**ChatBot Recommendation System**

**ChatBot:**

**A chatbot is an artificial intelligence (AI) software that can simulate a conversation (or a chat) with a user in natural language through messaging applications, websites, mobile apps or through the telephone.**

**Why are chatbots important? A chatbot is often described as one of the most advanced and promising expressions of interaction between humans and machines. However, from a technological point of view, a chatbot only represents the natural evolution of a Question Answering system leveraging Natural Language Processing (NLP). Formulating responses to questions in natural language is one of the most typical examples of natural language processing applied in various enterprises’ end-use applications.**

**MODEL:**

**First we downloaded the NLTK(natural language toolkit) . Then we studied the NLP , in which we studied Text Preprocessing which includes:**

**1. Tokenization**

**2. Stemming**

**3. Lemmatization**

**4. Stop words**

**5. Tf Idf**

**Tokenization:**

**Given a character sequence and a defined document unit, tokenization is the task of chopping it up into pieces, called tokens , perhaps at the same time throwing away certain**

**characters, such as punctuation. Here is an example of tokenization:**

**Input: Friends, Romans, Countrymen, lend me your ears;**

**Output: [‘Friends’, ‘Romans’, ‘Countrymen’, ‘lend’, ‘me’, ‘your’, ‘ears’]**

**Stemming:**

**Stemming is the process of reducing a word to its word stem that affixes to suffixes and prefixes or to the roots of words known as a lemma. Stemming is important in natural**

**language understanding (NLU) and natural language processing (NLP). Stemming is also a part of queries and Internet search engines.**

**Lemmatization:**

**Lemmatization usually refers to doing things properly with the use of a vocabulary and morphological analysis of words, normally aiming to remove inflectional endings only and to**

**return the base or dictionary form of a word, which is known as the lemma .**

**Stop Words:**

**In natural language processing, useless words (data), are referred to as stop words. Stop Words: A stop word is a commonly used word (such as “the”, “a”, “an”, “in”) that a search**

**engine has been programmed to ignore, both when indexing entries for searching and when retrieving them as the result of a search query.**

**TF IDF:**

**The Inverse Document Frequency is the the number of times a word occurs in a corpus of documents. tf-idf is used to weight words according to how important they are. tf-idf is**

**used in a number of NLP techniques such as text mining, search queries and summarization.**

**After Text preprocessing for understanding classification techniques we worked on Sentiment Analysis , which predicted whether the sentence is positive , negative or neutral, also for but we didn’t used it in our model .**

**Sentiment Analysis:**

**the process of computationally identifying and categorising opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular**

**topic, product, etc. is positive, negative, or neutral.**

**With the recent advances in deep learning, the ability of algorithms to analyse text has improved considerably. Creative use of advanced artificial intelligence techniques can be an**

**effective tool for doing in-depth research. We believe it is important to classify incoming customer conversation about a brand based on following lines:**

* **Key aspects of a brand’s product and service that customers care about.**
* **Users’ underlying intentions and reactions concerning those aspects.**

**Then we studied Logic Adapters and Storage adapters of Chatterbot.**

**Logic Adapters:**

**ChatterBot Logic Adapters. Logic adapters determine the logic for how ChatterBot selects a response to a given input statement. The logic adapter that your bot uses can be**

**specified by setting the logic\_adapters parameter to the import path of the logic adapter you want to use.**

**Storage Adapter:**

**Storage adapters provide an interface that allows ChatterBot to connect to different storage technologies. The storage adapter that your bot uses can be specified by setting the**

**storage\_adapter parameter to the import path of the storage adapter you want to use.**

**Then we also studied StatsModel but didn’t apply in our working model .**

**StatsModel:**

**Used for estimation of many different statistical models, as well as for conducting statistical tests and statistical data exploration. An extensive list of result statistics are**

**available for each estimator . Results are tested against existing statistical packages to ensure that they are correct .**

**Then we made Feedback system in which we were training the model and creating the database by our own questions and answers and then if the feedback was given as yes then the particular set of question answer was saved in the database and if no was given as the feedback then the set of question and answer was not saved in the database . Earlier we were working on SQL Storage Adapter , then we installed MONGODB and started our training on MongoDatabase Storage Adapter. We also studied how to change The Uri of the storage adapter also in learning phase we created our own database using module PyMongo. Then we worked on same feedback system with the MongoDatabase Storage Adapter and made a separate file which could access the same feedback system uri.**

**After this for recommendation we studied K Means, Decision Tree, MultinomialNB Classifier**

**K Means:**

**K refers to number of clusters(groups) we expect to find in a dataset , Means refers to average distance of data to each cluster centre also known as centroid , which we are trying**

**to minimise . It works on the algorithm :**

**1. Place K random Centroids for initial clusters .**

**2. Assign data samples to nearest centroid .**

**3. Update centroids based on newly assigned samples .**

**In this technique we also did Elbow Test which actually in initial phase checks that how many clusters could be made .**

**We dropped down this technique because for small dataset Elbow Test was working fine but for large datasets it was actually difficult to predict number of clusters formed .**

**Decision Tree:**

**A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and**

**utility. It is one way to display an algorithm that only contains conditional control statements.**

**This technique uses Test size which determines amount of data to be trained and rest used as a data .**

**In this we used Gini Index which calculates the index values and the node having lowest index value is chosen as root node , and we calculated F1 Score , F1 score combines**

**precision and recall relative to a specific positive class -The F1 score can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value**

**At 1 and worst at 0.**

**MultinomialNB Classifier:**

**This classifier is commonly used when we have discrete data (e.g. movie ratings ranging 1 and 5). Multinomial Naive Bayes is a specialised version of Naive Bayes that is designed**

**more for text documents. It estimates the conditional probability of a particular word given a class as the relative frequency of term t in documents belonging to class(c). The**

**variation takes into account the number of occurrences of term t in training documents from class (c),including multiple occurrences.**

**We used MultinomialNB Classifier for our recommendation system . Afterwards moving towards our main Model we changed the parameters of our mongo database such as from in\_resdponse\_to changed to question and text changed to answer . After that we started working with the actual dataset according to which the response should be given regarding any query . We actually did this because in our actual dataset the parameters were different but we dropped down this approach because of the reason that we needed not just two parameters but we required more parameters**

**For our main model we first created a separate python file which actually extracted the fields such as Problem Description , Solution , Technology stack etc. and created a YAML file where the statements of those selected fields were saved . Then in our main python file we created a new uri for mongodatabase and trained it using that YAML file that we created. Atlast our Python file was created and worked accordingly.**