**IITD Machine Learning-01(chatbot)[INTP23-ML-1]**

**Day6:**

On day 6, of IITD-AIA FSM internship, I learnt NLP module.

**Learning:**

So basically I learnt NLTK module, where I studied various new things tokenizer, stemming, Regular expressions use of ‘re” library and how to extract data using regular expression

Also learnt the lemmatization and difference between lemmatization and stemming.

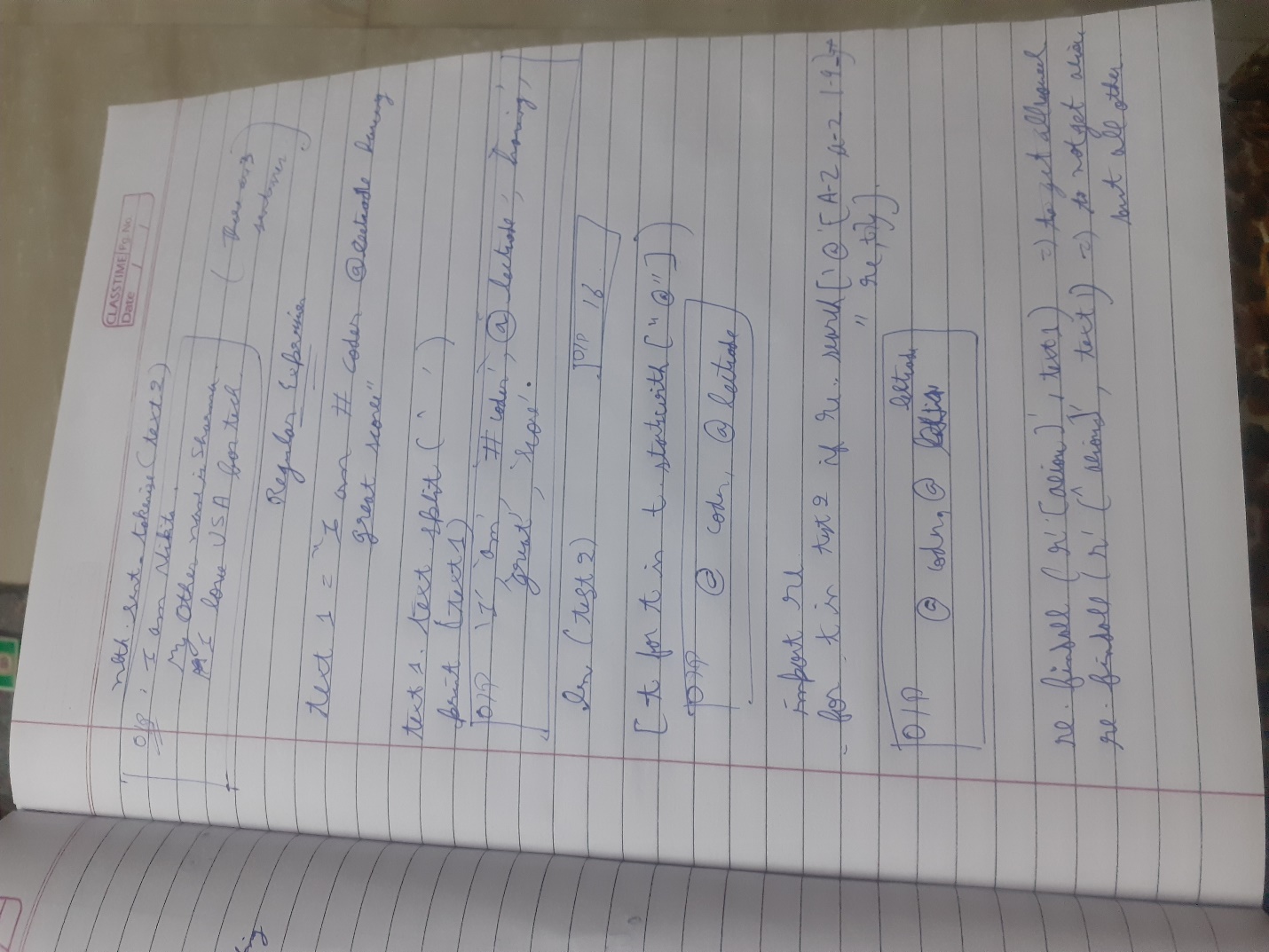
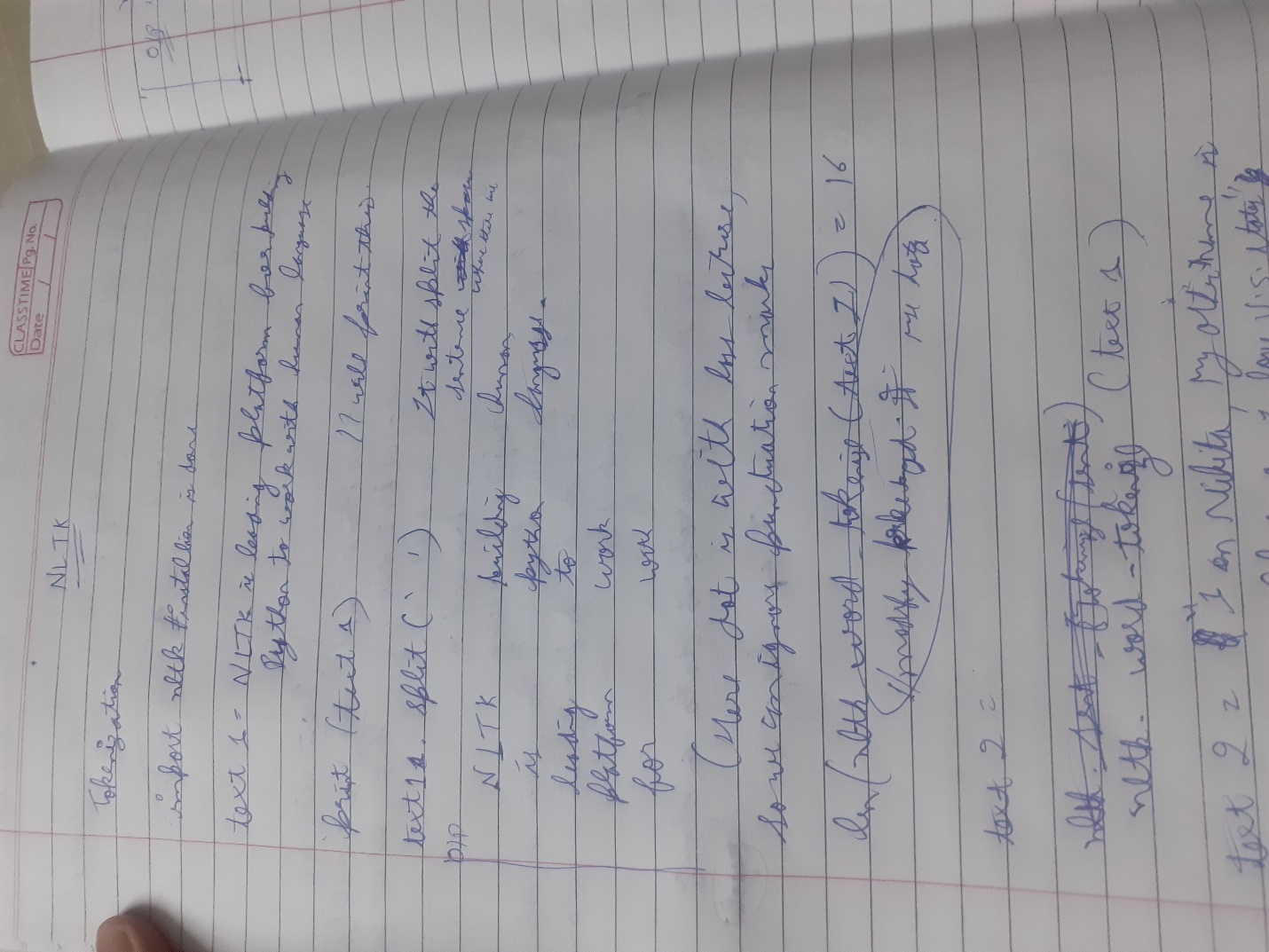
And after reading comment/feedback on yesterday’s report I researched for research papers that will improve the efficiency of model.

While searching for research papers, meeting was arranged with my mentor, where I got various doubts resolved. Like web scrapping for data collection here we got two ideas either web-scrap the whole site and store in “.json” to further process the data or copy the website content and then paste in document and process that data.

Along with this I got various topics to look upon for providing output and increasing accuracy that were : attention mechanism, transformer and study how ChatGPT worked and do something related to that. Also, my mentor mentioned to learn and use RNN for learning and predicting/ generating answers.

Also, We got idea for basic common questions like greeting and goodbye we can take data from Kaggle so I go through that data.

I learned new term EDA (exploratory data analysis).So now I will go through new deep learning concepts and implement the same using Kaggle today and tomorrow and after successful web scraping I will add this scrapped data to further train the model.



**Work Done/learning Implemented:**

Today, I learnt exactly how tokenization, stemming, lemmatization is done. Also, I learnt new terms and concept to increase the accuracy of model and got resolved various doubts.

**Is Progress As per Track?**

Yes, I tried with NLP part and now after writing the report I will continue learning some new concepts which I got to know in meet.

**Issues Faced Today:** issue was about data collection, search text is n

**Issues Closed Today:**

**Highlights:**

So basically I learnt NLTK module, where I studied various new things tokenizer, stemming, Regular expressions use of ‘re” library and how to extract data using regular expression

Also learnt the lemmatization and difference between lemmatization and stemming. use RNN for learning and predicting/ generating answers.

**Concluding the day:**

Today, I got to know various new terms to learn, which will improve our accuracy for model /bot and also, I learnt how properly tokenization, stemming, lemmatization is done and how human text is handled.