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

**Day 12:**

On day 12, of IITD-AIA FSM internship, I scrapped out almost complete website.

**Learning:**

Today, I solved the problem mentioned in yesterdays’ diary, first is to extract out all the “a” tags even after utf-8 and print it, for this first I changed the scraped data encoding to “utf -8”

Using” encode()” function and then I found all the “a” tags are get them using for loop.

2nd is I extracted out all the “href” by using get () function and also searching only “http” links and scraped them also while scraping dashboard’s link we get the information of each page and that pages’ link are also extracted.

So, now I scrapped out whole “FSM” site

**Tomorrow’s Objective: now, I have to store this scrapped out information in “json” file.**

**Problem: what information to be stored is not clear to me, right this moment will take assistance from my mentor in that information.**

**Work Done/learning Implemented:**

Implemented web-scrapping on the FSM website through Beautiful Soup, and scrapped complete site.

**Is Progress As per Track?**

Yes, the processing of the data, in order to apply NLP, EDA

**Issues Faced Today:** putting the scrapped data into “.json” file

**Issues Closed Today:** solved encoding problem, extracted all links

**Highlights:**

Today, I solved encoding problem, it was not printing the data but now it is printing links present in “dashboard” page and all these links information is parsed and completed scrapping of website.

**Concluding the day:**

So today, I applied and solved the web scrapping of site.