Salient features-

1. I have used os.chidr from os library and for loop with a range of 0 to the number of program files, to traverse through the files.
2. After traversing through the files, I concatenated all the data.txt files from the program files into one data frame called df. After which I replaced N/A values with 0.001.
3. I have read the features.txt file into a table using pd.read\_table and removed the duplicates.
4. Using the features table, I have selected the desired features from df.
5. After which I created the required text files after finding the maximum values and normalizing the data
6. To create the data.txt file, I had to use a for loop and a simple logic to draw the desired data from df3 (which is a data frame containing the normalized values.)

Challenges Faced-

As this is my first time working on an assignment in python, I have faced a lot of difficulties, which I have listed below and also how I overcame them.

1. The first biggest challenge I have faced is, to find a way to traverse through the directory using the programnames.txt file to collect the desired data. I tried using readme but that didn’t workout for me. Finally, after searching the internet I found a way to traverse through the directory using os.chdir (I know this is not an effective way of traversing through the directory but it worked out for me).
2. The next challenge I faced is the duplicates in the features.txt file, which I didn’t take into account due to which I didn’t get the desired output. But, after removing the duplicates I was back on track.
3. The third challenge I have faced is while calculation the maximum values for the features. For some reason after executing the code python showed an error stating that my values column datatype was string. To overcome this I have changed my value column data type to int.
4. My last challenge was to create data.txt file from df3(data frame that contained normalized values). It was problematic because the data frame contained all the normalized values of all the features which made it difficult for me separate the values for a particular program name. I over came this by using a for loop and a simple logic.

GITHUB-LINK –

https://github.tamu.edu/n26111997/-CSCE633\_S20\_KIRAN\_KONDISETTI.git