**PROJECT REQUIREMENT AND SPECIFICATION**

**ON**

**CHATBOT USING PYTHON**

**CSE  VI-TH SEMESTER MINI PROJECT**

**2021-2022**



**Submitted to:       Submitted by:**

DR. JYOTI AGARWAL      MR KUSHAGRA TRIVEDI

DR NARAYAN CHATURVEDI University Roll. No.: 2015079

CSE-DS- VI-th Sem

GUIDED BY:

MR. YUVRAJ JOSHI

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**GRAPHIC ERA DEEMED TO BE UNVERSITY, DEHRADUN.**

* 1. **About Project**

Asking a question to a machine and receiving an answer was always the stuff of sci-fi in the not too distant past. Now, things have changed, and we find ourselves using Chatbot systems everywhere without even realizing it. Google,Siri,Alexa are the best example of modern day chatbots used in our day to day activities. At the most basic level, 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.

**1.2 METHODOLOGY:**

* **NLP CONCEPTS BEING USED**

**NLP:**

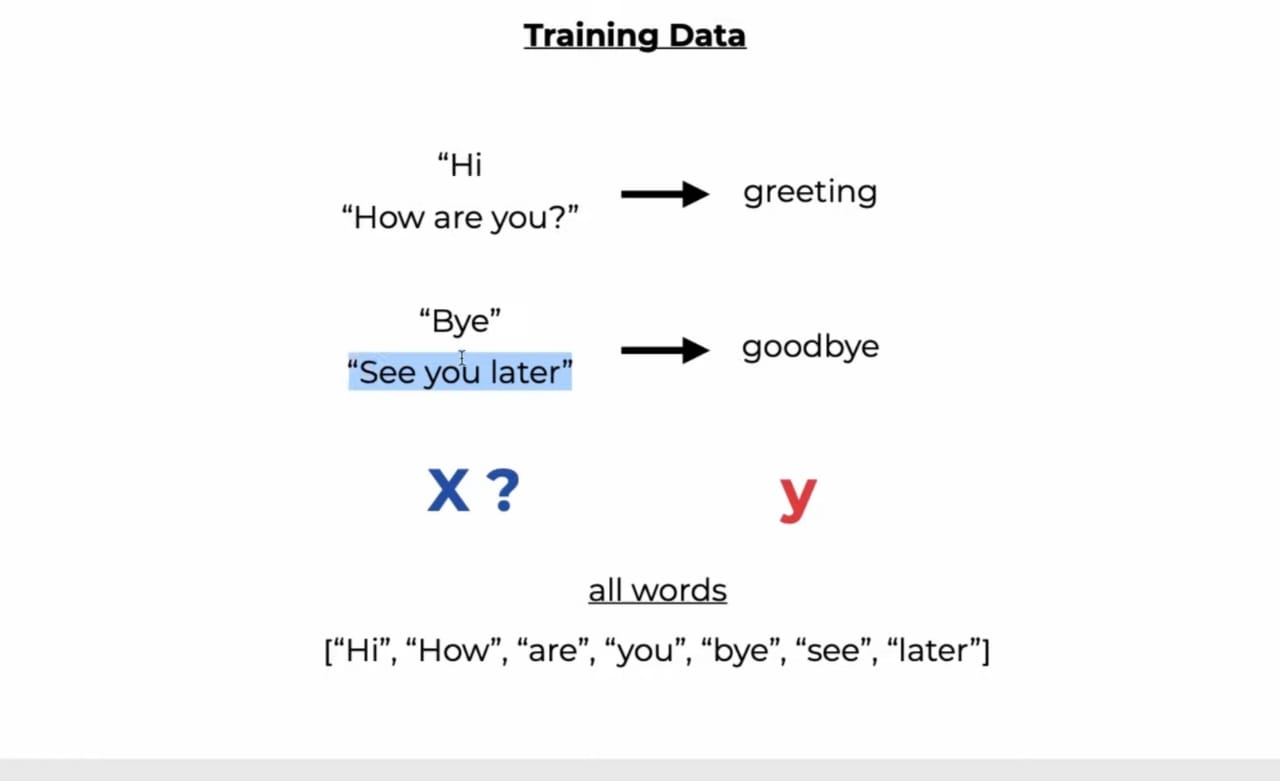
Natural language processing (NLP) is the ability of a computer program to understand human language as it is spoken and written referred to as natural language. It is a component of artificial intelligence.Nlpenables computers to understand natural language as humans do. Whether the language is spoken or written, natural language processing uses artificial intelligence to take real-world input, process it, and make sense of it in a way a computer can understand. Just as humans have different sensors -- such as ears to hear and eyes to see -- computers have programs to read and microphones to collect audio. And just as humans have a brain to process that input, computers have a program to process their respective inputs. At some point in processing, the input is converted to code that the computer can understand.

### 

### **Why is natural language processing important?**

Businesses use massive quantities of unstructured, text-heavy data and need a way to efficiently process it. A lot of the information created online and stored in databases is natural human language, and until recently, businesses could not effectively analyze this data. This is where natural language processing is useful.There are two main phases to natural language processing: **Data Preprocessing and algorithm development.**

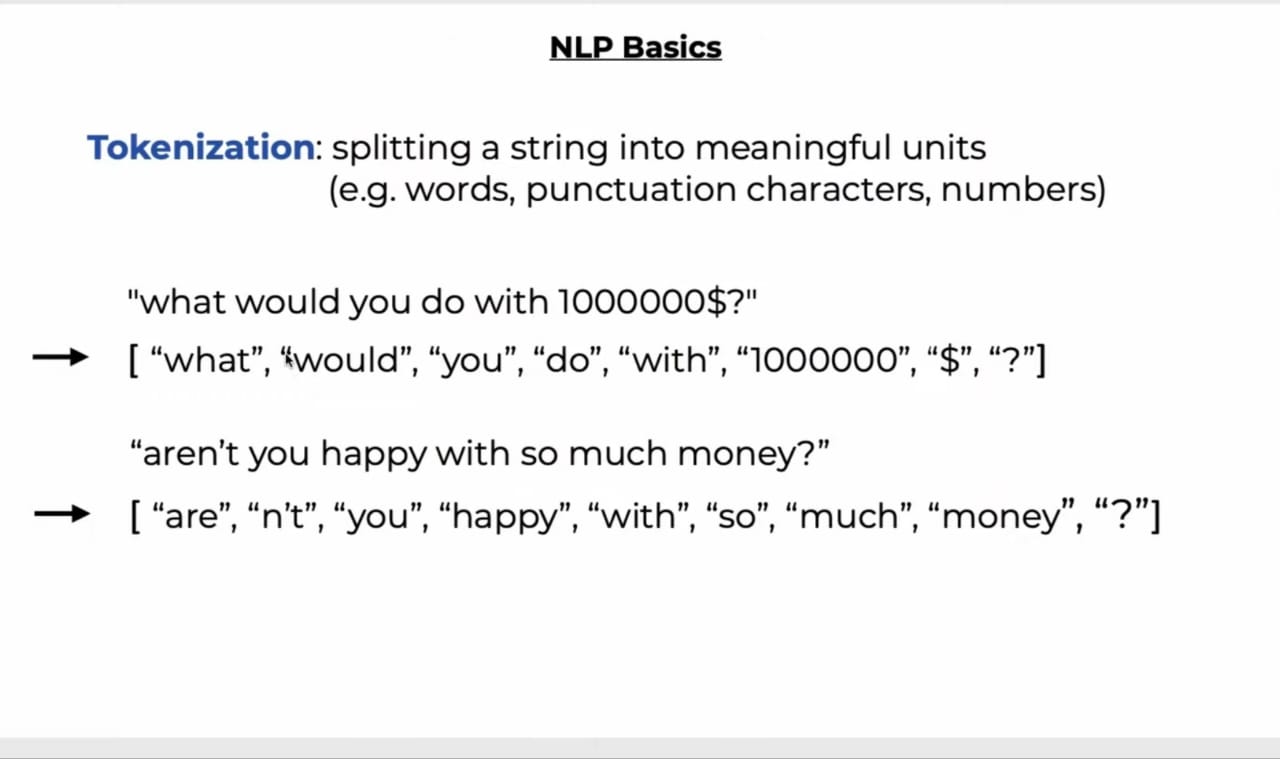
* **Data preprocessing and Creating Training Data:**

****

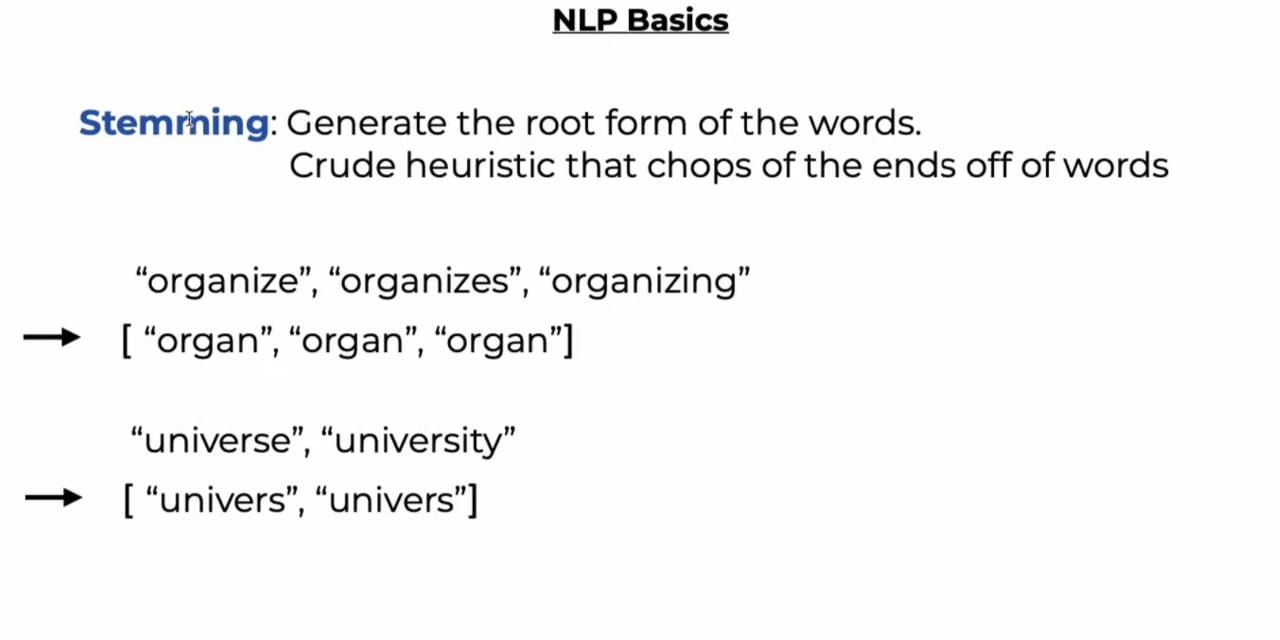
Involves preparing and "cleaning" text data for machines to be able to analyze it. preprocessing puts data in workable form and highlights features in the text that an algorithm can work with.We have **Intents.json** file as our Initial data to train our model with which we pass through the nlp preprocessing pipeline,There are several ways this can be done, including:

* **Tokenization:**

This is when text is broken down into smaller units to work with.

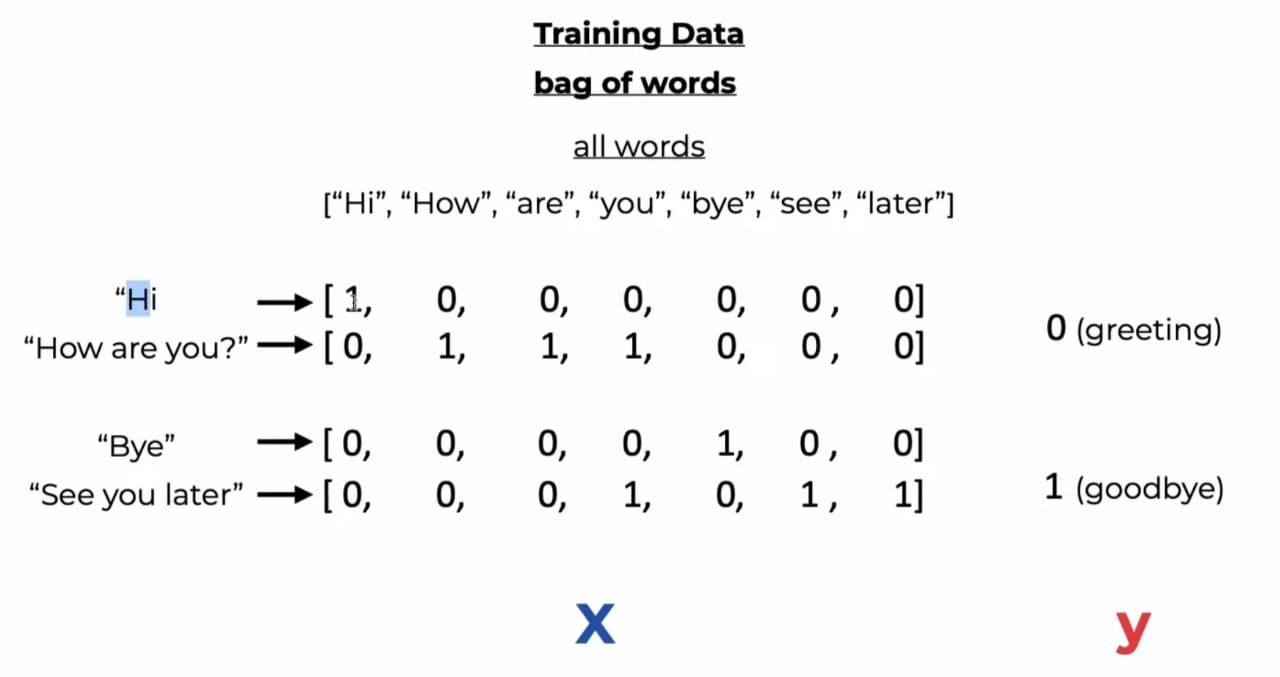


* **Stop word removal.** This is when common words are removed from text so unique words that offer the most information about the text remain.
* **Lemmatization and stemming.** This is when words are reduced to their root forms to process.



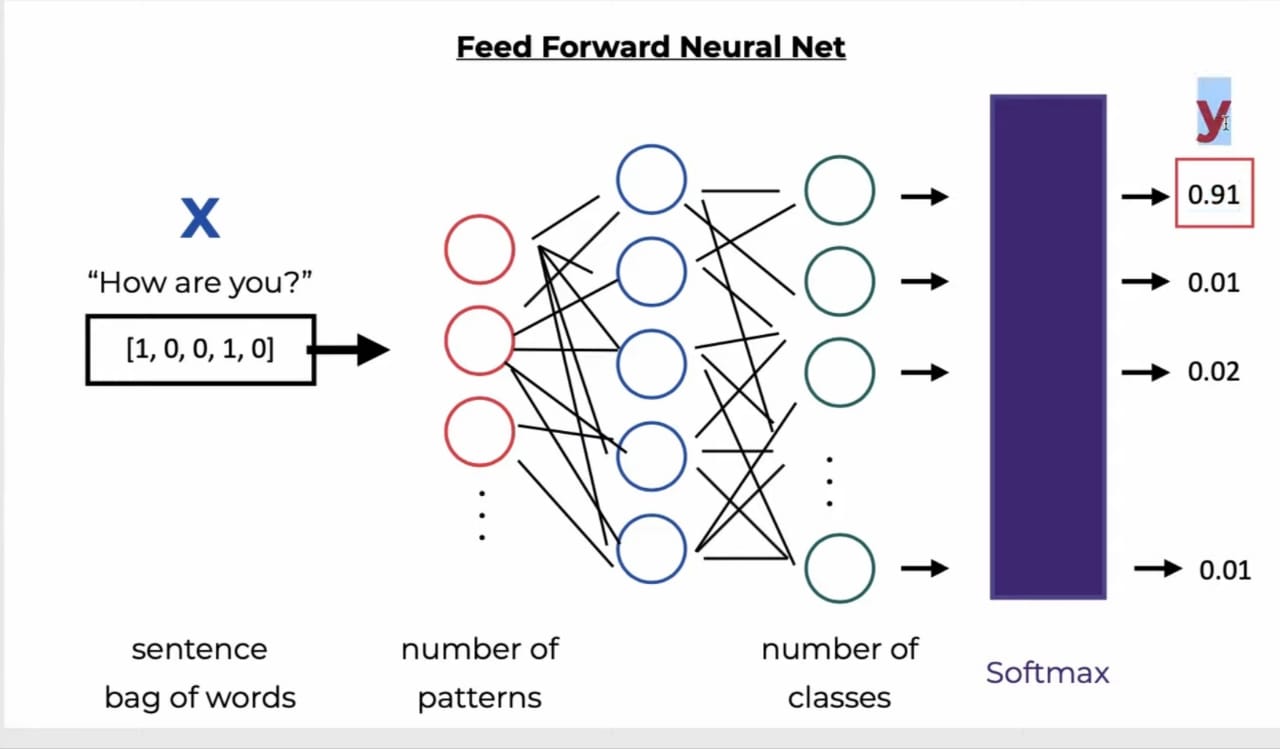
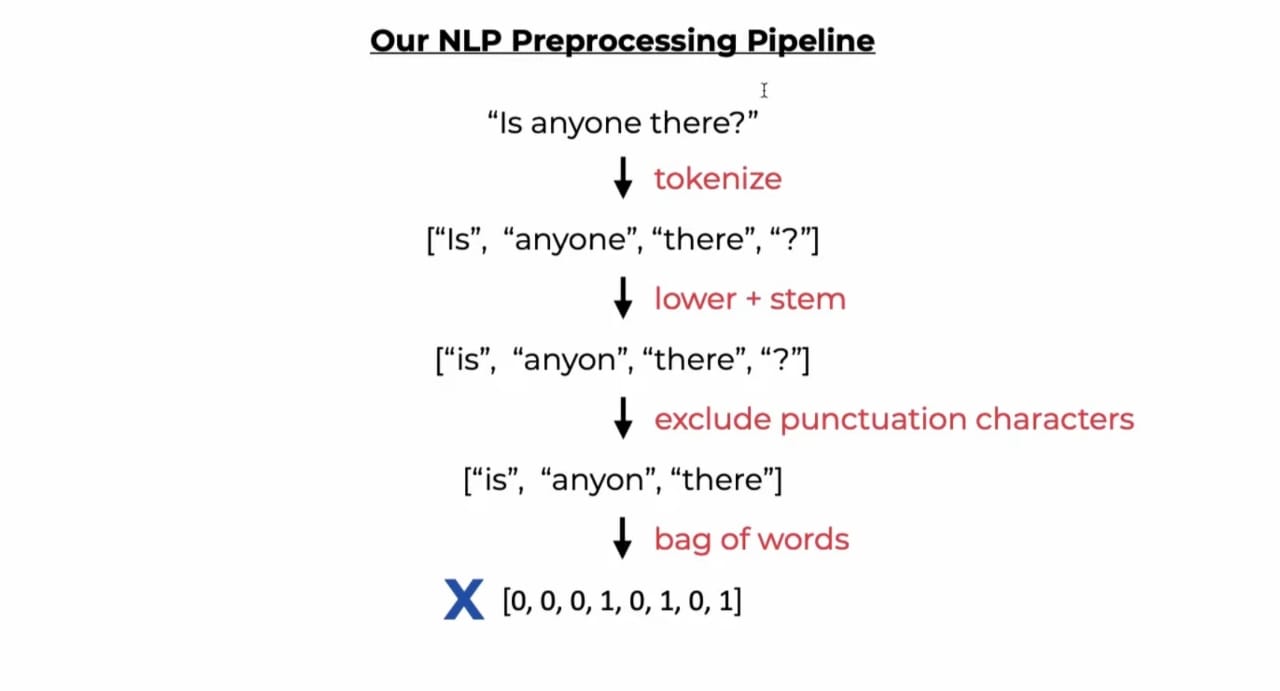
* **Bag of Words:**

A Natural Language Processing technique of text modeling known as Bag of Words model. Whenever we apply any algorithm in NLP, it works on numbers. We cannot directly feed our text into that algorithm. Hence, Bag of Words model is used to preprocess the text by converting it into a *bag of words*, which keeps a count of the total occurrences of most frequently used words.



* **Algorithm Development:**

Once the data has been preprocessed, an algorithm is developed to process it. There are many different natural language processing algorithms, but two main types are commonly used:

* **Rules-based system.** This system uses carefully designed linguistic rules. This approach was used early on in the development of natural language processing, and is still used.
* **Machine learning-based system.** Machine learning algorithms use statistical methods. They learn to perform tasks based on training data they are fed, and adjust their methods as more data is processed. Using a combination of machine learning, deep learning and Neural Networks, natural language processing algorithms hone their own rules through repeated processing and learning.
* Here we are going to use **Feed Forward Neural Net** to process our Tokens.A Feed Forward Neural Network is an artificial neural network in which the connections between nodes does not form a cycle. The opposite of a feed forward neural network is a Recurrent Neural Network, in which certain pathways are cycled. The feed forward model is the simplest form of neural network as information is only processed in one direction. While the data may pass through multiple hidden nodes, it always moves in one direction and never backwards. 
* **Complete NLP Preprocessing Pipeline: **
* **Feeding Trained Data to Models:**

And with the help of this nlp preprocessing pipeline we create our training data which is feeded in the models to train them so they can make predictions.Here we have **train.py** file doing the job for us.

* **NLTK Library and Methods:**

The Natural Language Toolkit (NLTK) is a platform used for building Python programs that work with human language data for applying in statistical natural language processing (NLP). It contains text processing libraries for tokenization, parsing, classification, stemming, tagging and semantic reasoning.In our project we create nltk\_utils.py file to implement methods like tokenize,stem and bagofwords.

* **Pytorch Model and Training with Model.py:**

Here in model.py ,we are implementing our pytorch model and giving functionality to out neural network model and building training loops and epochs to be trained on.

* **Save/Load Model and Implement the chat with Chat.py:**

Here we are finally compiling all the files to create the final chatbot output so we are going to import all the files and we are going to give our chatbot the name and the behavior text it is going to output based on the questions asked.

**There we have it! A working Chatbot system based on intents and self learning through intents.**

* **Requirements of Project:**
  1. **Hardware Requirement:**
* Processor: min 1 GHz ,recommended 1.60 GHz or more
* Ethernet connection or a wi-fi
* Hard Drive: min 32GB ,recommended 64 GB or more
* Memory(RAM):8 GB recommended
* OS:WINDOWS 10
* System type: 64 bit recommended.
* GPU:16GBrecommended

1. **SoftwareRequirements:**

* Python 3s version
* Visual Studio
* Pytorch
* Nltk Library

**3. Modules of Project:**

* Intents Creation
* Data Preprocessing
* Feeding & Training Feed Forward Neural Network Model
* Pipeline and Prediction

**4.BRIEF MOTIVATION:**

I want to thank Computer Science Engineering Department who provided me this

opportunity to do the mini project .I also want to thank my teachers who provided

the lecture to understand all important topics related to project and taught us to

write and run the program, Completing this project has  build a confidence in me

and now i am able to perform more projects like these.

**5.  REFERENCES:**

**1.**Sites like Deepai.org,AnalyticsVidhya and techtarget for understanding Concepts.

**2.**Hugging Face Documentation

**3.**Google.com for some explanations

**4.**Towards DataScience Documentation

**THANK YOU!**