Working of the project

TRAINING DATA :

We have created a intent.json on which we have loaded our Training Data with a different classes.

It consist of A keyword and response to provide with a appropriate results

Data Pipeline :

PyTorch is created for training the model and function of the automation

Our Model is a **sophisticated,** this has feed forward neural network and 2 hidden layers

We use PyTorch to create a data pipeline to store the pattern and the formula about the machine learning about how to react for given text. This file has to be trained again after creating changes in the Data.

We also added softmax for probability function

SCOPE

We are gonna further implement with tensor flow for creating a bot with larger datasets and providing with most probable (0.75) accurate results

Tokenisation :

So we split a single sentence into a string of array

With a training data we collect a particular set bag of words which further is matched with the string in array to create a pattern

Stemming :

Stemming is a NLP technology in which the process of reducing the words to its simplest term so the tenses won’t affect the recognition systems

Concept of Bag of words :

The bag of words is another set of array consist of words which are stored after Stemming

This pattern creates response which is recognised with the data pipeline

So The NLP preprocessing pipeline works in a order which is depending on flow.

Tokenise → stemming → excluding punctuation characters → matching with bag of words

Technology used :

- Natural Language toolkit → NLTK
  - PorterStemmer
- JSON for training Data
- NumPy
- PyTorch
  - Neural Network
  - utils data for data loader
-