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 OpenAl 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 Task Task Name Start End Durato | | | | | | | | | | | | | | | | G | ΔN | TT | CH | ΔR | T F | OR | INT | P2: | R-M | I -Oʻ | 1 | | | | | | | | | | | | | | | | | | | | — | |
|------------|--|-----------------------|------------------------|--------------|----------|----------|----------|--------|-----------|--------|--------|--------------|--------|----------|----------|----------|----------|--------|----|----------|--------|-----------|-----------|--------|-----------|--------|--------------|--------|--------------|-----------|--------------|-----------|-----------|-----------|--------|--------|--------|----------|----------|--------------|--------|---------|---------|---------|----------|----------|-----------|-----------|---------|
| Task ID | Task Name | Start date | End date | Duratio n | 01-Jun | 02-Jun | 03-Jun | 04-Jun | 05-Jun | unf-90 | 08-lun | unf-60 | 10-Jun | 11-Jun | 12-Jun | 13-Jun | | 15-Jun | _ | 17-Jun | 18-Jun | | | | | 24-Jun | | 26-Jun | 27-Jun | 28-Jun | 29-Jun | 30-Jun | 01-Jul | 02-Jul | 04-Jul | Inf-S0 | Inf-90 | 07-Jul | lnf-80 | lnf-60 | 10-Jul | 13-14 | 12.70 | 13-Jul | 15-Jul | 16-Jul | 17-Jul | 18-Jul | lul-91 |
| 1 | Deep Learning | 01- 06-23 | 01-06- 23 | 1 | 0 | 0 | 0 | 0 | 0 | 0 0 | 0 | 10 | - | 1 | н | - | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 2 | 7 | 2 | 2 | 2 | 2 | · | | , , | | - | 0 | 0 | | - | | - | 1 | | | н | - | | |
| 2 | Python Libraris | | 02-06- 23 | 1 | | | | | | | | T | | | T | | | | | | | | | | | | | | | | | | | | | | | | ı | 1 | | | | | | | \top | \top | |
| 3 | Linear modelling | 03- 06-23 | 03-06- 23 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 | Deep Neural Network | 06-23 | 04-06- 23 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 5 | Problems in network | 06-23 | 05-06- 23 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 6 | Prevention Techniques | 06-23 | 06-06- 23 | 1 | | | | | _ | | | 1 | | | | | | | | | | | | | | | ╙ | | Ш | _ | | | | | | | | | Ц | 4 | | | 1 | | | | _ | \perp | _ |
| 7 | Resoruces (Project) | 06-23 | 07-06- 23 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | 4 | | | | | | | | | 4 | | | | \perp | | | | 4 | 4 | _ |
| 8 | NLP | 06-23 | 08-06- 23 09-06- | 1 | | | 4 | | 4 | _ | | Ł | | \vdash | - | _ | | | | | | \perp | | 4 | 4 | | _ | | | \dashv | | 4 | _ | _ | - | | | _ | | 4 | | + | \perp | | _ | Н | 4 | \perp | _ |
| 9 | NLTK,Scipy libraries | 06-23 | 23 | 1 | H | | _ | | 4 | _ | - | H | | | | | | | | | | | | 4 | _ | + | ╀ | | \vdash | 4 | | + | - | | - | | | | \dashv | 4 | _ | + | + | \perp | | | \dashv | \dashv | _ |
| 10 | Models for DL | 06-23 | 23 11-06- | 1 | | | \dashv | | \dashv | | | + | | H | | | | | | | | \dashv | | + | | | \vdash | | | \dashv | | + | + | | | | | _ | + | + | | + | + | | | | \dashv | \dashv | - |
| 11 | Gradient Descent Deep Reinforcement | 06-23 12- | 23 12-06- | 1 | H | | | | - | + | | + | | | | | | | | | | - | | + | + | | \vdash | | H | \dashv | | - | | | | | | | \dashv | + | + | | + | | | | \dashv | + | + |
| 12 | learning Deep Learning | 06-23 13- | 23 13-06- | 1 | \vdash | | \dashv | - | \dashv | + | + | + | + | + | | | \vdash | | | Н | _ | \dashv | + | + | + | + | + | | \vdash | \dashv | + | + | + | + | + | | | \dashv | + | + | + | + | + | + | \vdash | Н | + | + | - |
| 13 | Limitations | 06-23 14- | 23 14-06- | 1 | | | 1 | | \dashv | | | + | | + | | | | | | | | \dashv | | + | | | + | | | + | | + | | | | | | | | + | | | + | | | Н | + | + | - |
| 14 | Python Libraries Data extraction | 06-23 15- 06-23 | 23 15-06- 23 | 1 | | | 7 | | \dashv | | | † | | \vdash | | | | | | | | | | $^{+}$ | | | \vdash | | | \forall | | \top | | | | | | | | † | | | $^{+}$ | | | | \forall | \forall | + |
| 16 | Text cleaning | 16- 06-23 | 16-06- 23 | 1 | T | | T | | \exists | \top | | T | | T | | | | | | П | | T | | T | | \top | † | | \Box | \forall | | 1 | | | | | | | Т | 7 | | | Ť | \top | | П | \exists | \top | |
| 17 | Text cleaning Text cleaning | 17- 06-23 | 17-06- 23 | 1 | | | | | | | | T | | | | | | | | | | П | | 1 | | | | | | 7 | 1 | 1 | | | | | | | T | 1 | | | 1 | | | | \forall | \top | \top |
| 18 | Word2Vec model | 18- 06-23 | 18-06- 23 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | T | П | |
| 19 | Corpus making | 19- 06-23 | 19-06- 23 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 20 | EDA | 06-23 | 20-06- | 1 | | | | | | _ | | 퇶 | | L | | | | | | | | | | | _ | | L | | | \perp | | \perp | | | | | | | Ц | 4 | | | \perp | | | | | Ц | |
| 21 | EDA | 06-23 | 21-06- | 1 | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | _ | \perp | |
| 22 | EDA | 06-23 | 22-06- 23 23-06- | 1 | | | | | _ | _ | | 1 | | _ | | | | | | | | | | | | | ╄ | | \sqcup | 4 | | 4 | | | | | | | 4 | 4 | _ | _ | \perp | _ | | | 4 | \perp | _ |
| 23 | EDA and twilio | 06-23 | 23 24-06- | 1 | | | _ | | _ | + | | + | | - | | | | | | | | | | _ | | | | | \vdash | 4 | | _ | - | | | | | | | 4 | _ | - | + | - | | | 4 | \dashv | _ |
| 24 | EDA | 06-23 | 23 | 1 | \vdash | | _ | _ | \dashv | + | _ | + | | ╁ | - | | | | | | | \dashv | _ | + | + | | | | \vdash | + | | + | + | | + | | | | + | + | + | + | + | + | | Н | + | + | + |
| 25 | EDA | 06-23 26- | 23 26-06- | 1 | | | \dashv | | \dashv | + | | + | | \vdash | | | | | | | | \dashv | | + | + | | | | | \dashv | | + | - | | | | | | \dashv | + | | | + | | | | \dashv | \dashv | + |
| 26 | Task no 26 | 06-23 27- | 23 27-06- | 1 | \vdash | | + | | \dashv | + | | + | | \vdash | | | | | | | | \dashv | | + | | | \vdash | | | + | | + | + | | | | | | + | + | | | + | | | Н | + | + | - |
| 27 | Task no 27 Task no 28 | 06-23 28- 06-23 | 23 28-06- 23 | 1 | | | \dashv | | \dashv | + | | + | | + | | | | | | | | \dashv | | + | + | + | + | | | | | + | + | | + | | | | \dashv | + | + | + | + | + | | | \dashv | + | _ |
| 29 | Task no 29 | | 29-06- 23 | 1 | | | 1 | | | | | T | | | | | | | | | | | | | | | | | | Т | | | | | | | | | | 7 | | | T | | | | T | \top | |
| 30 | Task no 30 | | 30-06- 23 | 1 | | | | | | | | T | | T | | | | | | | | | | | | | | | | T | | | | | T | | | | | T | | | T | | | | T | Т | |
| 31 | Task no 31 | 01- 07-23 | 01-07- 23 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | T | | |
| 32 | Task no 32 | 02- 07-23 | 02-07- 23 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 33 | Task no 33 | 07-23 | 03-07- 23 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | L | | | | | _ | | | | | | | | Ц | |
| 34 | Task no 34 | 07-23 | 04-07- 23 | 1 | | | | | | | | 1 | 1 | | | | | | | Ш | | | | 4 | 1 | | 1 | | Ц | _ | \perp | \perp | _ | 1 | | | | | \perp | 1 | | 1 | 1 | | | Ш | \perp | \perp | \perp |
| 35 | Task no 35 | 07-23 | 05-07- 23 06-07- | 1 | | | 4 | | _ | _ | - | + | - | - | - | | | | | | | \square | | 4 | 4 | | - | | \sqcup | 4 | \downarrow | \perp | + | + | - | | | | 4 | 4 | | \perp | + | | | Ш | 4 | 4 | + |
| 36 | Task no 36 | 07-23 | 23 | 1 | \vdash | | _ | | _ | + | _ | + | - | - | - | _ | | | | Н | | | _ | + | + | + | - | | \vdash | \dashv | _ | + | + | + | + | | | | + | + | _ | + | + | + | - | Н | \dashv | \dashv | + |
| 37 | Task no 37 | 07-23 08- | 23 08-07- | 1 | | | + | | \dashv | + | | + | + | \vdash | - | | | | | | | Н | - | + | + | | \vdash | | \mathbb{H} | + | + | + | + | + | - | | Н | | | + | | + | + | | | Н | \dashv | + | + |
| 38 | Task no 38 | 07-23 09- | 23 09-07- | 1 | \vdash | \vdash | - | | - | + | + | + | + | + | \vdash | \vdash | \vdash | | | \vdash | | \vdash | \dashv | + | + | + | + | | \vdash | + | + | + | + | + | + | | | - | | | + | + | + | + | \vdash | \vdash | \dashv | + | + |
| 39 | Task no 39 | | 23 | 1 | \vdash | \vdash | + | | \dashv | + | + | + | + | + | \vdash | \vdash | | | | Н | | \dashv | \dashv | + | + | + | + | | \forall | \dashv | + | + | + | + | + | | | - | + | ١ | | + | + | + | + | Н | + | + | + |
| 40 | Task no 40 | 07-23 11- | 23 11-07- | 1 | \vdash | H | 1 | | \dashv | + | + | + | + | + | | | | | | Н | | H | 1 | + | + | + | + | | H | \dashv | + | + | + | + | + | | | | + | + | | | + | + | | Н | + | + | + |
| 41 | Task no 41 Task no 42 | 07-23 12- 07-23 | 12-07- | 1 | | | + | | + | | | $^{+}$ | | + | H | | | | | Н | | \vdash | \dashv | + | + | | + | | \forall | \dashv | \dashv | + | + | | + | | | | 1 | \dagger | | | | | | Н | + | + | + |
| 43 | Task no 42 | | 13-07- 23 | 1 | | | 1 | | \neg | \top | | T | | T | | | | | | Н | | Н | | \top | \dagger | | | | H | \dashv | \dashv | \dagger | \dagger | | 1 | | | | 1 | \dagger | | | | | | Н | \forall | \forall | + |
| 44 | Task no 44 | 14- 07-23 | 14-07- | 1 | П | П | | | | \top | T | T | T | T | T | | | | Г | П | | П | \exists | \top | \top | | T | | | \dashv | \forall | \top | \top | \dagger | T | | | | T | 1 | | \top | | | | П | \forall | \top | \top |
| 45 | Task no 45 | 15- 07-23 | 15-07- 23 | 1 | П | | | | | | T | T | | | | | | | | П | | П | | | | | | | П | 1 | | 1 | | | | | | | T | 1 | | | | ſ | | П | \top | \top | 1 |
| 46 | Task no 46 | 16- 07-23 | 16-07- 23 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 47 | Task no 47 | 17- 07-23 | 17-07- 23 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | I | | | I | I | | | | | | |
| 48 | Task no 48 | 07-23 | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | Ш | | | |
| 49 | Task no 49 | 07-23 | | 1 | | | | | | | | 1 | | | | | | | | | | | | 4 | 1 | | | | | _ | | 4 | | 1 | | | | | 4 | 4 | | | 1 | | | | \perp | _ | |
| 50 | Task no 50 | 20- 07-23 | 20-07- | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | \Box | |