

To,

IITD-AIA Foundation of Smart Manufacturing

Subject: **Weekly Progress Report for Week 3**

Dear sir,

Following is the required progress report to the best of my knowledge considering relevant topics to be covered.

What's happening this week:

- Studied Corpus and Made it.
- Made dataset.
- Performed EDA.
- Tried to implement a normal chatbot using twilio and flask.
- Learnt How to process large text data in NLP.

Weekly Progress:

June 19:

Studied various topics like corpus and it's preprocessing.

- Studied more detailed about the corpus.
- Made corpus from the fsm file and stored it into the text file.
- Also done its cleaning by removing punctuations and numbers from converting them into tokens and doing stemming (Porter stemming).
- I have done this on div tags of the website and removed every extra character or space.

June 20:

Understand different concepts of doing exploratory data analysis (EDA).

- Studied exploratory data analysis (EDA) from which we can get statistics of our data and relationship between them. It helps us to retrieve the main feature from the data.
- tried to visualize the most common words from the 1st corpus set but results are not coming appropriate.
- Trying to do eda on text file of the corpus. but stucked with there and trying to figure it out.

June 21:

Learnt about EDA and corpus. Also faced some issues related to corpus files.

- Studied exploratory data analysis (EDA) and how to do it on text file.
- try to convert my corpus text file into csv file.
- corpus contains much duplicate data so i tried to remove that duplicate data but unable to do that.

June 22:

Tried to do EDA with Feature engineering but faced some issue which was resolved later.

- Done EDA with dataset (Half done).
- Learn Feature engineering but didn't get what feature should be created for our dataset.
- divided the project into sections as per guided.

June 23:

Tried to explore the attributes in the dataset as total words of the content. Also tried Twilio with OpenAI and flask.

- Tried to visualize the content of the dataset and find the patterns or relationship between them.
- But only able to visualize the total number of words with the content.
- tried to build a normal bot using Twilio and OpenAI but finding difficulty in making virtual environment.

June 24:

Things became clear after the meeting so Did EDA and tried to visualize the contents.

- Tried to make a normal whatsapp bot with twilio and flask but it didnt workout because of some openai api issue.
- Almost done with EDA.
- Visualized it into word cloud,done bi-gram and uni-gram on it.
- Drawn a graph of word frequency. Also tried to do sentiment analysis but it came to be neutral.

June 25:

Done with EDA and learnt many things about NLP.

- Completed Exploratory Data Analysis.
- you can view it on [GitHub](#).
- Learnt how to do EDA on the text data which is used in NLP.

GANTT CHART FOR INTP23-ML-01

