# INTEGRAL UNIVERSITY,LUCKNOW

# 

****

**Major Project Report**

Submitted in partial fulfilment for the award

Of the degree of BACHELOR OF TECHNOLOGY

In

Computer Science & Engineering

Project Title

iBOT

Under the supervision of Mr. AFTAAB YASEEN

**Submitted by :-**

**Name: Arshiya Dilshad**

**Roll no.:- 1801012026**

**Name: Athar Zahid Usmani**

**Roll no.:- 1801012032**

**Name: Aamna Akhtar**

**Roll no.:- 1801012003**

**Year 2021-2022**

**UNDERTAKING**

We, Aamna Akhtar, Arshiya Dilshad and Athar Zahid Usmani, students of B.Tech (CSE), hereby declare that the work detailed in this Project entitled “ **iBot** ” submitted to the Department of Computer Science and Engineering, Integral University, Lucknow for the award of the Bachelor of Technology degree is our original work. We have neither plagiarized nor submitted this work for the award of any other degree.

**Date:-** 12/05/2022

**Name:-**

Aamna Akhtar (Roll no. 1801012003)

Arshiya Dilshad (Roll no. 1801012026)

Athar Zahid Usmani (Roll.no 1801012032)

**STATEMENT OF ORIGINALITY**

This is to certify that to the best of my knowledge, the content of this project is our own work. This project has not been submitted for any degree or other purposes. I certify that the intellectual content of this project is the product of our work and that all the assistance received in preparing this project and sources have been acknowledged.

**Aamna Akhtar (sign)**

**Arshiya Dilshad**

**Athar Zahid Usmani**

**CERTIFICATE**

This is to certify that the project entitled, **“ iBot ”**, which is being submitted here with for the award of the degree of B.Tech in Computer Science and Engineering, is the result of the work completed by **Aamna Akhtar, Arshiya Dilshad** and **Athar Zahid Usmani** under my supervision and guidance within the four walls of the institute and the same has not been submitted elsewhere for the award of any degree.

DATE :- 12/05/2022

NAME OF THE SUPERVISOR :- Mr. Aftab Yaseen

SIGNATURE OF THE SUPERVISOR :-

……………………………………….

**RECOMMENDATION**

On the basis of the declaration submitted by Aamna Akhtar, Arshiya Dilshad and Athar Zahid Usmani, internal assessment carried out by the department on date 22/03/2022 and the certificate issued by the Project Supervisor- Mr. Aftab Yaseen, the project- “ iBot ” submitted to the department of Computer Science and Engineering, is recommended for final examination.

……………………………………. …………………………..

(Signature of Project Coordinator) (Signature of HOD)

**ACKNOWLEDGEMENT**

Thanking everybody is such a crucial task. And without performing that, any good work remains incomplete.

Every constructive work needs some fuel as its driving force. The fuel in our case was every little stone of help and appreciation that we received from our **Project Supervisor - Mr. Aftab Yaseen ,** our team mates and our parents all through the way while completing this skillfull project on- **“ iBot ”**

Thanking you

AAMNA AKHTAR

ARSHIYA DILSHAD

ATHAR ZAHID USMANI

**Chatbot**

Chatbots is also called chatterbots. It is a form of artificial intelligence (AI) used in messaging apps. This tool helps add convenience for customers they are automated programs that interact with customers like a human would and cost little to nothing to engage with.





**ABSTRACT-RESUME OF THE PROJECT**

iBot is the name of our chatbot. A chatbot is an intelligent piece of software that is capable of communicating and performing actions similar to a human. These days, chatbots are used almost everywhere, from customer interaction to marketing on social network sites and instantly messaging the client. A chatbot is a computer program that simulates and processes human conversation -either written or spoken, allowing humans to interact with digital devices as if they were communicating with a real person. Chatbots can be as simple as rudimentary programs that answer a simple query with a single-line response, or as sophisticated as digital assistants that learn and evolve to deliver increasing levels of personalization as they gather and process information.

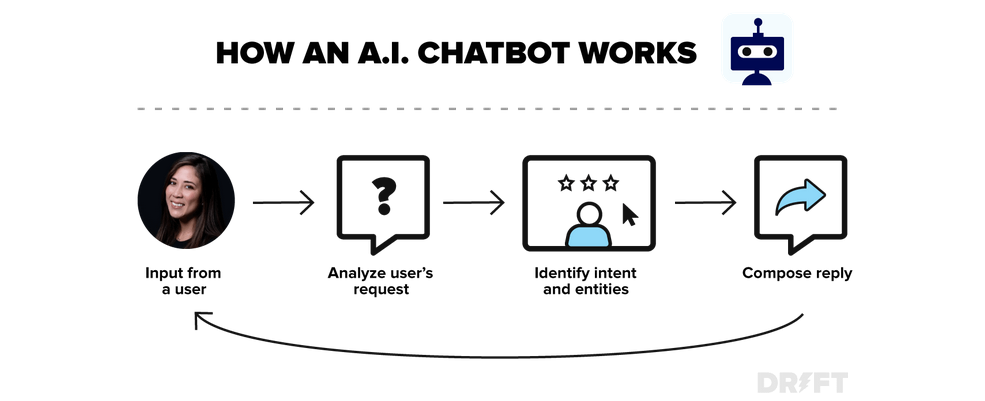
Chatbots allow businesses to connect with customers in a personal way without the expense of human representatives. For example, many of the questions or issues customers have are common and easily answered. That’s why companies create FAQs and troubleshooting guides. Chatbots provide a personal alternative to a written FAQ or guide and can even triage questions, including handing off a customer issue to a live person if the issue becomes too complex for the chatbot to resolve. Chatbots have become popular as a time and money saver for businesses and an added convenience for customers.

Chatbots are frequently used to improve the IT service management experience, which delves towards self-service and automating processes offered to internal staff. With an intelligent chatbot, common tasks such as password updates, system status, outage alerts, and knowledge management can be readily automated and made available 24/7, while broadening access to commonly used voice and text based conversational interfaces.

**How chatbot works**

The most typical chatbot interaction occurs on a business site. These customer service bots usually pop up after a human user navigates around a site for a few minutes or exhibits behaviors that show that they have become “lost” or are having trouble connecting with the information they need

Once the chatbot window presents itself, the user can enter their question in plain, syntactical English. [The bot’s language recognition functions break down the question](https://www.sofbang.com/utility-chatbot/) and, at the speed of light, compares the query to its data bank of previously asked questions to look for ways customers have achieved satisfying results in similar situations.



**REQUIREMENTS**

**Required moduls:-**

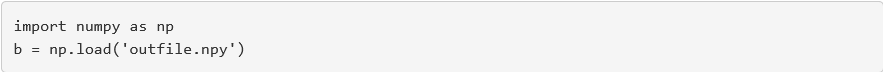
Following modules needs to be installed for it to work properly:\_

* Numpy
* Tenserflow
* Re
* Time

**Numpy:-**

It is a Python library that provides a multidimensional array object, various derived objects. NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices.

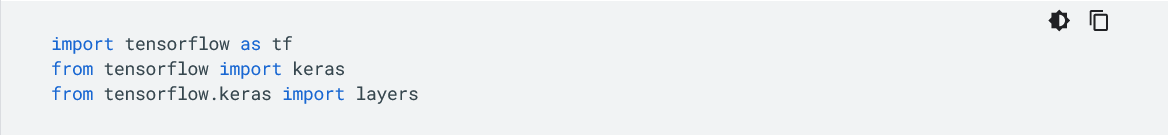
It can be installed using “pip install numpy”



**Tenserflow:-**

TensorFlow is a Python library for fast numerical computing created and released by Google. It is a foundation library that can be used to create Deep Learning models directly or by using wrapper libraries that simplify the process built on top of Tenserflow. We have used keras model for sequential model.

A Sequential model is appropriate for **a plain stack of layers** where each layer has **exactly one input tensor and one output tensor**.

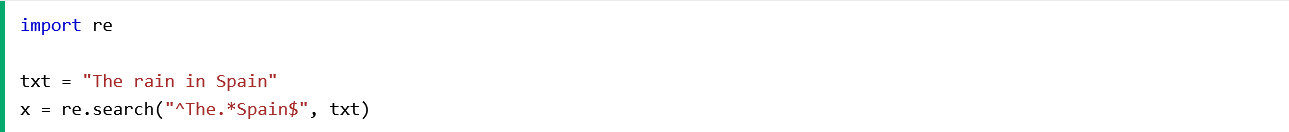


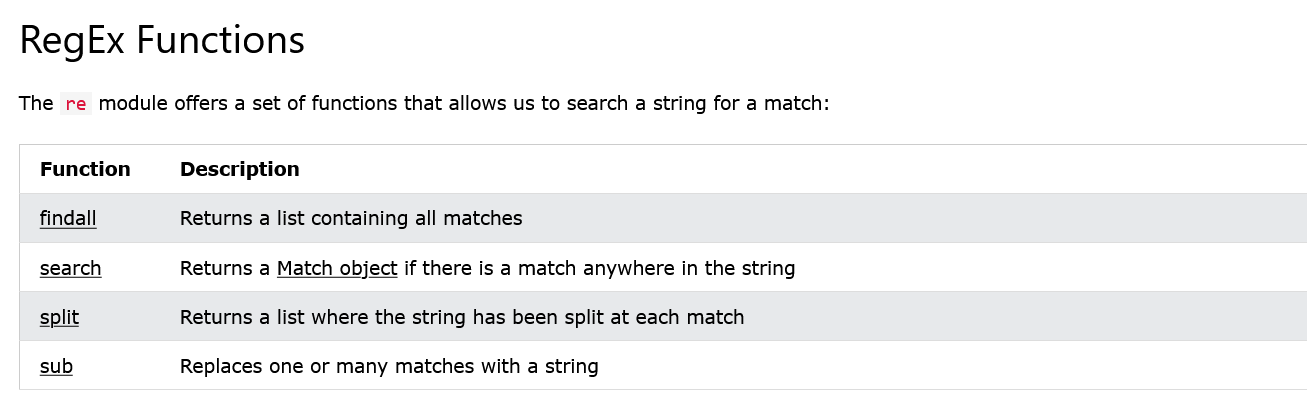
**Regular Expression (re) :-**

Regular expression (or RE) specifies a set of strings that matches it; the functions in this module let you check if a particular string matches a given regular expression (or if a given regular expression matches a particular string, which comes down to the same thing).

Regular expressions are a powerful language for matching text patterns. This page gives a basic introduction to regular expressions themselves sufficient for our Python exercises and shows how regular expressions work in Python. The Python "re" module provides regular expression support

Regular expressions can be concatenated to form new regular expressions; if *A* and *B* are both regular expressions, then *AB* is also a regular expression. In general, if a string *p* matches *A* and another string *q* matches *B*, the string *pq* will match AB. This holds unless *A* or *B* contain low precedence operations; boundary conditions between *A* and *B*; or have numbered group references. Thus, complex expressions can easily be constructed from simpler primitive expressions like the ones described here.





**Time:-**

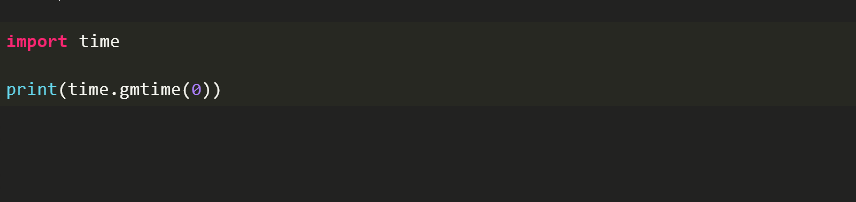
As the name suggests Python time module allows to work with time in Python. It allows functionality like getting the current time, pausing the Program from executing, etc. So before starting with this module we need to import it.

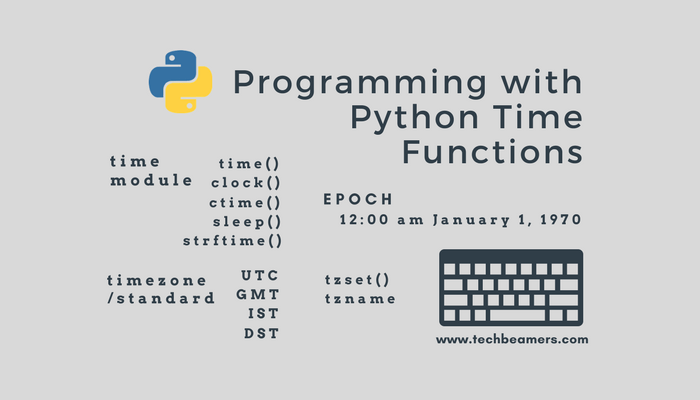
The Python time module provides many ways of representing time in code, such as objects, numbers, and strings. It also provides functionality other than representing time, like waiting during code execution and measuring the efficiency of your code.



**Epoch**

The epoch is the point where the time starts and is platform-dependent. On Windows and most Unix systems, the epoch is January 1, 1970, 00:00:00 (UTC), and leap seconds are not counted towards the time in seconds since the epoch. To check what the epoch is on a given platform we can use [time.gmtime(0)](https://www.geeksforgeeks.org/python-time-gmtime-method/).





**TECHNOLOGY USED**

Technology used in chatbot are:-

1. Artificial Intelligence(AI)
2. Natural Language Processing(NLP)
3. Deep Learning(DL)
4. Bag of words(BOW)
5. Seq2Seq

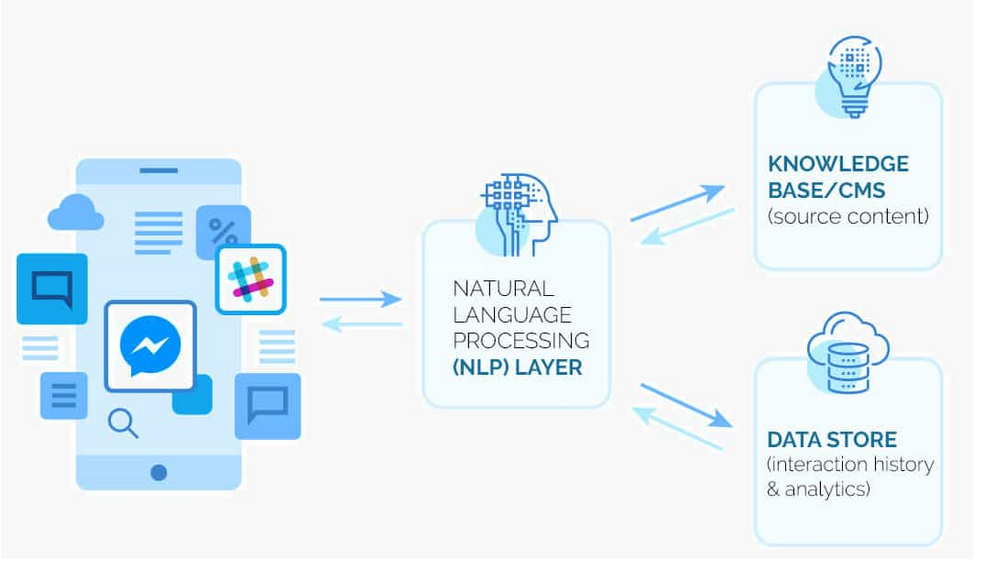
**Artificial Intelligence**

Chatbots is also called chatterbots. It is a form of artificial intelligence (AI) used in messaging apps. This tool helps add convenience for customers they are automated programs that interact with customers like a human would and cost little to nothing to engage with.

Artificial intelligence chatbots are text- or voice-based interfaces that provide support and connect human users with the services or information they need by simulating a traditional person-to-person conversation.

Text-based chatbots are often deployed online on websites and social media platforms to provide customer support and outreach. Voice-based chatbots, on the other hand, are most typically used for call deflection and sorting or over-the-phone customer service.

Most smartphones come equipped with a built-in chatbot, and smart speakers with chatbot functionality have been trendy gift-giving items for several years.



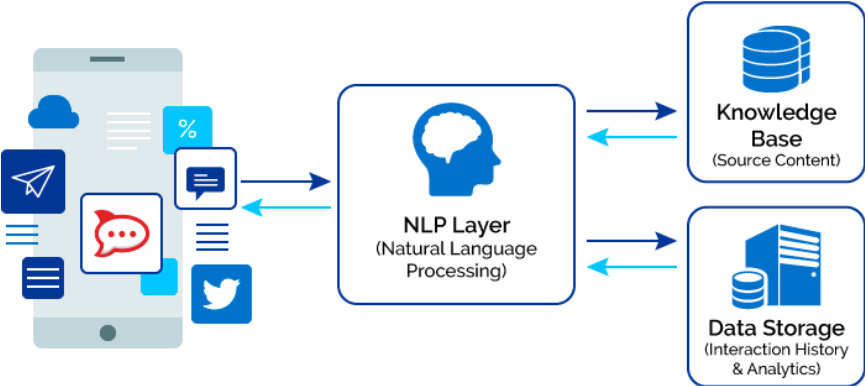
**Natural Language processing**

Natural language processing strives to build machines that understand and respond to text or voice data and respond with text or speech of their own in much the same way humans do.

Natural language processing (NLP) refers to the branch of computer science and more specifically, the branch of artificial intelligence concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

NLP combines computational linguistics rule-based modeling of human language with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text or voice data and to ‘understand’ its full meaning, complete with the speaker or writer’s intent and sentiment.

NLP drives computer programs that translate text from one language to another, respond to spoken commands, and summarize large volumes of text rapidly—even in real time. There’s a good chance you’ve interacted with NLP in the form of voice-operated GPS systems, digital assistants, speech-to-text dictation software, customer service chatbots, and other consumer conveniences. But NLP also plays a growing role in enterprise solutions that help streamline business operations, increase employee productivity, and simplify mission-critical business processes.

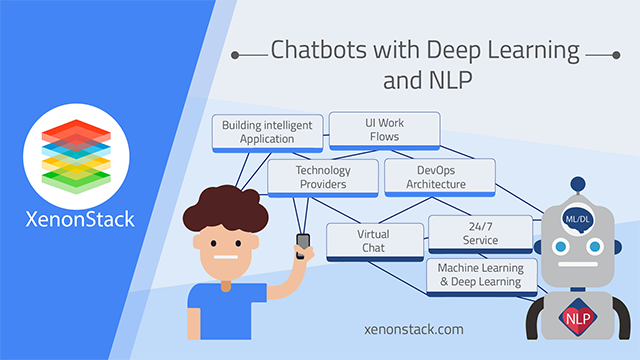


**Deep Learning:-**

**Deep Learning** is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called **artificial neural networks**.

**Deep learning** also known as **deep structured learning** is part of a broader family of [machine learning](https://en.wikipedia.org/wiki/Machine_learning) methods based on [artificial neural networks](https://en.wikipedia.org/wiki/Artificial_neural_network) with [representation learning](https://en.wikipedia.org/wiki/Representation_learning). Learning can be [supervised](https://en.wikipedia.org/wiki/Supervised_learning), [semi-supervised](https://en.wikipedia.org/wiki/Semi-supervised_learning) or [unsupervised](https://en.wikipedia.org/wiki/Unsupervised_learning).

Deep-learning architectures such as [deep neural networks](https://en.wikipedia.org/wiki/Deep_learning#Deep_neural_networks), [deep belief networks](https://en.wikipedia.org/wiki/Deep_belief_network), [deep reinforcement learning](https://en.wikipedia.org/wiki/Deep_reinforcement_learning), [recurrent neural networks](https://en.wikipedia.org/wiki/Recurrent_neural_networks) and [convolutional neural networks](https://en.wikipedia.org/wiki/Convolutional_neural_networks) have been applied to fields including [computer vision](https://en.wikipedia.org/wiki/Computer_vision), [speech recognition](https://en.wikipedia.org/wiki/Speech_recognition), [natural language processing](https://en.wikipedia.org/wiki/Natural_language_processing), [machine translation](https://en.wikipedia.org/wiki/Machine_translation), [bioinformatics](https://en.wikipedia.org/wiki/Bioinformatics), [drug design](https://en.wikipedia.org/wiki/Drug_design), [medical image analysis](https://en.wikipedia.org/wiki/Medical_image_analysis), [climate science](https://en.wikipedia.org/wiki/Climatology), material inspection and [board game](https://en.wikipedia.org/wiki/Board_game) programs, where they have produced results comparable to and in some cases surpassing human expert performance.



**Bag of words:-**

The **bag-of-words** (BOW) model is a representation that turns arbitrary text into **fixed-length vectors** by counting how many times each word appears. This process is often referred to as **vectorization.**

Let’s understand this with an example. Suppose we wanted to vectorize the following:

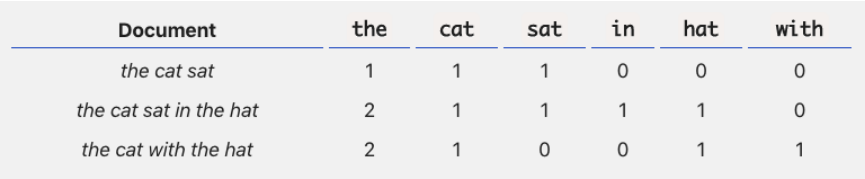
* the cat sat
* the cat sat in the hat
* the cat with the hat

# Step 1: Determine the Vocabulary

We first define our **vocabulary**, which is the set of all words found in our document set. The only words that are found in the 3 documents above are: the, cat, sat, in, the, hat, and with.

# Step 2: Count

To vectorize our documents, all we have to do is **count how many times each word appears**:



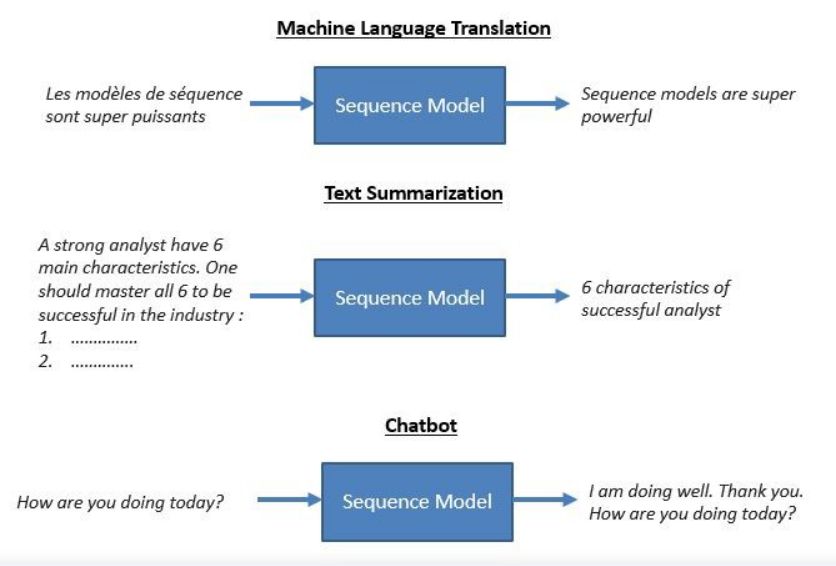
Now we have length-6 vectors for each document!

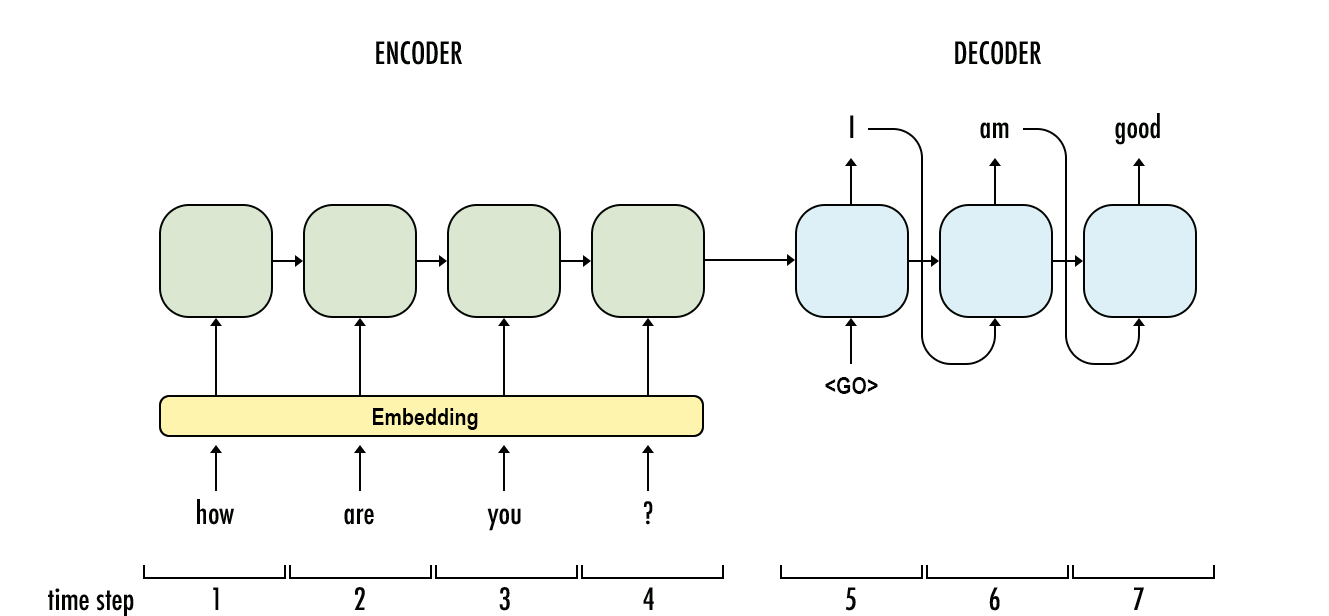
* the cat sat*:* [1, 1, 1, 0, 0, 0]
* the cat sat in the hat*:* [2, 1, 1, 1, 1, 0]
* the cat with the hat*:* [2, 1, 0, 0, 1, 1]

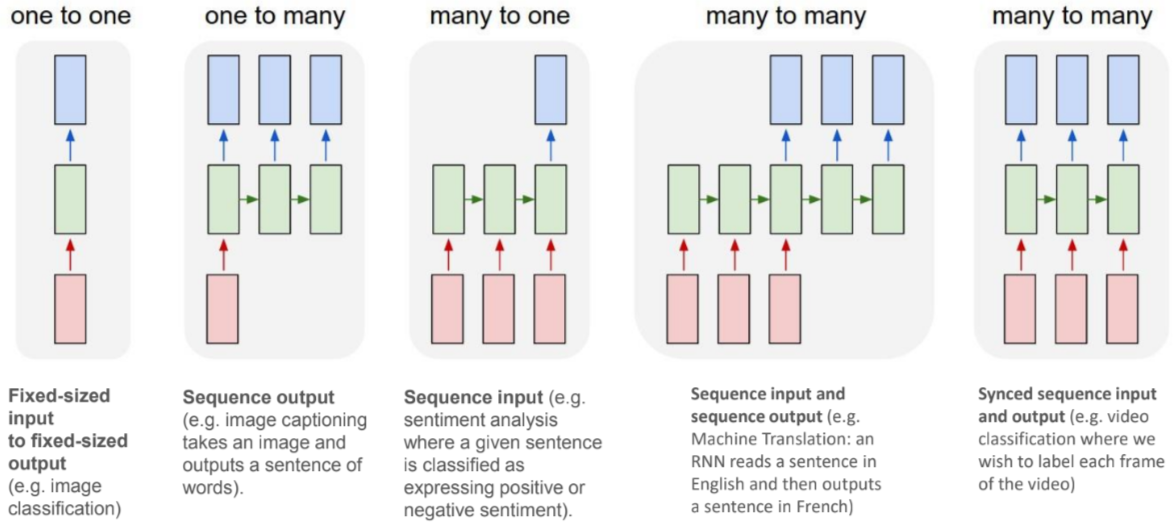
Notice that we lose contextual information, e.g. where in the document the word appeared, when we use BOW. It’s like a literal **bag**-of-words: it only tells you what words occur in the document, not where they occurred.

**Seq2Seq:-**

Sequence to Sequence (often abbreviated to seq2seq) models is a special class of Recurrent Neural Network architectures that we typically use (but not restricted) to solve complex Language problems like Machine Translation, Question Answering, creating Chatbots, Text Summarization, etc.

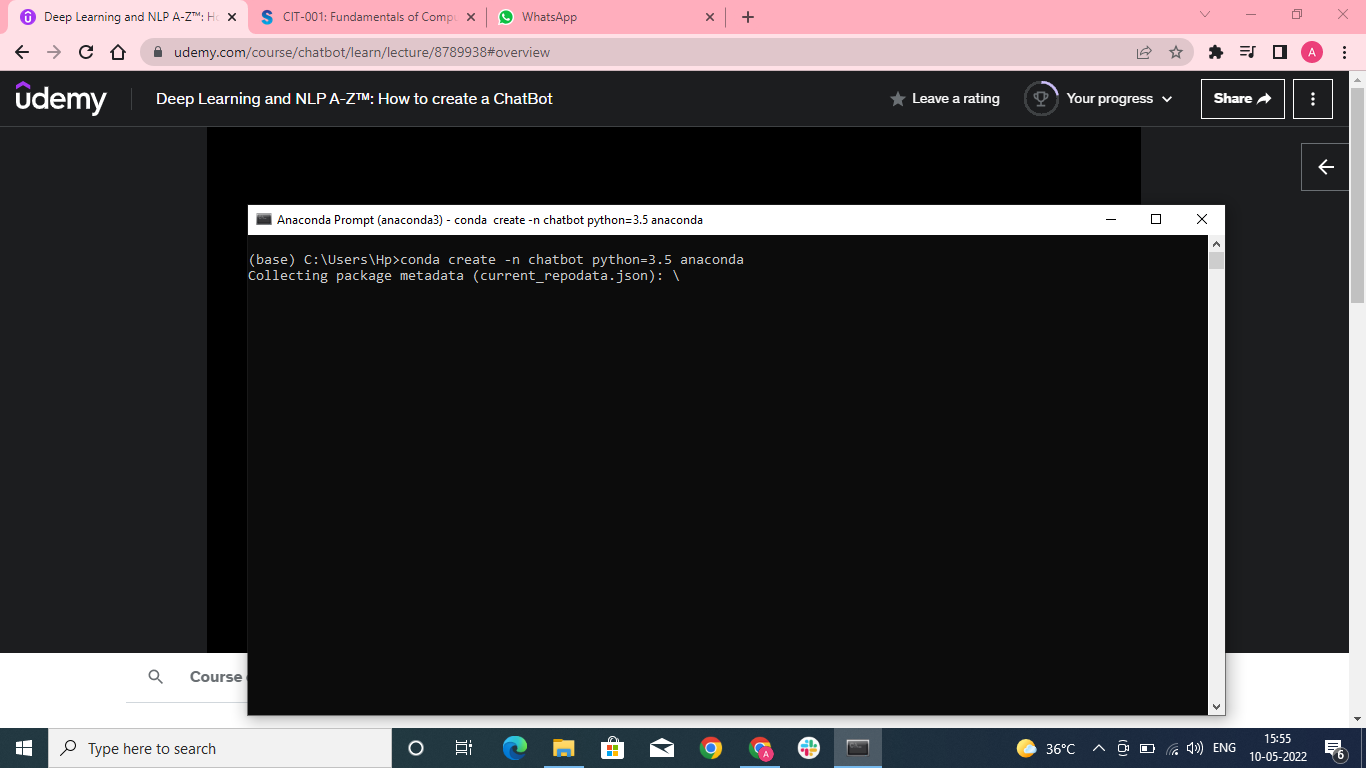


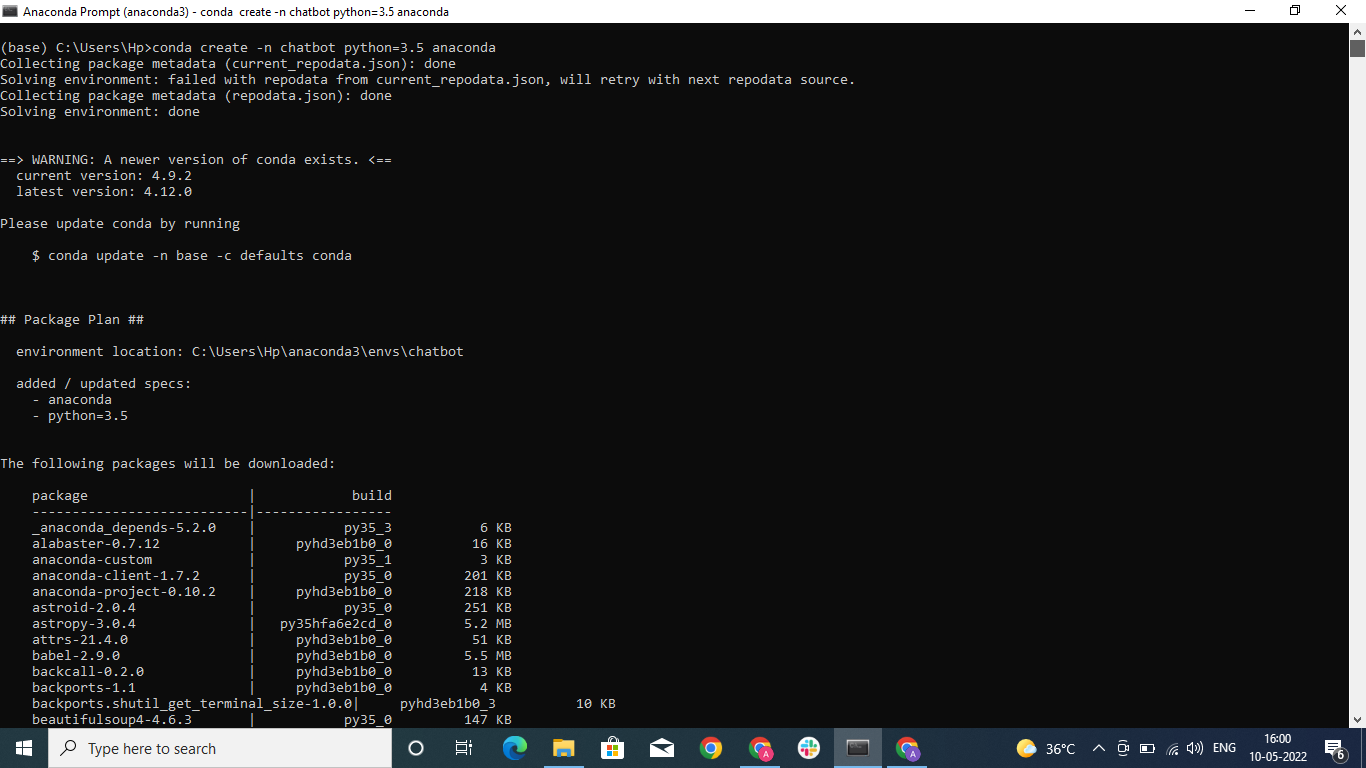




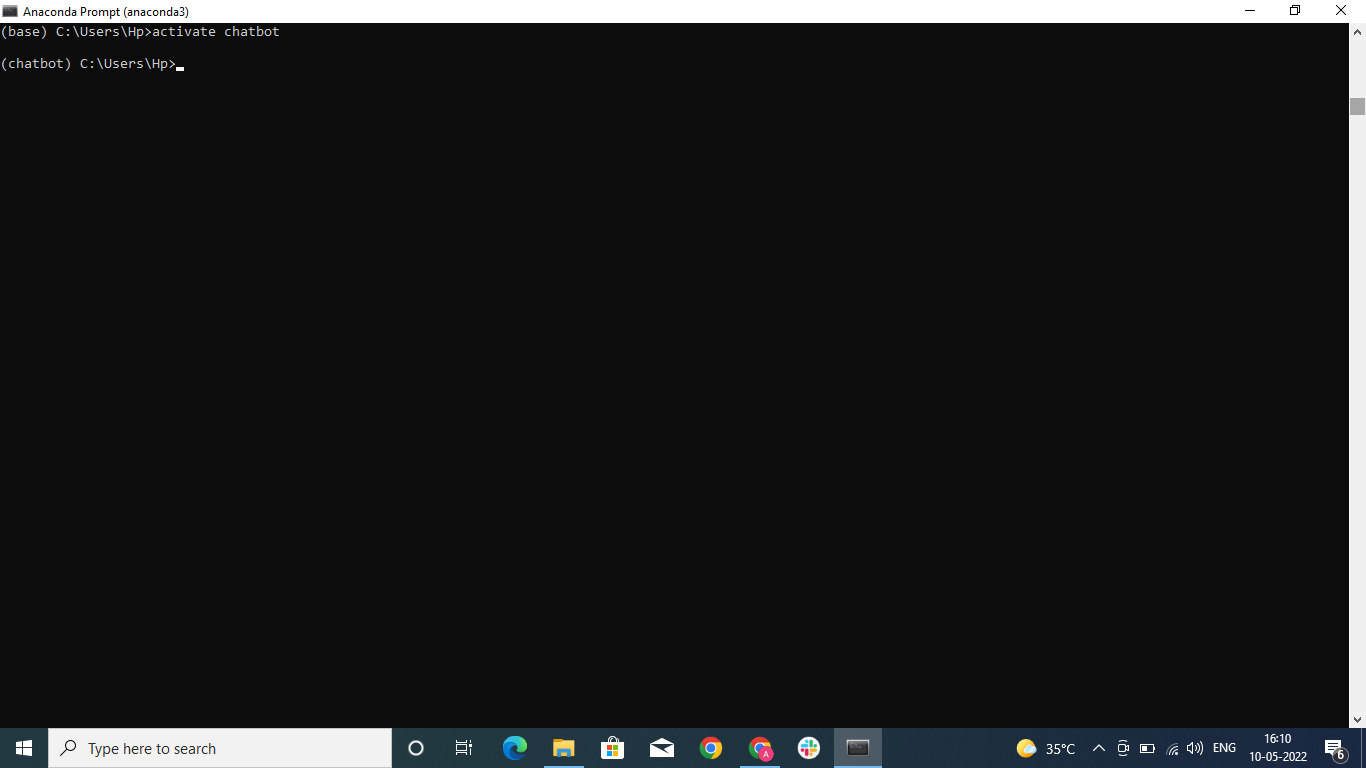
**Building A chatbot using Deep NLP**

**Step 1:-** Creating environment for chatbot.

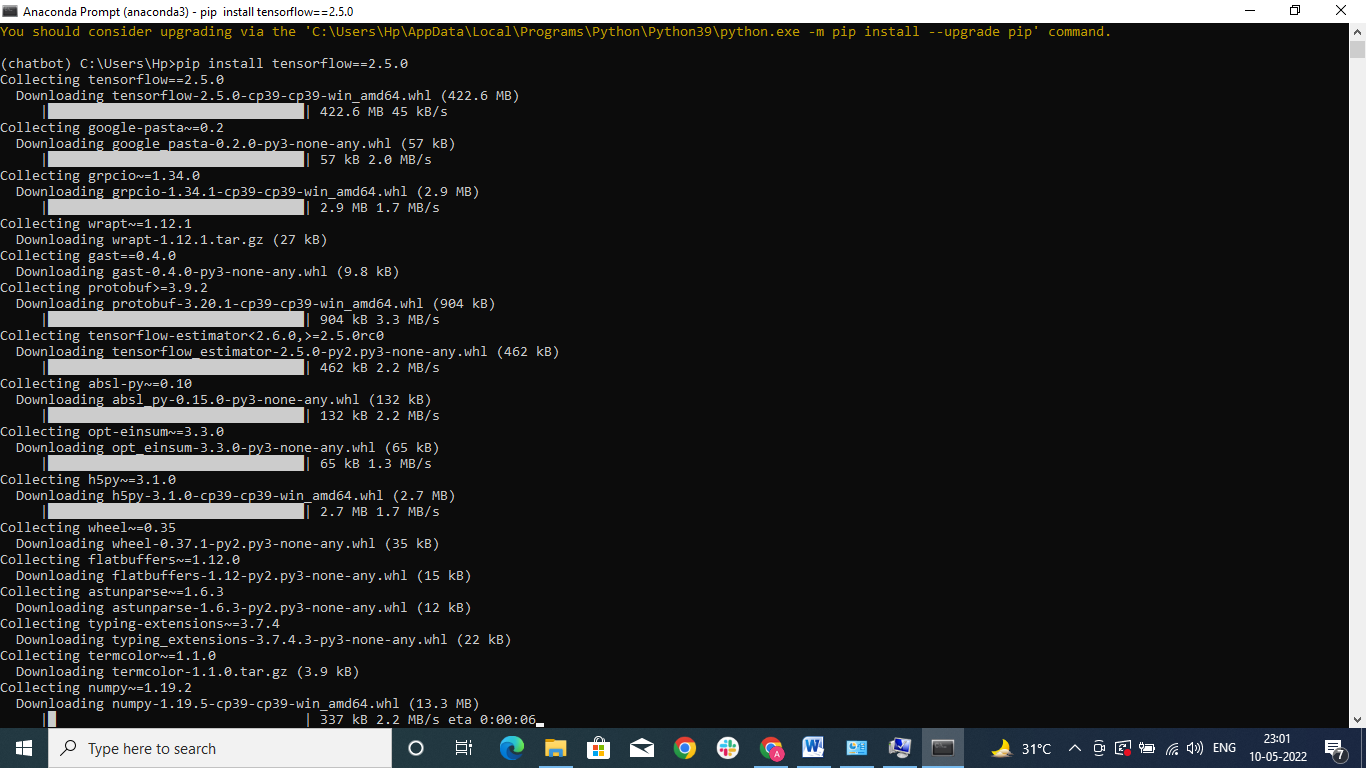




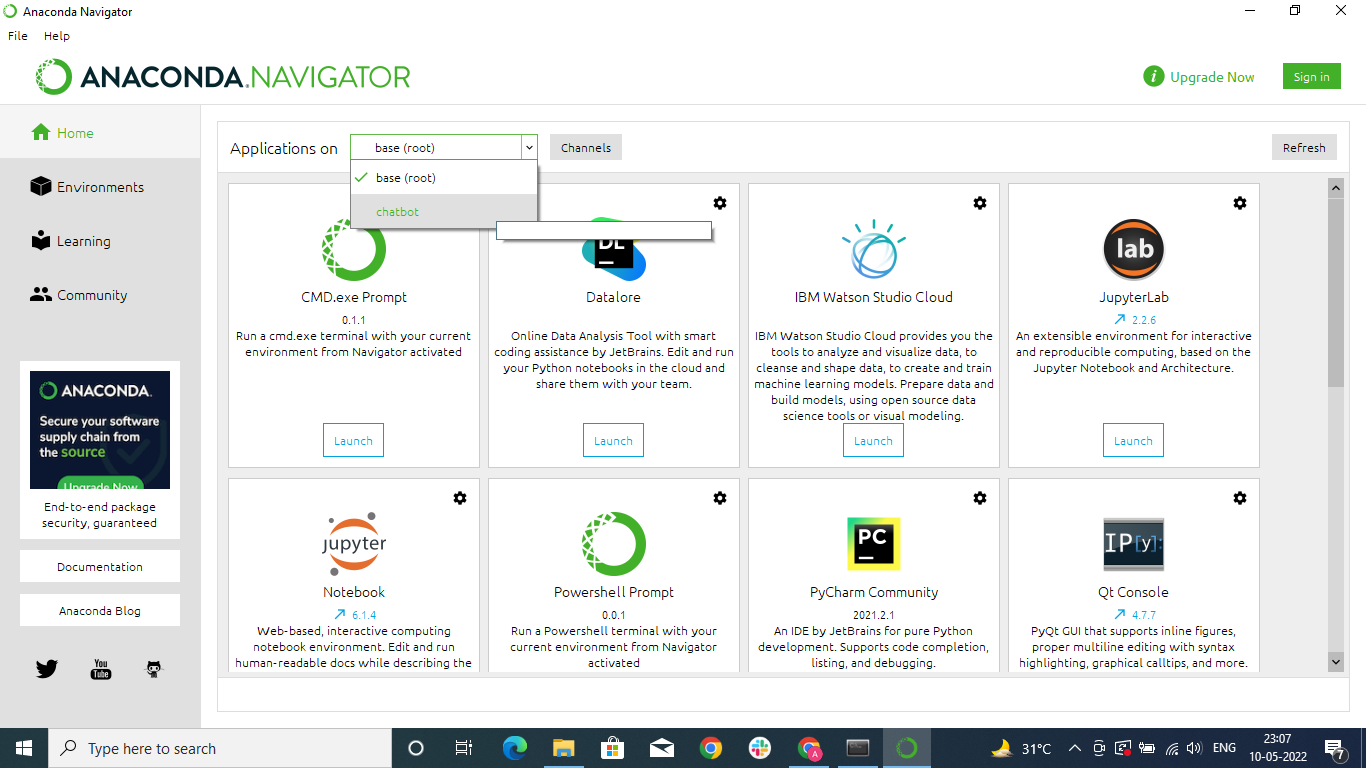
We created a virtual environment name chatbot.



Chatbot environment has been created and we are now in the chatbot environment.

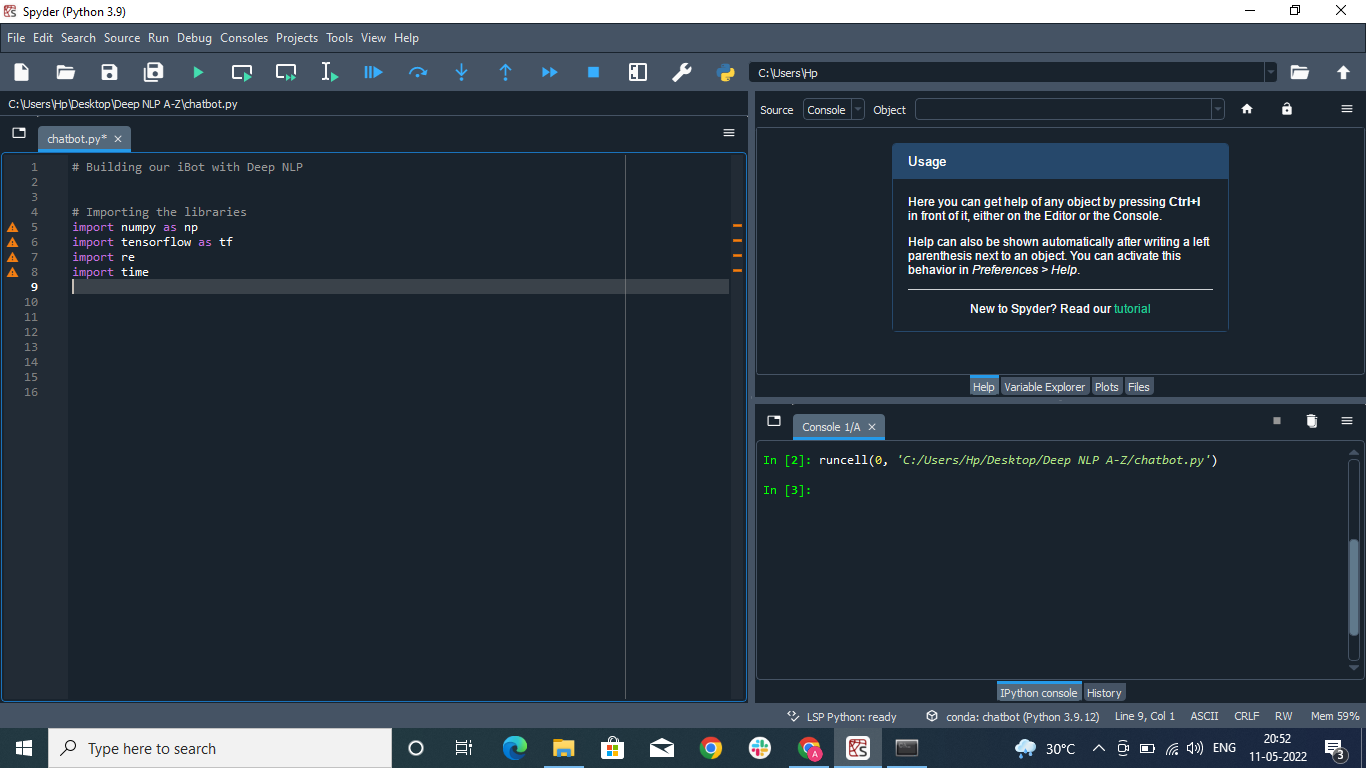


Installing tensorflow in the chatbot environment.

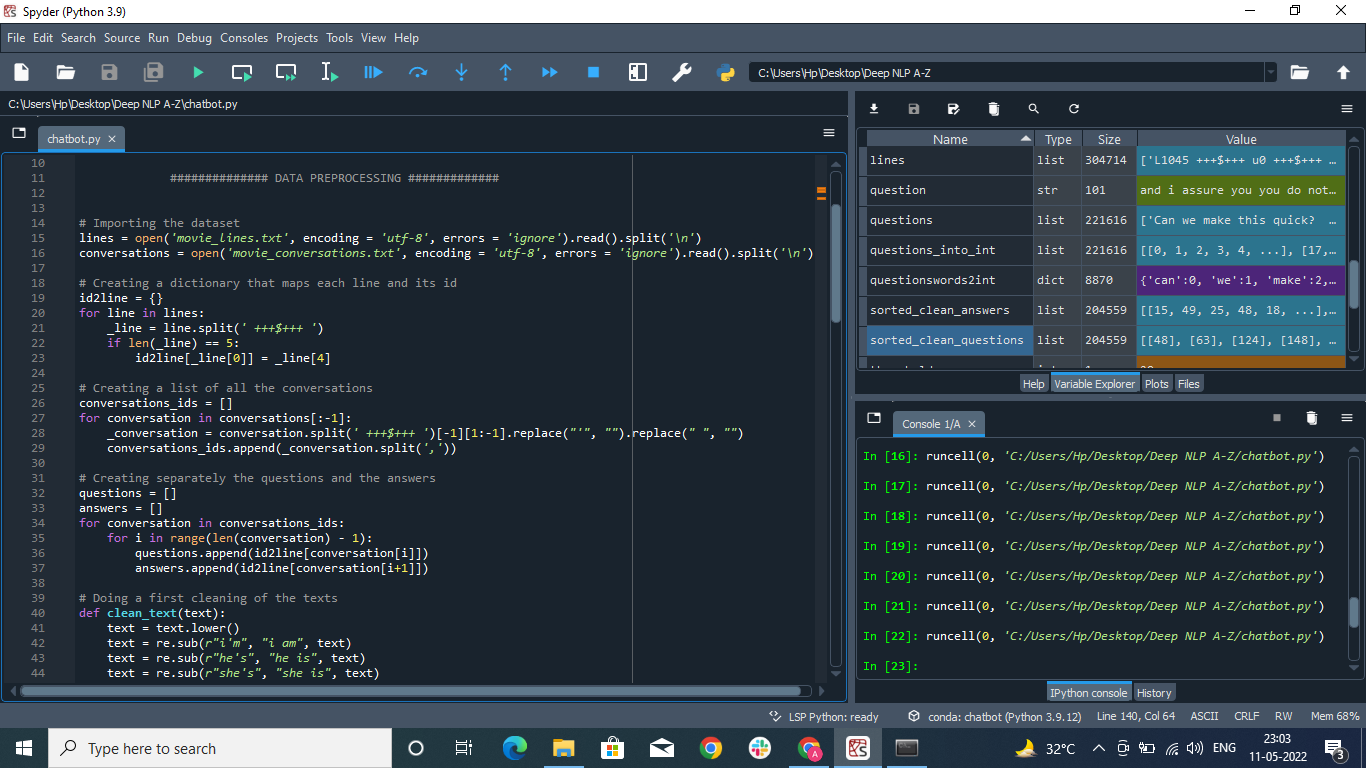


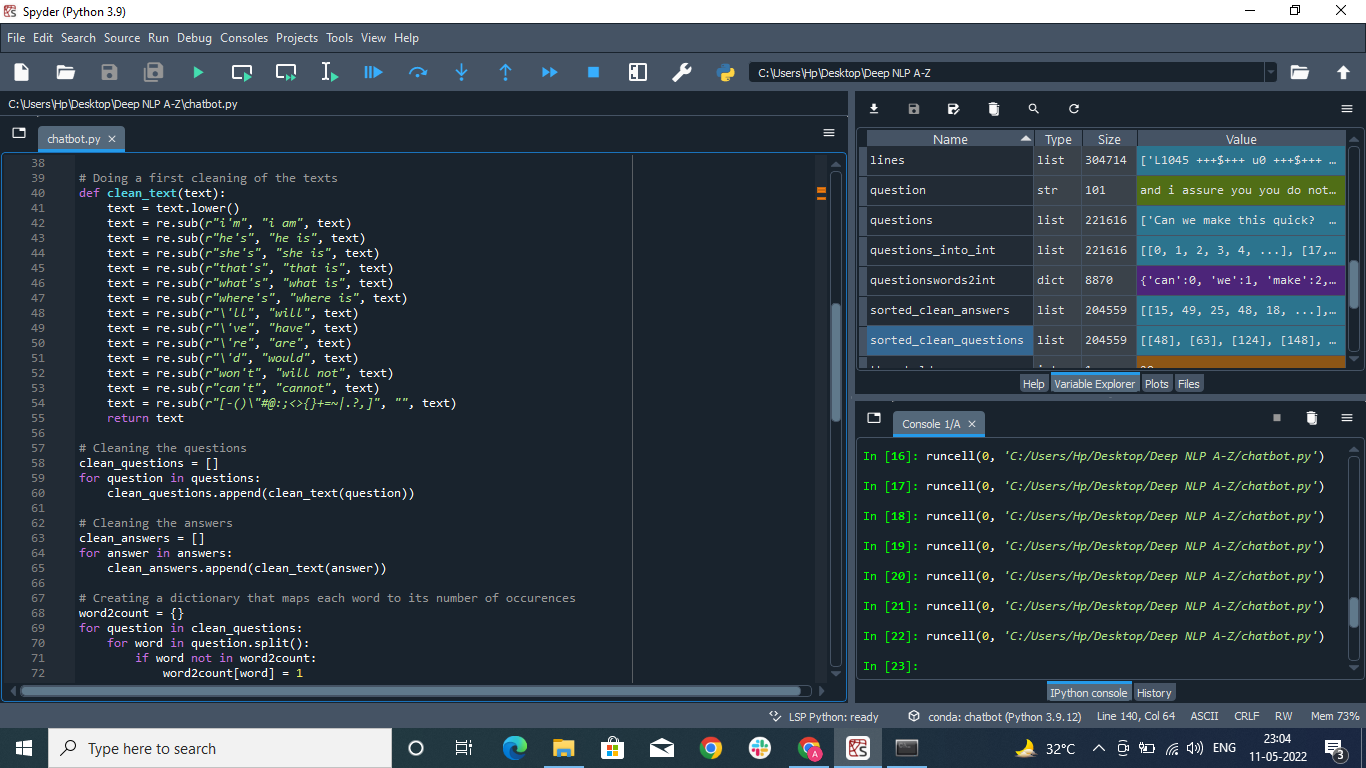
Now opening chatbot environment in Anaconda Navigator in Spyder.

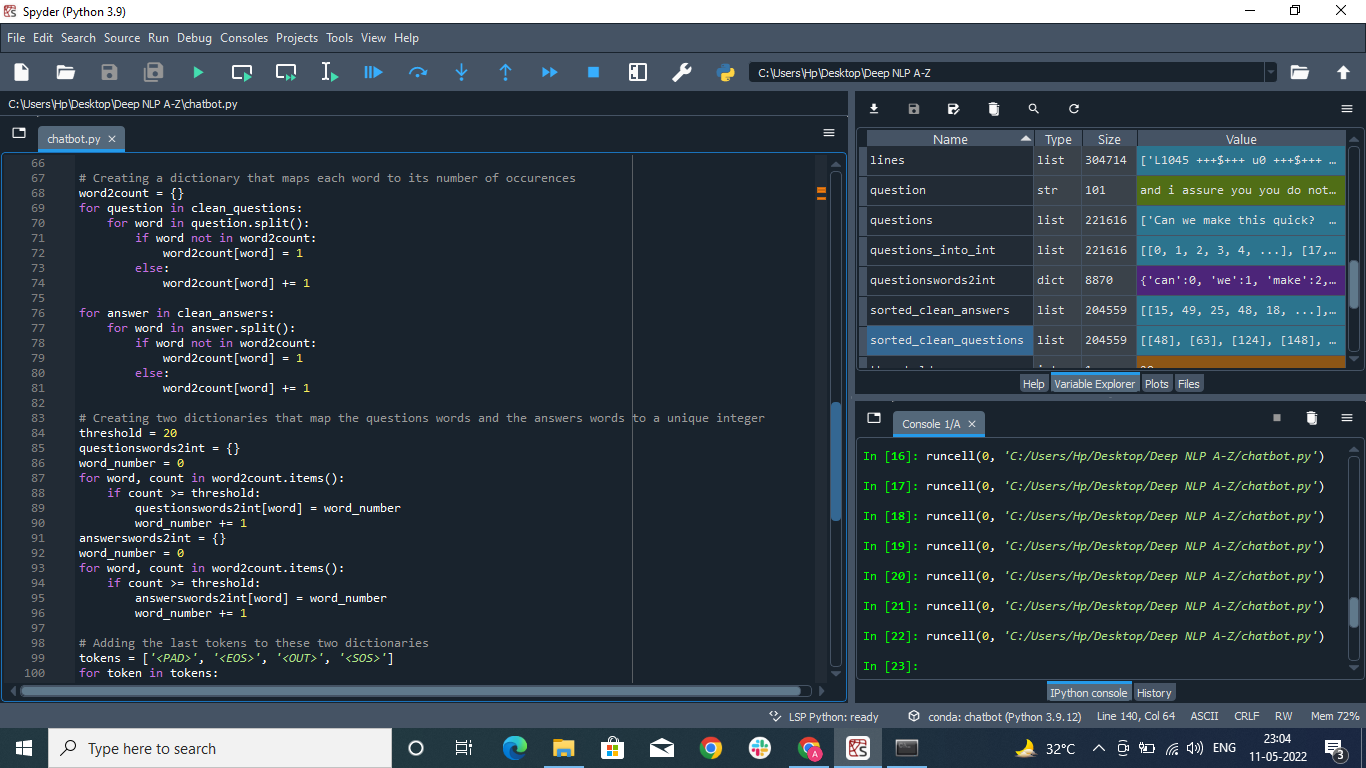
**Step 2:-** Importing the required libraries.

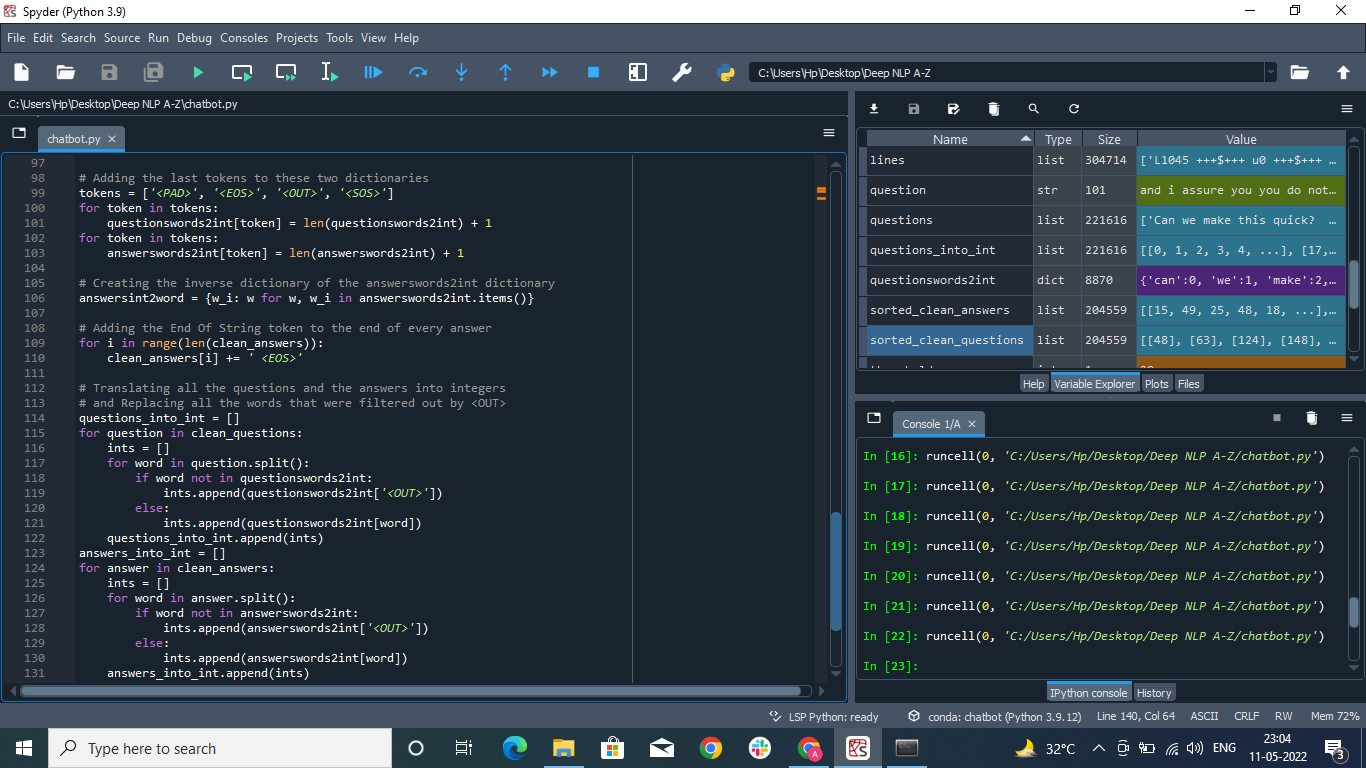


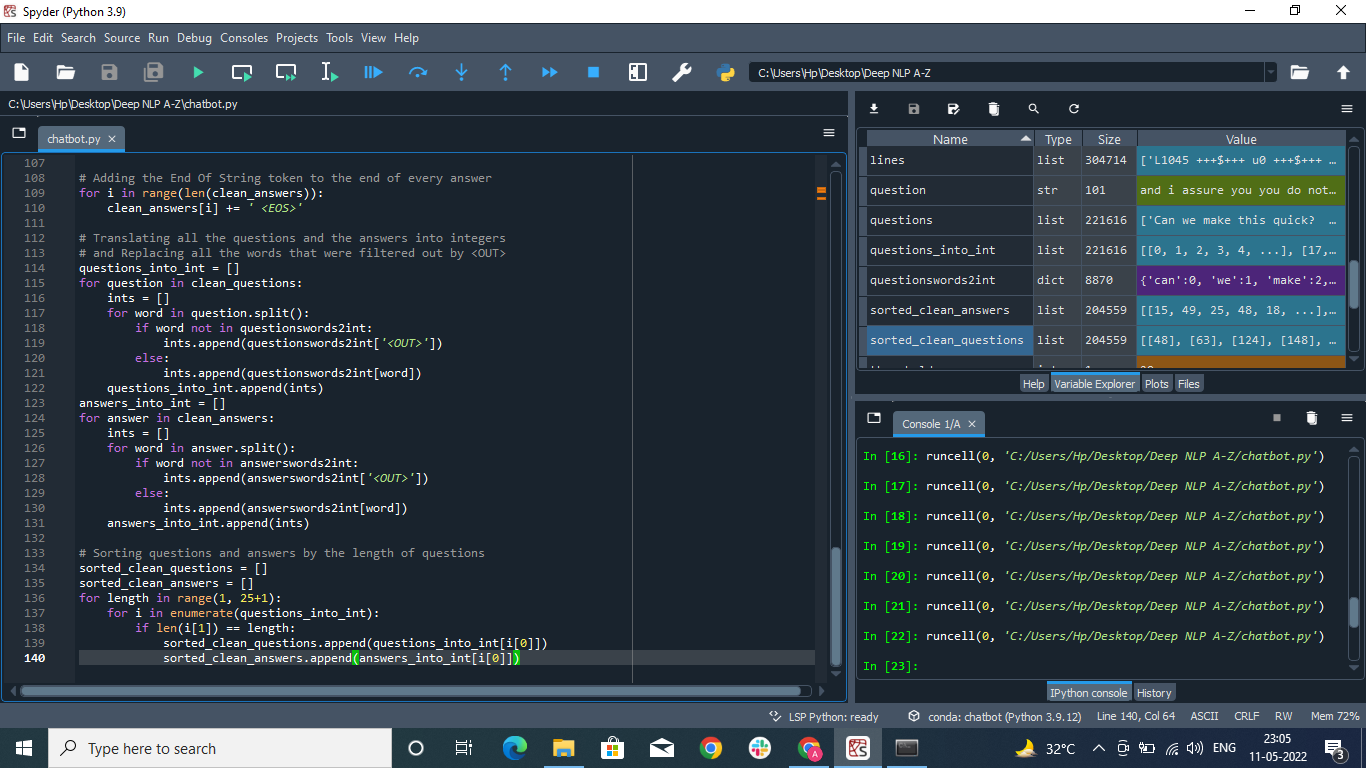
**Step 3:-** Data preprocessing











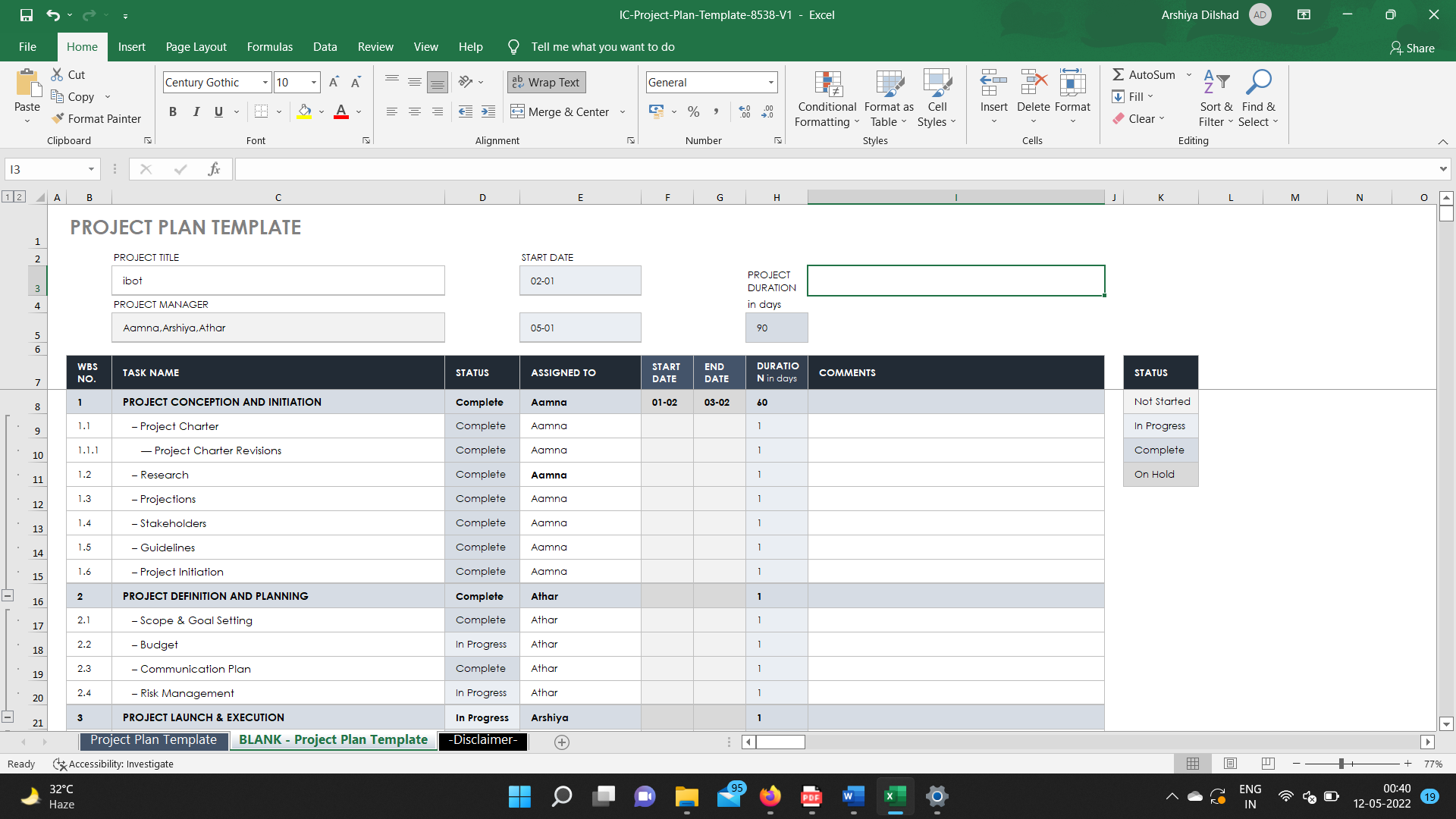
**PROJECT PROGRESS**

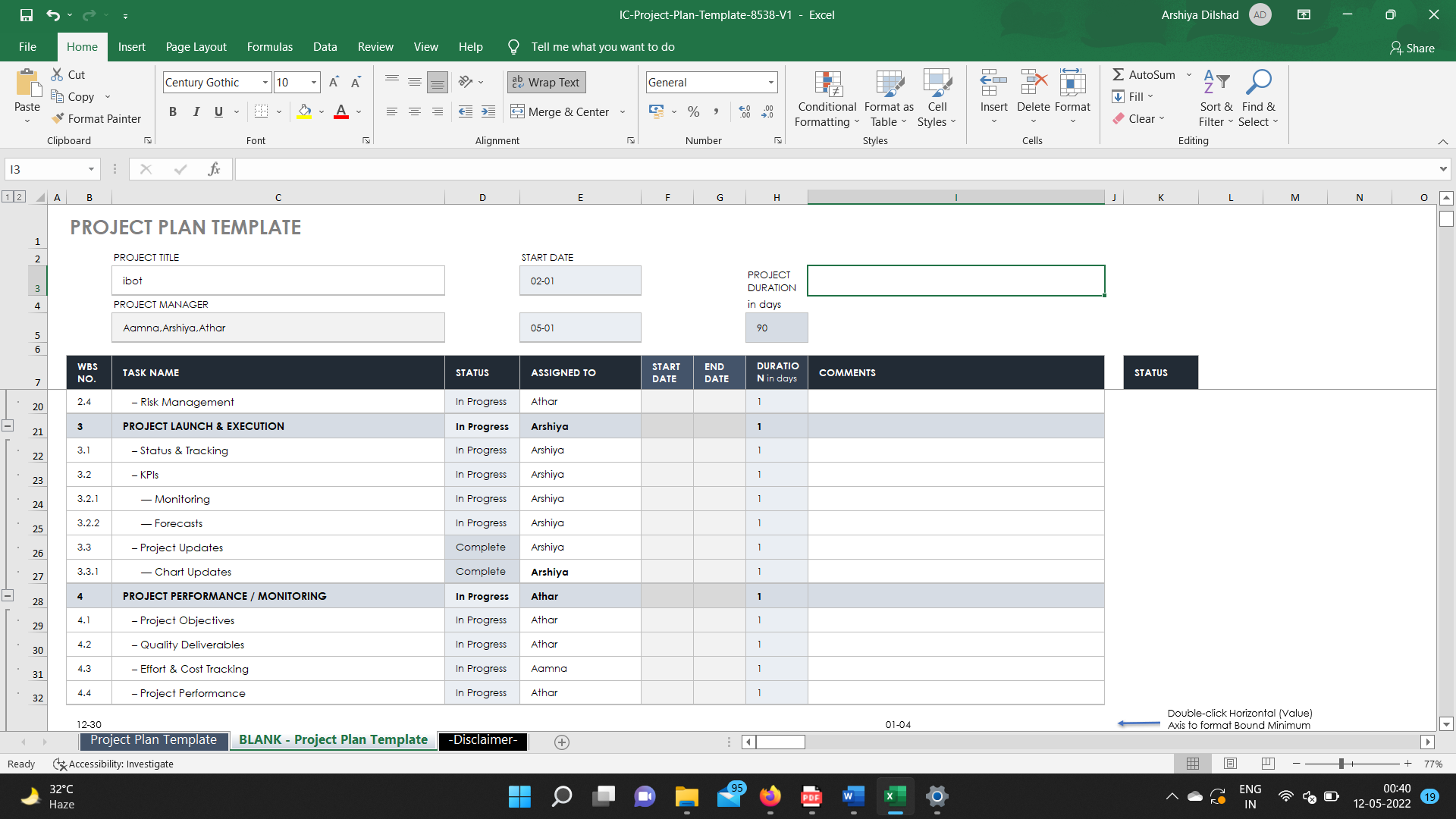
Initially, we tried to build our iBot using Tensorflow library and intents.json dataset. But the problem with that model came out while interacting with users. iBot does not seem to be well trained and accurate in its responses.

So, as an improvement to our bot, we are trying a totally new model using Deep Learning, NLP (Natural Language Processing), Bags of Word Model and Seq2Seq Model. Besides, this time instead of working on intents.json dataset we have taken a Cornell Movie--Dialogs Corpus. The reason for this is that in a movie there are lots and lots of conversations taking place, so modelling the bot on such a big dataset would improve the training of the bot. As far as project progress is concerned, we are done with Data Pre-processing part of it. We are hoping to get a better model with these steps and can successfully make a powerful chatbot.

We are also trying different neural network architectures with different hyperparameters.

And at last but not the least, we would also consider training our iBot with emojis when building the final model.





**ROLES OF THE MEMBERS INDIVIDUALLY**

**Aamna Akhtar:-**

Backened and model selection with better prediction results.

**Athar Zahid Usmani:-**

Frontend with user interfaces and UX design.

**Arshiya Dilshad:-**

Testing and Automation