**AI CHATBOT USING NLP**

**Contents**

1. Project Objective………………………………………………………..3
2. Data Understanding……………………………………………………..5
3. Data Preparation…….…………………………………………………..7
4. Model Building...………………………………….……………………10
5. Deployment ……….……………………………………………………12

1. Conclusion ……………………………………………………………..17
2. Future Scope……………………………………………………………18

**PROJECT OBJECTIVE**

* To provide superior user experience and support 24/7 efficiently.
* To maximize real-time engagement with users.
* To minimize cost and time needed for maintaining a support team.
* Chabot’s can help businesses improve efficiency and accuracy all while reducing costs.
* Provides better customer service and better customer experience.
* Enhanced Chabot service improves customer satisfaction.

**Constraints:**

* Chabot’s Don't Understand Human Context.
* They Have Zero Research Skills.
* Chabot’s Have No Emotions.
* Domain specific
* Needs maintenance
* Chabot’s are often repetitive

**Project overview:**

* Minimal human interference in the use of devices is the goal of our world of technology.

* Chabot’s can reach out to a broad audience on messaging apps and be more effective than humans are.

* At the same time, they may develop into a capable information-gathering tool. They provide significant savings in the operation of customer service departments.
* With further development of AI and machine learning, somebody may not be capable of understanding whether he talks to a Chabot or a real-life agent.

* In this project we extracted text data from Wikipedia and other web portals to create corpus in the form of JSON file format and created a Chabot for users to answer their queries.

**Project Scope:**

* Technology today is evolving at breakneck speeds, offering businesses multiple opportunities to market their brands and enhance the customer experience.
* A Chabot is one of the most prominent technologies among these advancements.

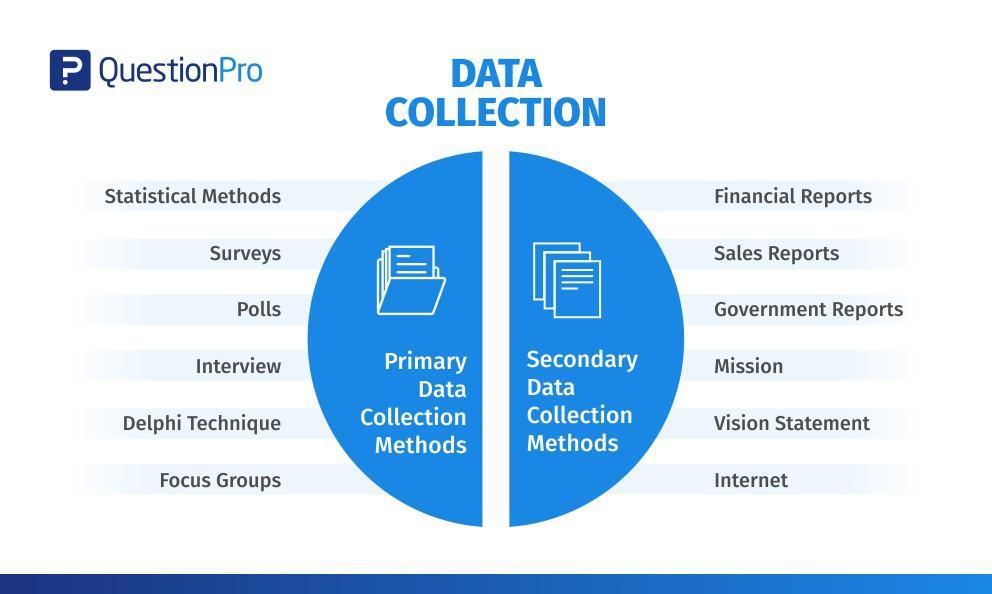
* Chabot’s are industry-agnostic and can be implemented across different verticals. Chabot’s not only help you save costs but, at the same time, ensure a superior customer experience that helps set your business apart.
* Main objective of this project is to give control to a user to get their query answered at any point of time.

**Data Understanding:**

|  |  |
| --- | --- |
| Source of the Data | 1. Collected data from use cases.  2. Gathered questions and answers to train the Chabot. 3. Identified most frequently asked questions and queries. 4. Did domain research based on Data Science. 5. Collected general scope questions, greetings, etc. |
| Type | Text data |
| About the Data | * The input is a JSON file format consisting of intents with unique tags. * It contains different variations of questions in patterns and respective answers in responses. * The File contains some of the questions and answers related to the Restaurants Food Service domain. |

**Data collection:**

Data collection is defined as the procedure of collecting, measuring and analysing accurate insights for research using standard validated techniques

* Restaurant-related queries were gathered to generate a corpus.
* Collected data on restaurant and customers possible interaction topics.
* Prepared possible questions and answers for each topic.
* Created a JSON file containing the full corpus.
* 

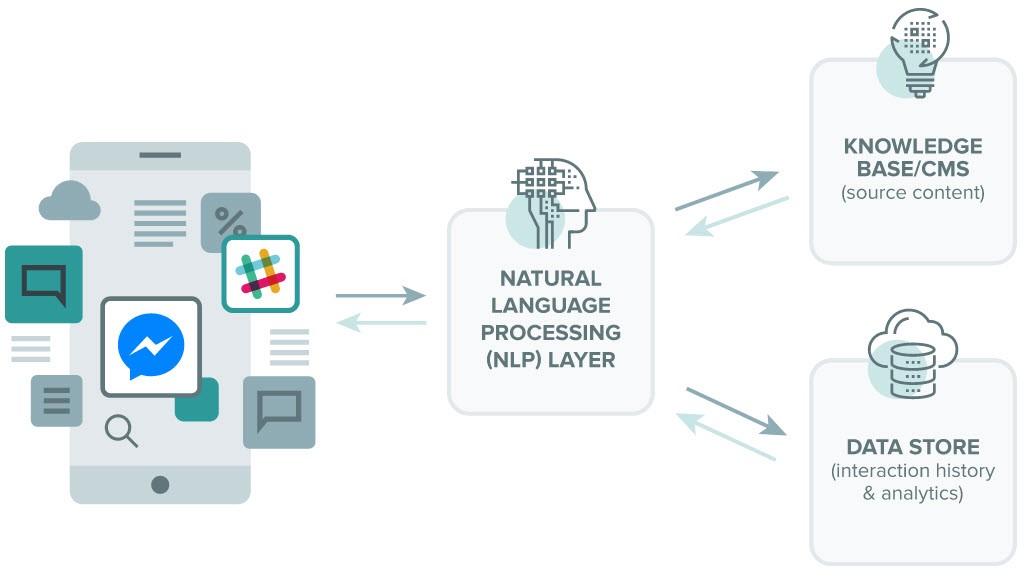
**Data Preparation:**

**Exploratory Data Analysis and Data Pre-processing**

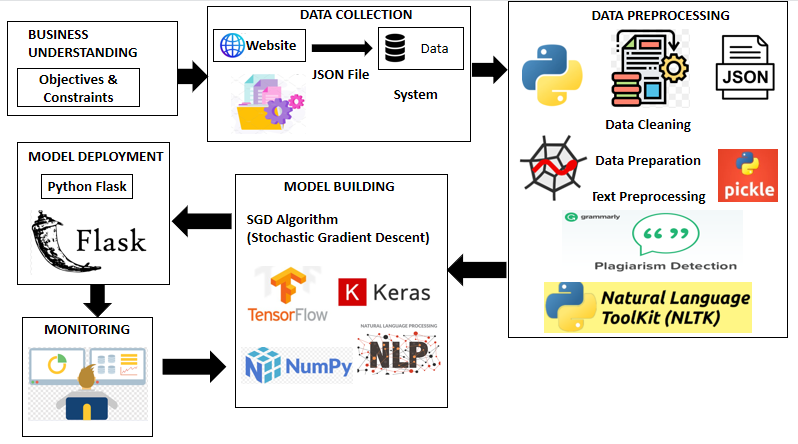
Pre-processing data

* The dataset originally comes in JSON format. We should convert our data into Tensor flow proprietary Example format.
* Removed the special characters like #, @, &, - from the data collected.
* The main benefit of this format is that it allows to load tensors directly from the input files and let Tensor flow handle the shuffling, batching and queuing of inputs. As part of the pre-processing, also create a vocabulary.
* NLTK was used to perform text pre-processing
* Lemmatized each of the words and saved in pickle format to use it further.

* Created training corpus using Bag of Words technique and then divided data into training and test data to train our model
* Use a lemmatizer to pre-process the words and remove duplicates.
* Tokenization.
* Creating an error-free huge corpus for the purpose of training the Chabot.
* Create a bag of words for the data.



**Project Architecture:**



**Model Building:**

**What is NLP?**

* Process information contained in natural language text
* Started off as a branch of Artificial intelligence
* Also known as Human Language Technology, Natural Language Engineering

**NLP Examples**

* Smart assistances like Apple’s Siri, Google voice assistant and Amazon’s Alexa
* Language translations like Google Translate
* Text prediction and text generation
* Chabot

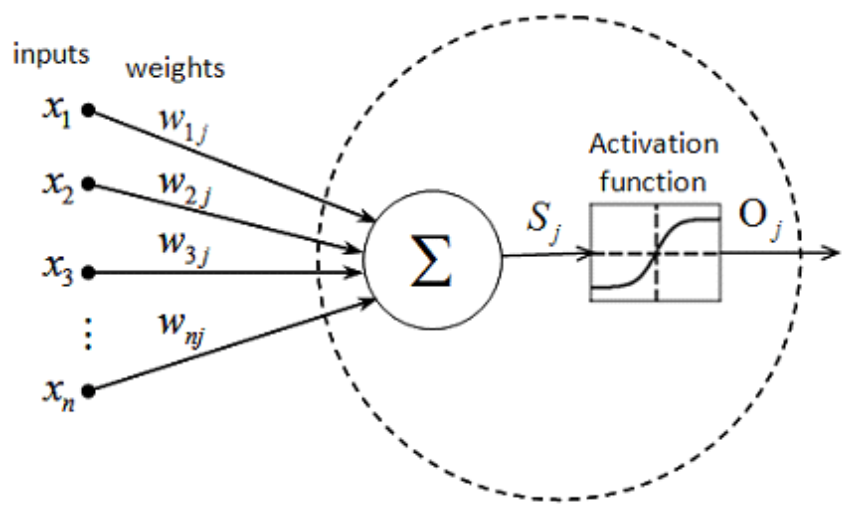
**Model Description:**

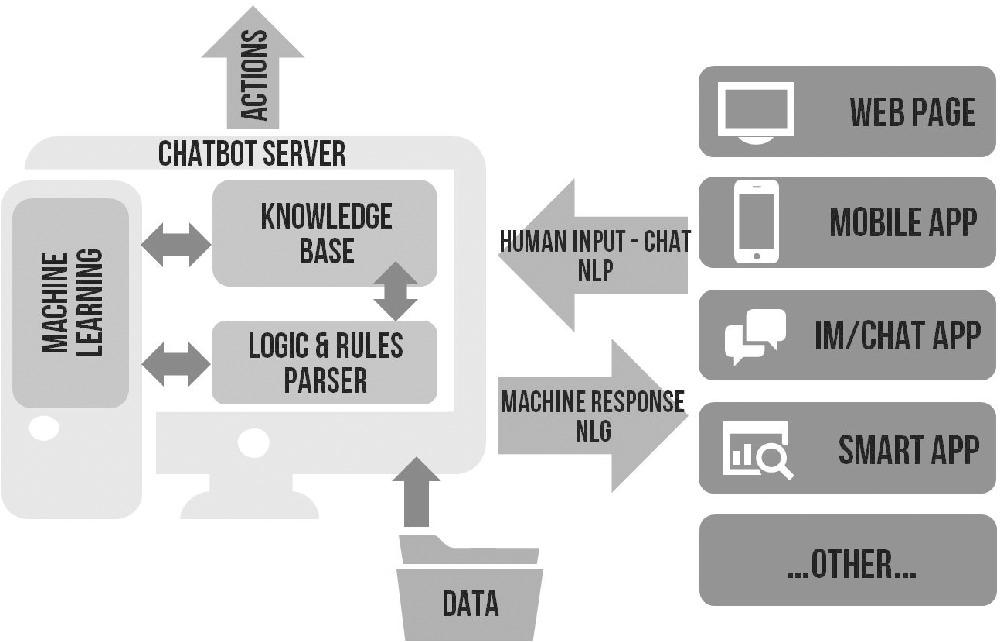
**Deep Neural Networks**

* The model is a Tensor flow based sequential model.
* Several dense layers are stacked together and trained.
* Each of the hidden layer contains different number of neurons it (1024, 512, 256, 128, 64)
* Pre-processed text is divided into training and test sets.
* The first layer takes the input text and weights get initialized, an activation function applied on to it.
* The output of each neuron from each layer transfers its output to the next layer which is densely connected.
* Forward and back propagation is done until training is finished.
* Model compilation.
* Fitting the model to the training and test data.

**Parameters**

* Activation function ‘relu’ used in hidden layers and ‘softmax’ in output layer
* Batch size – 32, Optimizer – SGD, Loss –  ‘categorical\_crossentropy’, Epochs – 200



****

**Model Evaluation:**

* After training the model we can evaluate it on the training data.
* This will run evaluation metrics on the test set instead of the validation set.
* We will get probability scores for unseen data.
* We could imagine feeding in 100 potential responses to context and then picking the one with the highest score.
* After hyper parameter tuning model accuracy is at 95%

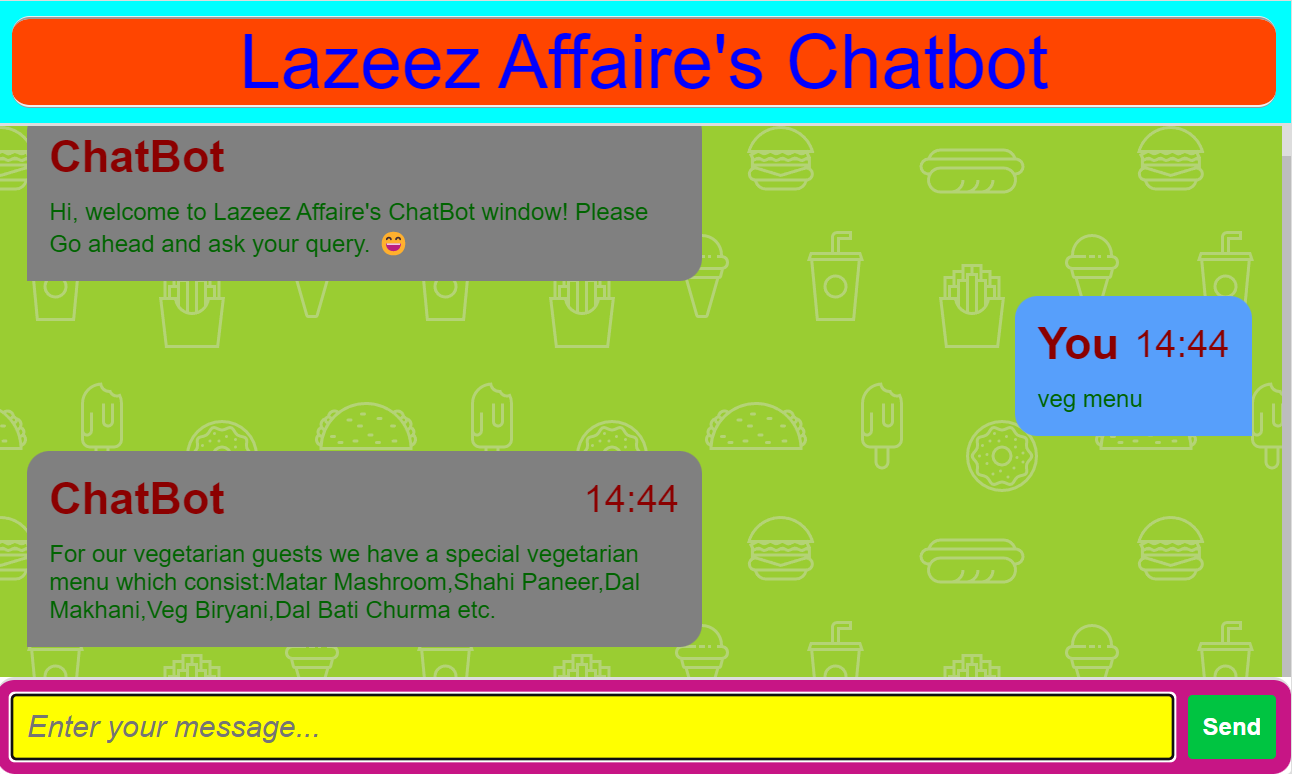
**Deployment:**

* Flask framework was used to create a web based application.
* Flask is a micro web framework written in python. It is classified as a micro framework because it does not require any particular tool or libraries.
* Other available frameworks are Streamlit, Django, etc.
* The Chabot uses HTML for web pages and CSS for styling.
* The web-based Chabot will take user question and provides an answer which is having the highest probability

**Web based:**

* Deployment using Flask Web Framework

**1. Web App** → here we can check the performance of Chatbot. Enter the required query and then hit the enter button to get the answer as output.

****

**Conclusion and Future Scope:**

* In this Assignment project, we understood about Chabot’s and implemented a deep learning version of a chatbot in Python which is accurate.
* We can customize the data according to business requirements and train the chatbot with great accuracy.
* Chatbots are used everywhere and all businesses are looking forward to implementing bot in their workflow.
* More efficient chatbot

* Voice chatting