**Project-2**

**NLP Report**

Swaraj Chirumamilla

Arun Soma

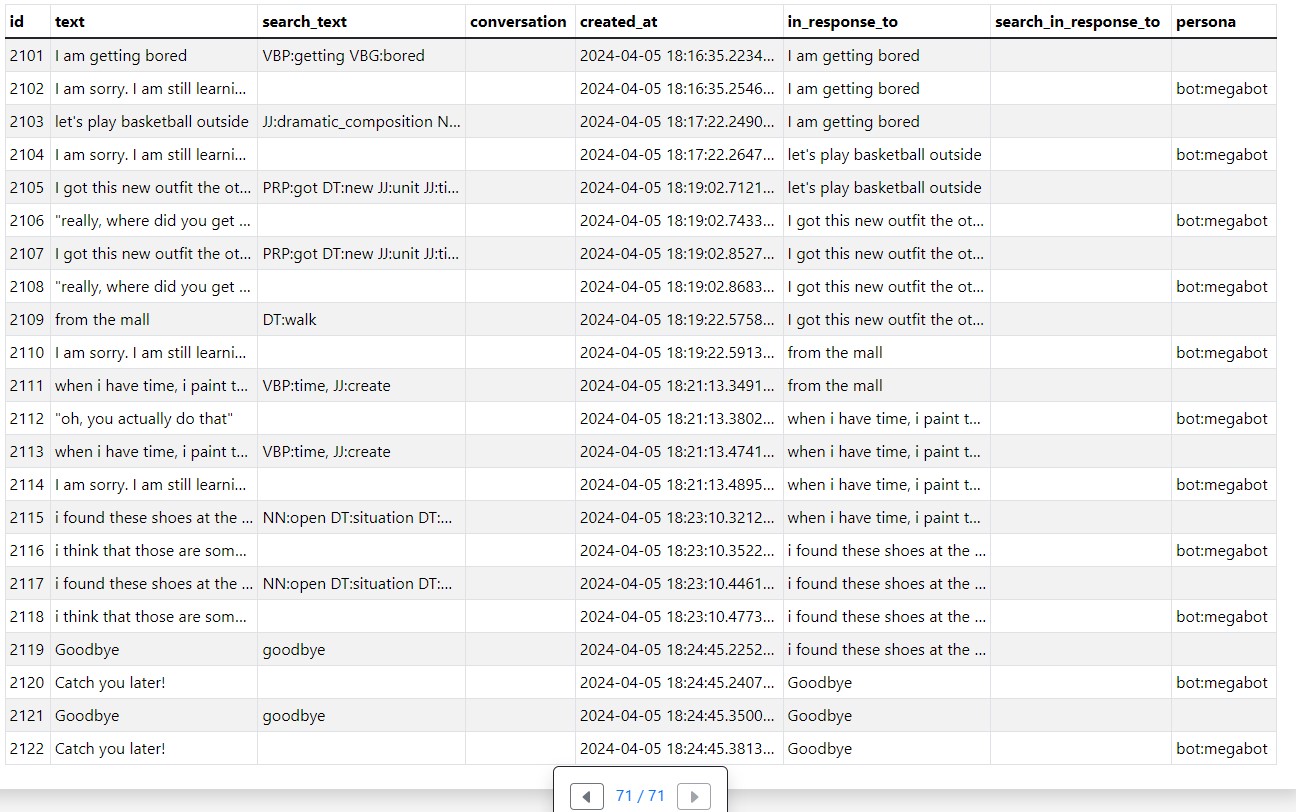
For our training data, we used a dialogs text file from Kaggle and pre-processed it by splitting the text in raw file by using regular expression pattern like periods and question marks effectively into a list of sentences, stripping leading and trailing whitespace in each sentence using the strip() method and removing duplicate sentences from the dataframe.

The custom features that we implemented are using storage adapters, logic adapters by using 2 unique bots where we set their confidence levels to 0.50 (megabot) and 0.80 (gigabot), where their responses would alternate between the two based on the prompts given by the user.



In the above conversation with the chatbot, the first four prompts given by us the chatbot responded with the default training data in the ListTrainer alternating between megabot and gigabot. For the responses with double quotes, the chatbot searches through the dataframe to find the exact response for the given prompt. When for the given chatbot megabot and gigabot, when the maximum confidence level falls below 0.50 and 0.80, the chatbot responds with the default response as 'I am sorry. I am still learning!' and 'I do not have an answer to that.'

For the prompts which are very similar to the actual text like, “i found these new shoes earlier at the store” we can see storage adapters being in use over there for the response.



In the above screenshot from the sqlite viewer, the 5 POS tags given here with their meanings are

1. VBP: Verb, present tense, other than third person singular
2. JJ: Adjective
3. PRP: Personal pronoun (PP)
4. DT: Determiner
5. NN: Noun, singular