. we will have class labels in the datasets and we use these to build the classification models this will mean that we will be using structured dataset for computing. Sensei will be a weak or narrow AI that means it will be limited to a specific area that is Career guidance, it simulates human intelligence thus have potential to benefit society by automating task and analyze data.
The purpose for building this Chabot will be to build a Cognitive computing system, because we are using Machine learning, the system shall Observe, Interpret, Evaluate and Decide on its own just like humans do.

To make our goal reality we have to follow certain steps, we will use python to code the Chabot, the reason behind using python is because it is an object oriented language so we can use benefits of OOPS, also python is easy to code by reducing lines of codes. We will use datasets in a structured format and Sensei will take raw information from that dataset and will give a personalized answer to the user.

Market Analysis

If we talk about the market size, then we have a huge market. In 2020, about 12 lakh students appeared in 12th exam also there is huge pool of students who take a 1-year gap or drop after 12th. These are our target audience.
There are not many AI based career assistants out in the market, so this is fresh concept, if implemented right this could turn into a well-functioning startup as it has a high market demand, less competition.

Problem Statement:

Many students are lost and confused for what to do after 12th, they don't know what works best for them and what their dreams and passion are. Many are forced down the lane build by their parents, many don't know what college to choose and which stream to apply in order to fulfil their dreams. Out of confusion students go to google and search randomly and decide their career based on a search. Many times this google information is miss leading so it leads to bad career decisions. While some do this other go to career consulators where they pay a huge chunk of money but only hear what the counselor wants them to hear, either these are biased opinions or they are limited to their knowledge.

Objectives:

To create a AI based career assistant which help students choose branch and college for engineering.

Literature Review:

AN INTELLIGENT WEB-BASED VOICE CHAT BOT
This paper presents the design and development of an intelligent voice recognition chat bot. The paper presents a technology demonstrator to verify a proposed framework required to support such a bot (a web service). While a black box approach is used, by controlling the communication structure, to and from the web-service, the web-service allows all types of clients to communicate to the server from any platform. The service provided is accessible through a generated interface which allows for seamless XML processing; whereby the extensibility improves the lifespan of such a service. By introducing an artificial brain, the web-based bot generates customized user responses, aligned to the desired character. Questions asked to the bot, which is not understood is further processed using a third-party expert system (an online intelligent research assistant), and the response is archived, improving the artificial brain capabilities for future generation of responses
https://www.researchgate.net/publication/224564336_An_intelligent_web-based_voice_chat_bot.

Designing a Chatbot for Diabetic Patients
Artificial Intelligence chatbot is a technology that makes interaction between man and machine possible by using natural language. In this paper, we proposed an architectural design of a chatbot that will function as virtual diabetes physician/doctor. This chatbot will allow diabetic patients to have a diabetes control/management advice without the need to go to the hospital. A general history of a chatbot, a brief description of each chatbots is discussed. We proposed the design of a new technique that will be implemented in this chatbot as the key component to function as diabetes physician. Using this design, chatbot will remember the conversation path through parameter called Vpath. Vpath will allow chatbot to gives a response that is mostly suitable for the whole conversation as it pecifically designed to be a virtual diabetes physician
https://www.researchgate.net/publication/266872926_Designing_a_Chatbot_for_Diabetic_Patients

Pharmabot: A Pediatric Generic Medicine Consultant Chatbot
The paper introduces a Pharmabot: A Pediatric Generic Medicine Consultant Chatbot. It is a conversational chatbot that is designed to prescribe, suggest and give information on generic medicines for children. The study introduces a computer application that act as a medicine consultant for the patients or parents who are confused with the generic medicines. The researchers use Left and Right Parsing Algorithm in their study to come up with the desired result.
http://www.joace.org/uploadfile/2014/0801/20140801025000959.pdf

Methodology:

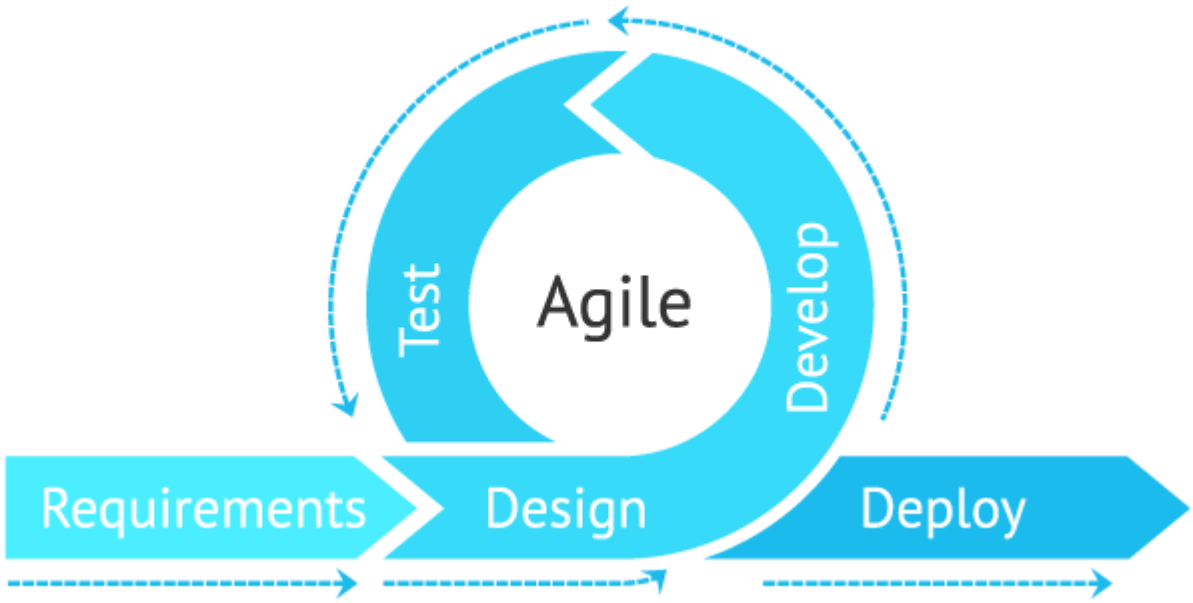
We will make our project using agile methodologies following agile values and principals.
Agile project management is a methodology that is commonly used to deliver complex projects due to its adaptiveness. It emphasizes collaboration, flexibility, continuous improvement, and high quality results.
The Agile Manifesto is comprised of four foundational values and 12 supporting principles which lead the Agile approach to software development.

Agile Values:

1. Individuals and Interactions Over Processes and Tools
2. Working Software Over Comprehensive Documentation
3. Customer Collaboration Over Contract Negotiation
4. Responding to Change Over Following a Plan

Agile Principals:

1. Customer satisfaction through early and continuous software delivery
2. Accommodate changing requirements throughout the development process
3. Frequent delivery of working software
4. Collaboration between the business stakeholders and developers throughout the project
5. Support, trust, and motivate the people involved
6. Enable face-to-face interactions
7. Working software is the primary measure of progress
8. Agile processes to support a consistent development pace
9. Attention to technical detail and design enhances agility
10. Simplicity
11. Self-organizing teams encourage great architectures, requirements, and designs
12. Regular reflections on how to become more effective



System Requirements: (Software/Hardware)

Software:

=> Windows (7 and above)/ Linux Operating System/macOS
=> Python versions: 2.7.X, 3.6.X
=> Anaconda

Hardware:

=> 2 GB RAM
=> 100 GB externally hard disk
=> 1 gigahertz (GHz) or faster 32-bit (x86) or 64-bit (x64) processor
=> Processors: Intel Atom processor or Intel Core™ i3 processor.

Execution

In this Minor 2, we are going to build a chatbot using deep learning techniques. The chatbot will be trained on the dataset which contains categories (intents), pattern and responses. We use a special recurrent neural network (LSTM) to classify which category the user's message belongs to and then we will give a random response from the list of responses.

Let me introduce 2 important terms over here, Retrieval based chatbot and Generative based chatbot. Retrival based chatbot uses predefine input patterns and responses while generative based chatbot is not based on predefined responses they are based on sequence to sequence neural network.

Making a retrieval based chatbot:

The journey of making a narrow AI chatbot or retrieval based chatbot is divided into 5 parts:

1. Importing and loading data file
2. Preprocessing
3. Training and testing data
4. Building the model
5. Predicting the response and giving output

I. Importing and loading data file

i. Create file name train_chatbot.py
ii. Create dataset file in .aiml format
iii. Create a file std_startup.aiml
iv. Call all our dataset in this file
v. Import aiml
vi. Import other necessary packages as well, nltk, pickle, numpy, keras, tenserflow
vii. Initialize variables, list, directories
viii. Call std_startup.aiml in our code to access the data.

II. Preprocessing

We need to preprocess data before making ML model, we start with tokenizing, it is the process of breaking the whole text into small parts like words. We will use nltk.word_tokenize( ) method over here.
Second part of preprocessing our data is lemmatizing it. Lemmatizing is a process of converting words into its lemma form and then create a pickle file to store python objects. Pickle file: python pickle module is used for serializing and deserializing a python object structure.

i. Import nlkt
ii. Run a for loop through data
iii. Tokenizng using nltk.word_tokenize( )
iv. Create a list
v. Add each word to the list
vi. Create a list of classes for our tags

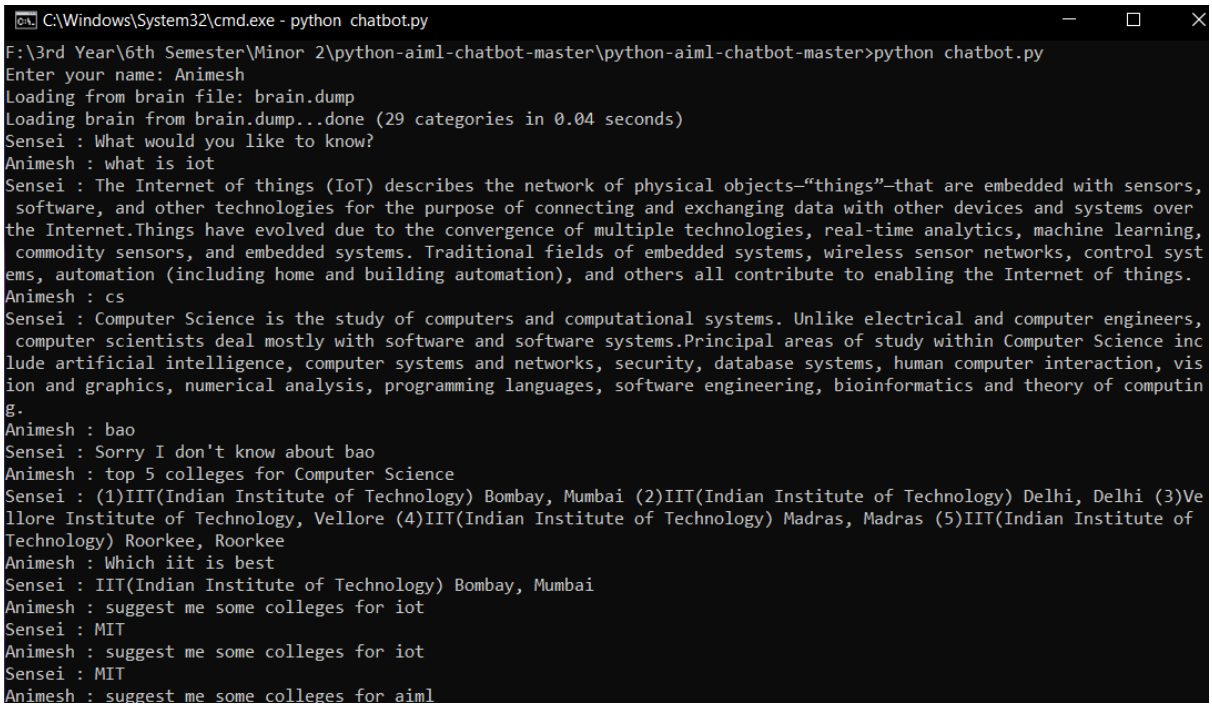
vii. Lemmatizing each word using WordNetLemmatizer
viii. Remove spaces and duplicate word
ix. Create pickle file to store python objects

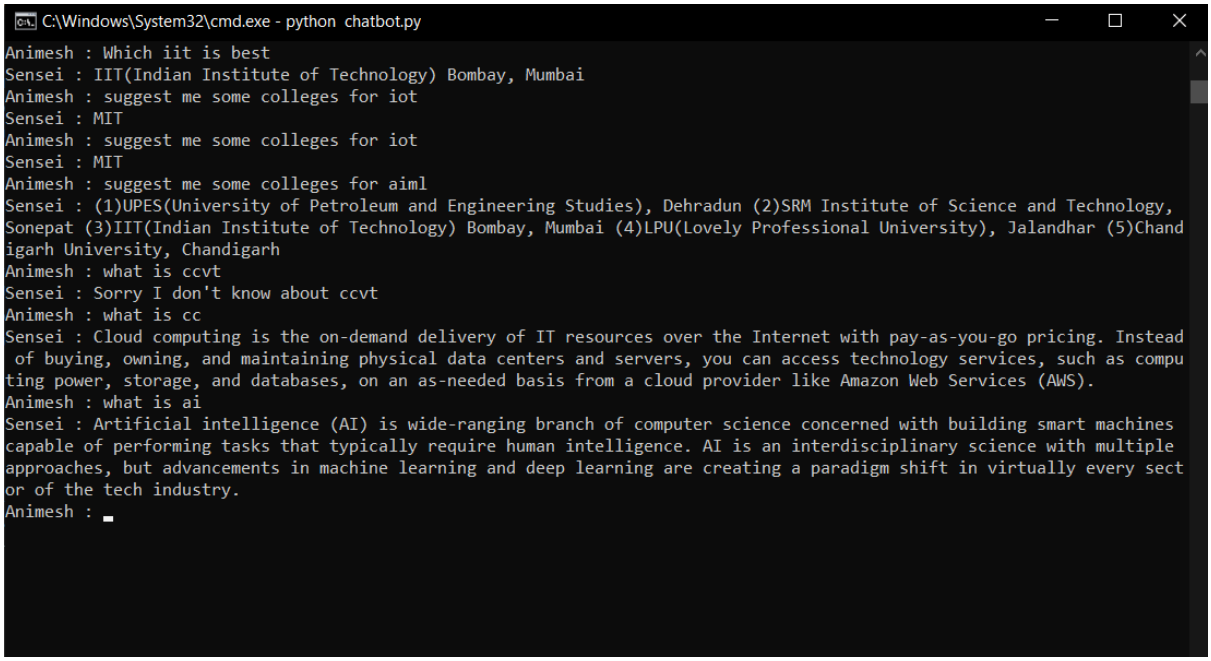
III. Training and testing data

We will create our training data, input will be pattern, output will be class our pattern belongs to.

i. Create list training
ii. create an empty array for our output
iii. training set, bag of words for each sentence
iv. initialize our bag of words
v. list of tokenized words for the pattern
vi. lemmatize each word - create base word, in attempt to represent related words
vii. create our bag of words array with 1, if word match found in current pattern
viii. output is a '0' for each tag and '1' for current tag
ix. shuffle our features and turn into np.array
x. create train and test lists

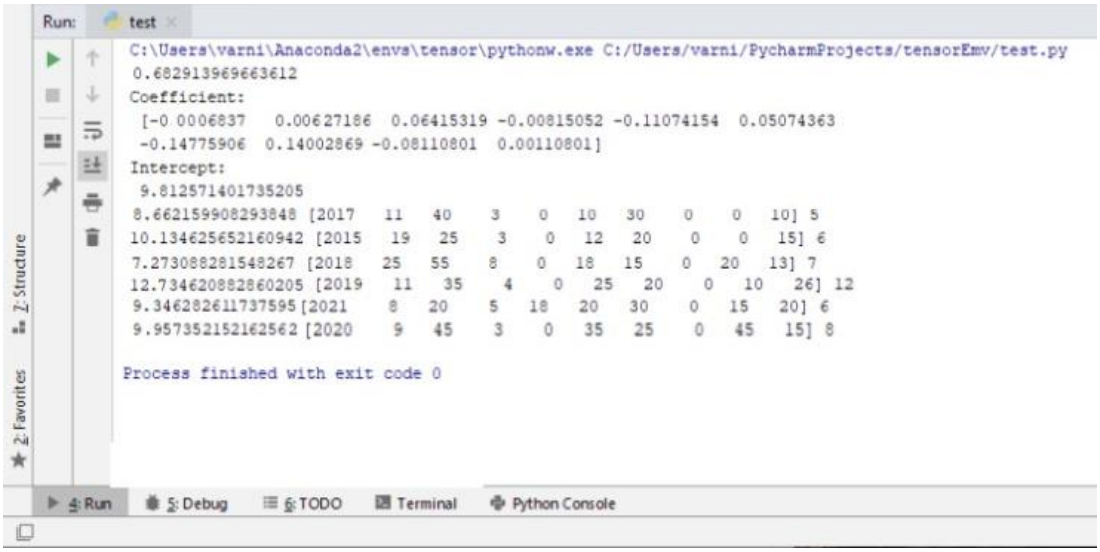
At this stage Sensei looks something like this:





IV. Building the model

i. Importing libraries
ii. Loading dataset
iii. Split to independent and dependent variables
iv. Splitting data into training and testing data
v. Choosing the Model
vi. Fit our model
vii. Predict the output
viii. Plot the graph



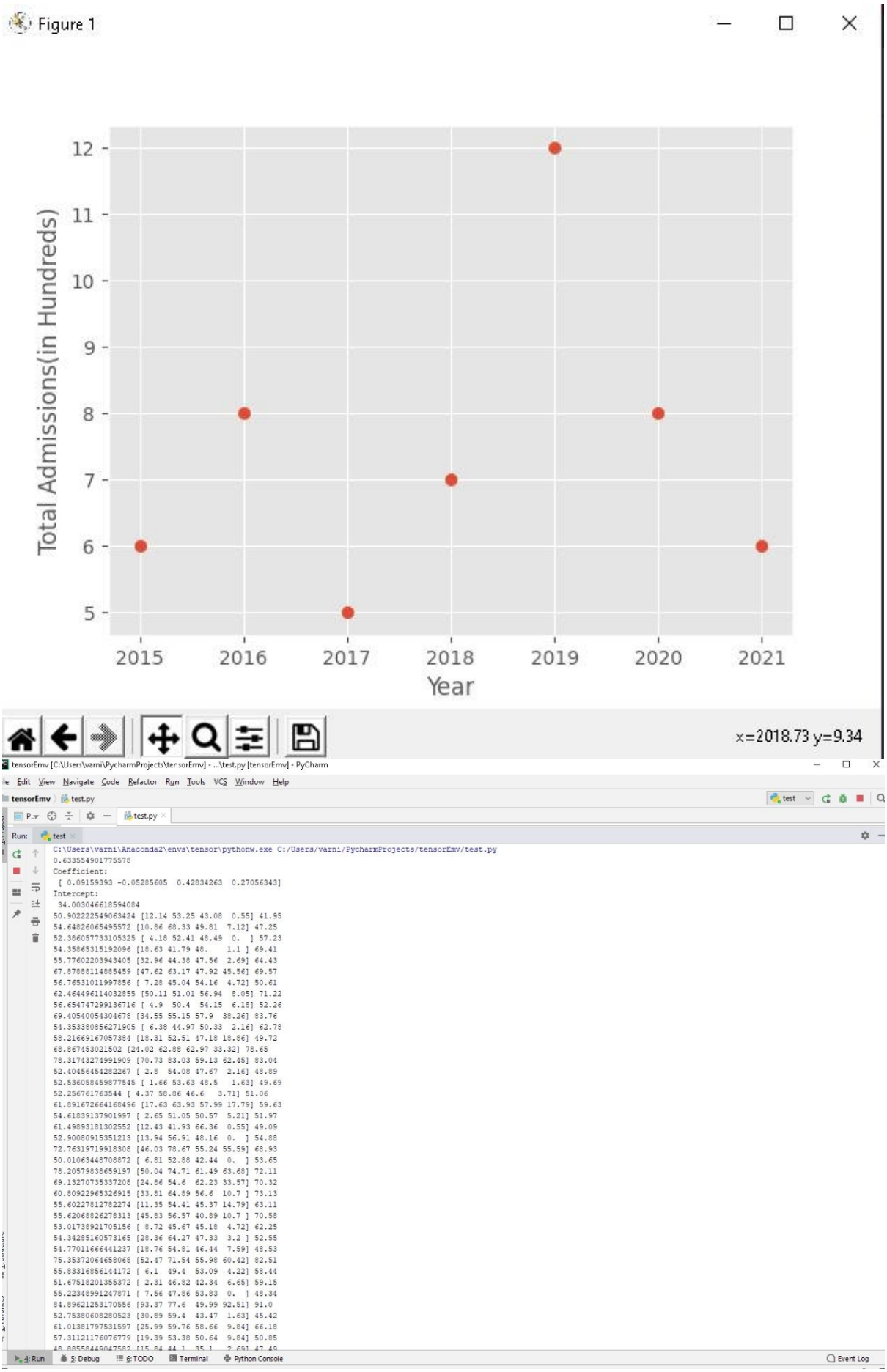
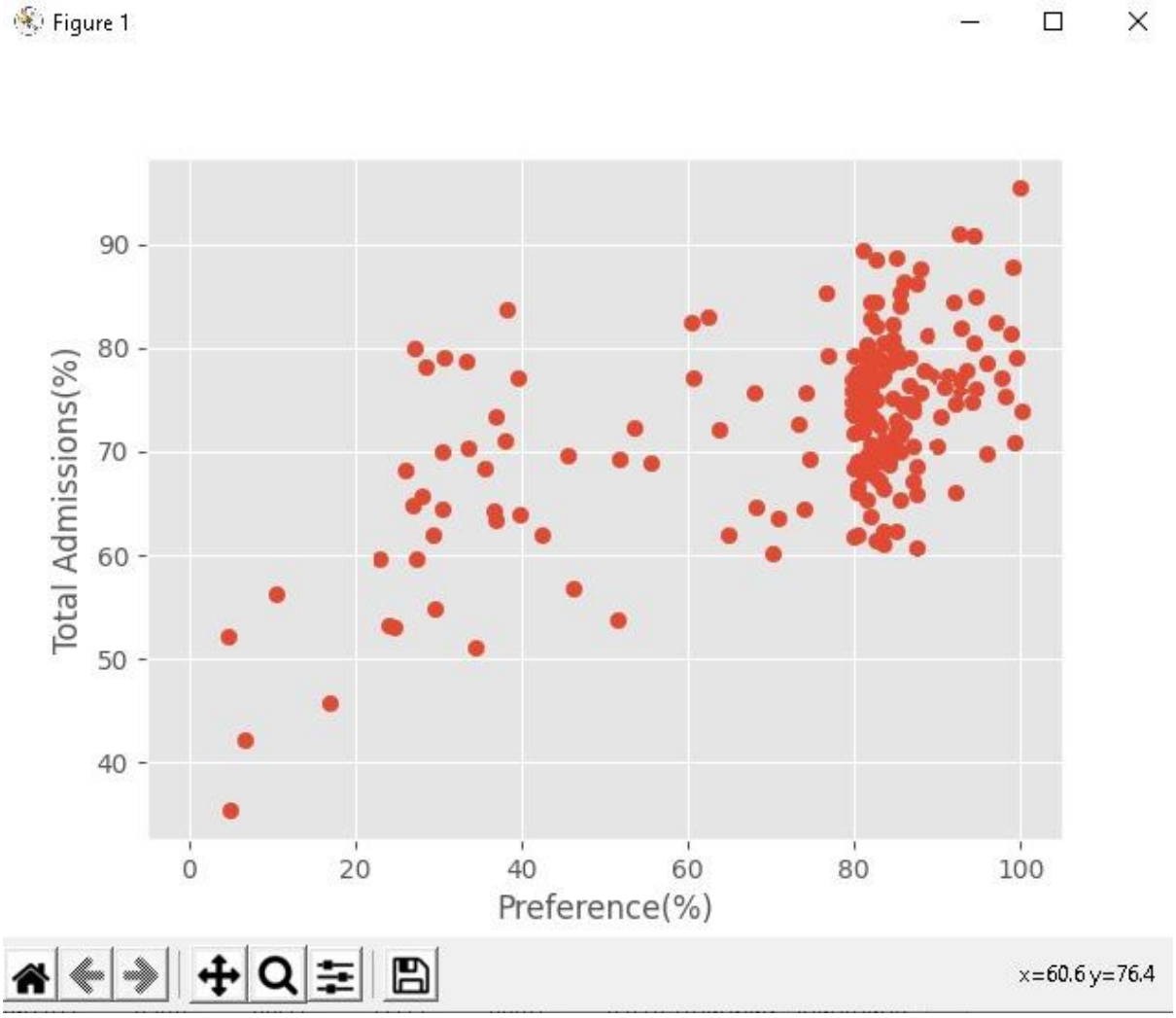


Figure 1
Total Admissions(in Hundreds)
Year
2015
2016
2017
2018
2019
2020
2021
x=2018.73 y=9.34
tensorEmv [C:\Users\varni\PycharmProjects\tensorEmv] - ...\test.py [tensorEmv] - PyCharm
C:\Users\varni\Anaconda2\envs\tensor\pythonw.exe C:/Users/varni/PycharmProjects/tensorEmv/test.py
0.633554901775578
Coefficient:
[0.09159393 -0.05285605 0.42834263 0.27056343]
Intercept:
34.0030466185940



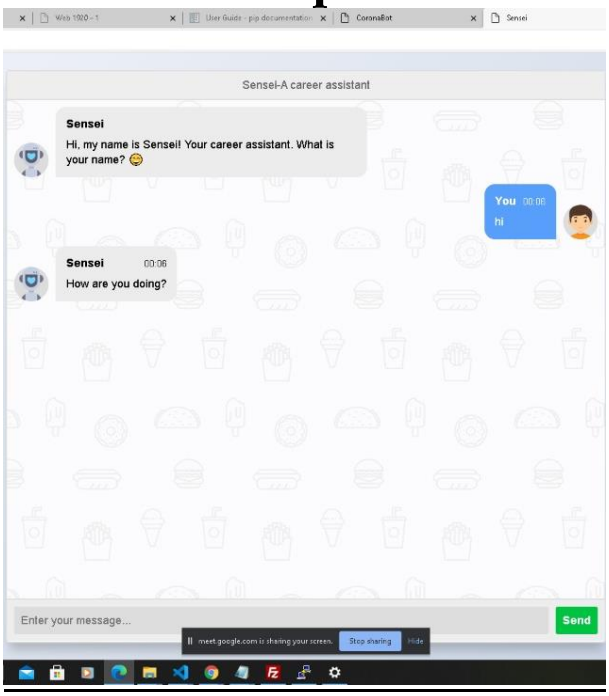
V. Predicting the response and giving output

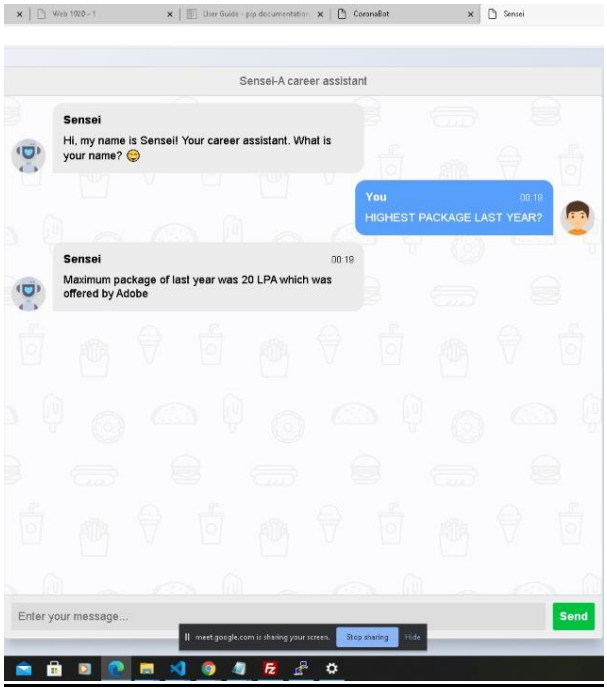
To predict the sentences and get a response from the user to let us create a new file 'chatapp.py'.

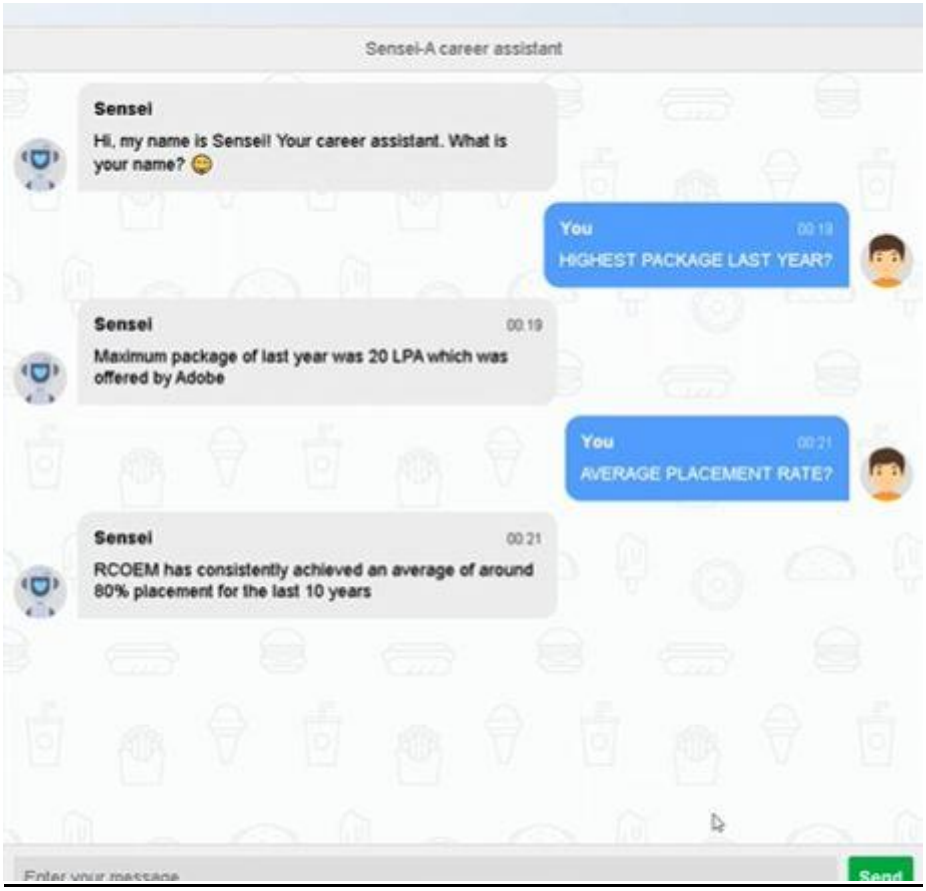
1. We will load the trained model and then use a graphical user interface that will predict the response from the bot.
2. so we will implement some functions which will identify the class and then retrieve us a random response from the list of responses.
3. Now create a new file app.py for flask server
4. Import flask and requirements.txt in your machine
5. In app.py import chatapp.py by 'from chatapp import chatapp'
6. Now create a index.html file for our frontend and store it in folder Templates
7. Style it with css, for this create a file style.css and write css command in it.

8. In index.html call our css file and display output
9. Go to http://localhost:5000/ and see the final output

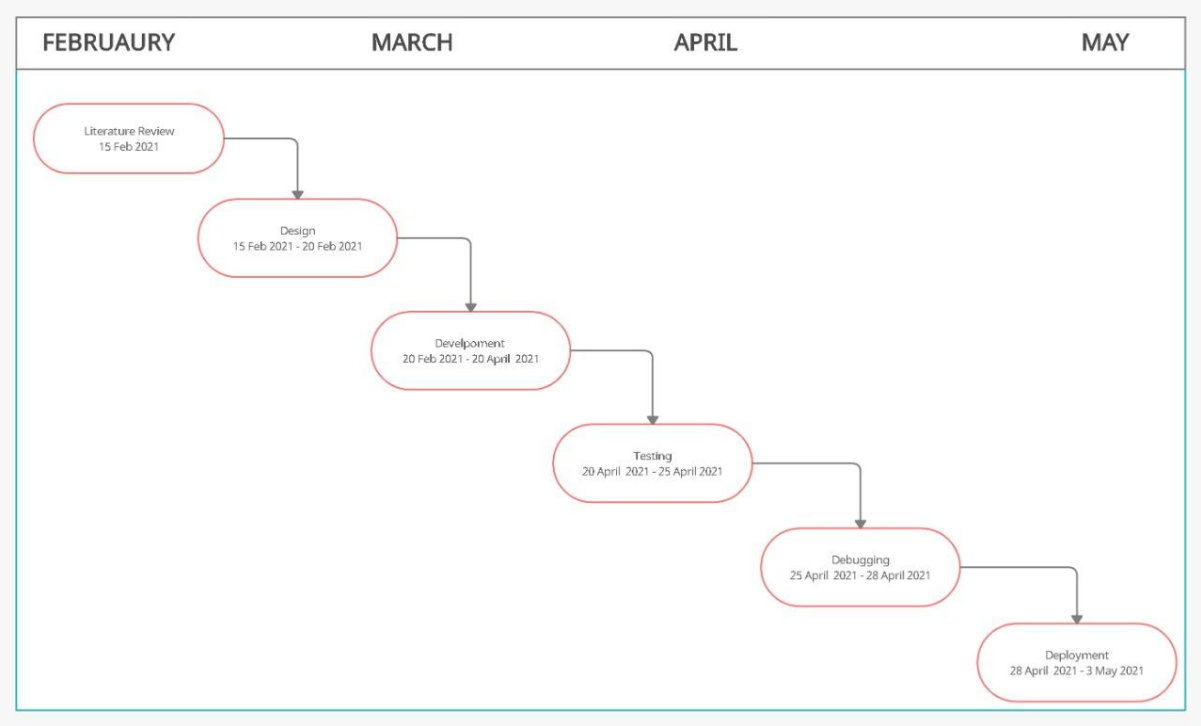
Output


Sensei-A career assistant
Sensei
Hi, my name is Sensei! Your career assistant. What is your name?
You 00:06
hi
Sensei 00:06
How are you doing?
Enter your message...
Send
meet.google.com is sharing your screen.
Stop sharing
Hide

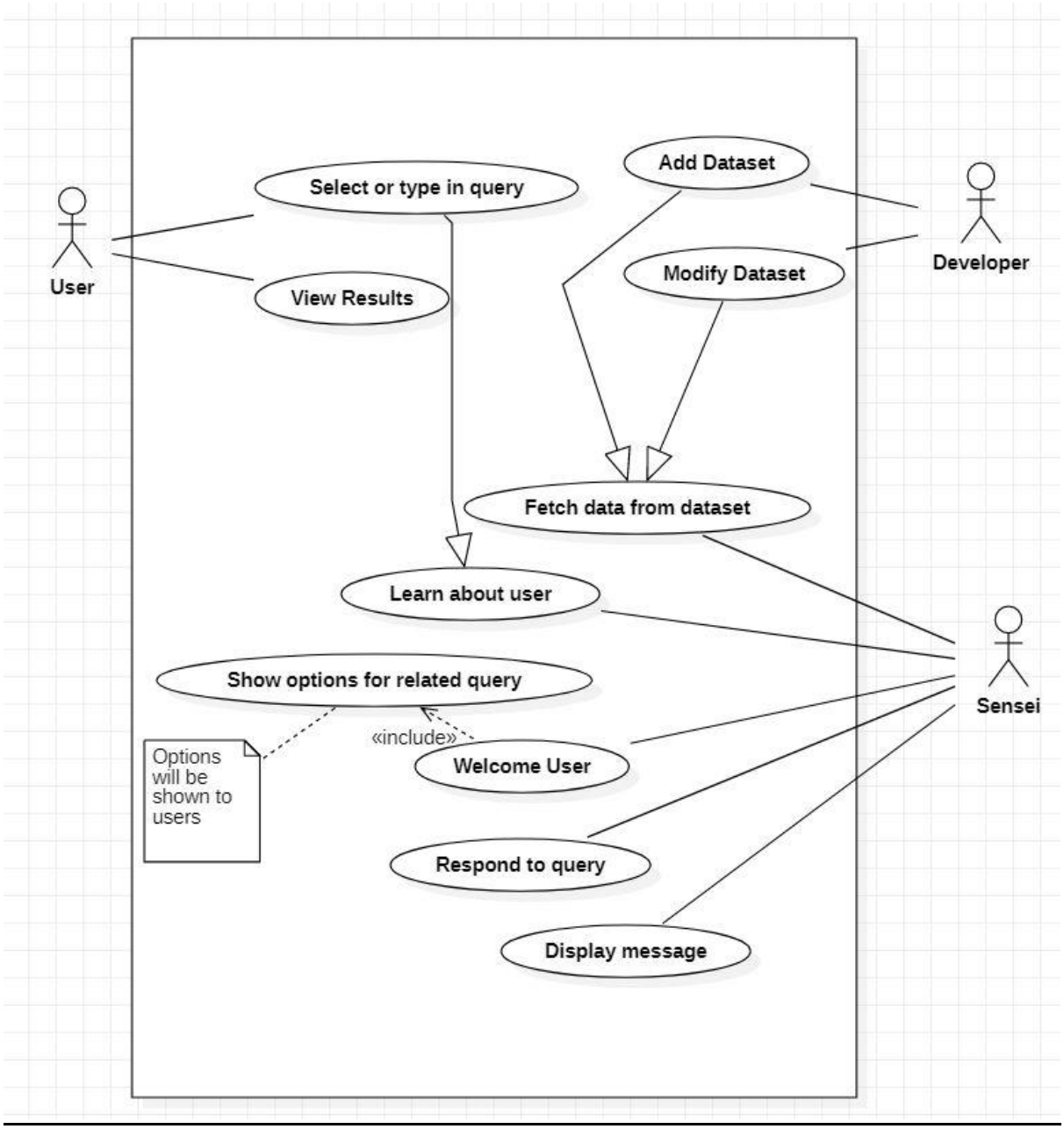

Sensei-A career assistant
Sensei
Hi, my name is Sensei! Your career assistant. What is your name?
You 00:19
HIGHEST PACKAGE LAST YEAR?
Sensei 00:19
Maximum package of last year was 20 LPA which was offered by Adobe
Enter your message...
Send
meet.google.com is sharing your screen.
Stop sharing
Hide


Sensei-A career assistant
Sensei
Hi, my name is Sensei! Your career assistant. What is your name?
You
HIGHEST PACKAGE LAST YEAR?
Sensei
00:19
Maximum package of last year was 20 LPA which was offered by Adobe
You
AVERAGE PLACEMENT RATE?
Sensei
00:21
RCOEM has consistently achieved an average of around 80% placement for the last 10 years
Send

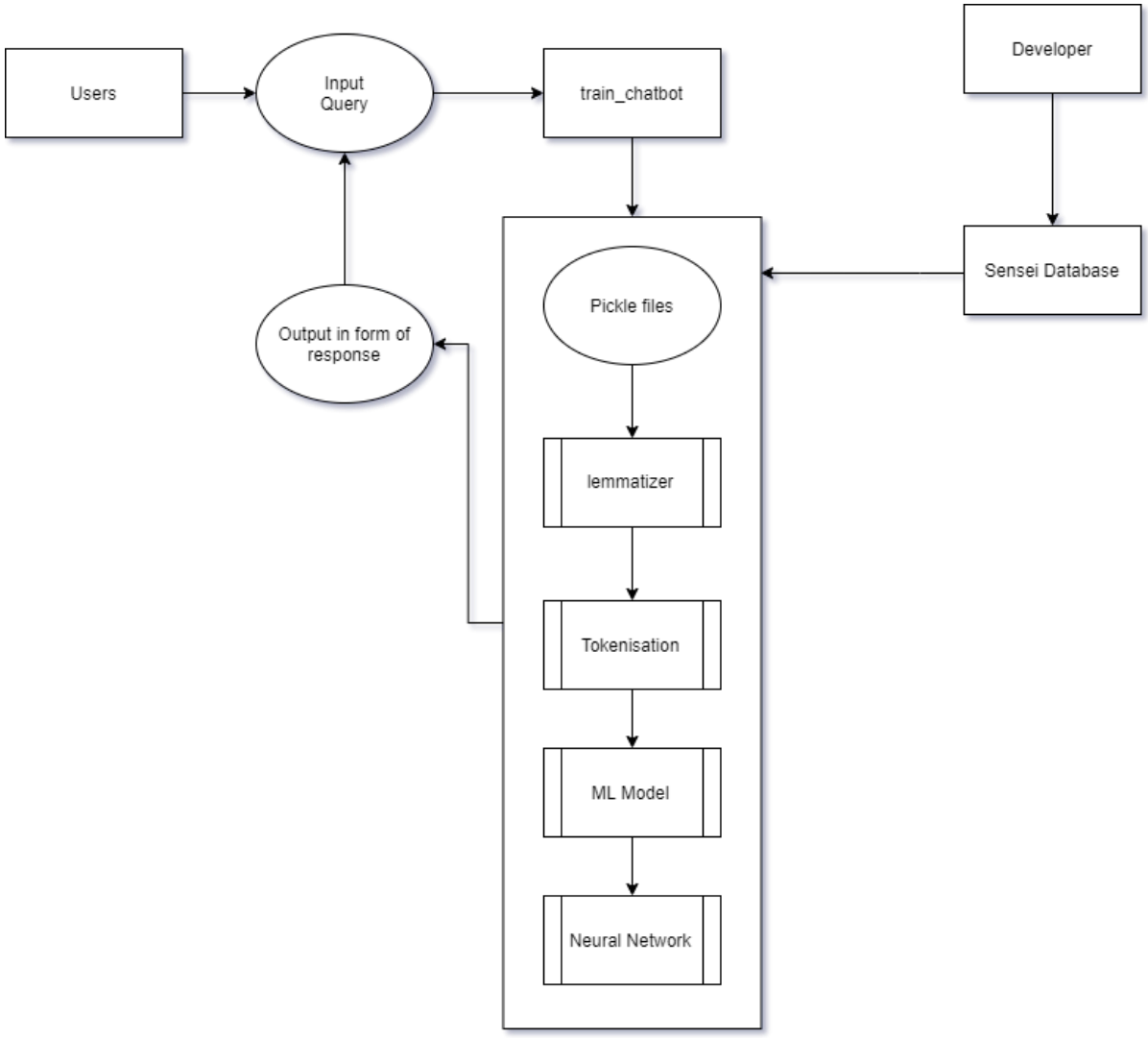
Schedule: (PERT Chart)



UML Diagram


Add Dataset
Select or type in query
Developer
User
Modify Dataset
View Results
Fetch data from dataset
Learn about user
Sensei
Show options for related query
«include»
Options will be shown to users
Welcome User
Respond to query
Display message

Dataflow Diagram



References:

1. Chatbot.com
2. Botsociety.com
3. Artificial Intelligence: A Modern Approach by Peter Norvig and Stuart J. Russell

* Whole Documents should not be more than 7 pages excluding Front Page
* The Front should contain Project Name, Partial Submission for Minor, Students name, Enrollment No, SAP Id no, Mentor Name

Approved By

Dr. Rohit Srivastava	**Dr. Monit Kapoor**
Assistance Professor	Head of Department Department of Cybernetics,
School of Computer Science	School of Computer Science
Project Guide	**Head of Department**