

Project Summary: Chatbot

1 Problem Statement

To develop a chatbot based on text sequence which interacts with users and respond to their queries

2 Dataset

We have used publicly available **Reddit comments** to make parent reply table. If a comment has more than one reply, we have selected most upvoted one.

- **Comment size:** 5GB
- **Period:** 2 months
- **URL for zipped comment:** [here](#)

3 ML Algorithm used

LSTM: Since the chatbot needs context for better conversation, So we have used RNN (LSTM) as it allows previous information to persist which was not possible in neural networks.

LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behavior.

- **Activation Function:** Sigmoid
- **Weight Initialization Function:** Glorot Normal
- **Number of nodes:** 300
- **Number of epochs:** 5000
- **Number of layers used:** 4

NLP: We have used NLP tools from NLTK library for language processing to better understand human language.

4 Platform used

NIMBIX Jarvice, IBM

- **RAM:** 512 GB
- **No of CPU cores:** 40
- **GPU:** 8x NVIDIA V100 NVLink

5 Time to train

Each batch (of 300) takes approximately 5 days for 500 epochs.

6 Results

The model produced after training on 5GB of comment data for 500 epochs is produced. We are using app interface to present the chatbot.