Telecom Churn Prediction Project Documentation:

During the development of this application I took lots of steps from data exploration to creating public git hub repository. Below are the steps that I did to accomplish the task. They are briefly explained below:

1. Data Load: First I download the data into my local machine and read the data using pandas read_csv method as the given data set was in csv format.
2. I explore the data like, total rows and columns, data types of each variable, null values, duplicates values, summary statistics etc.
3. Data Clean: In this step we had not much more to do as the data set was quite clean, no null values, no duplicates values, no need to change the data types of variable etc..
4. Data Visualization: I created lots of different charts to see the pattern and data variation in data set. Like box plot, count plot , hist plot, rec plot etc…
5. Data scaling: I perform data scaling methods to keep the respective data in same range.
6. Data transformation: I transforms some data as I found some variables needs to transform.
7. I apply feature selection method as to check the relationship between variables.
8. Data preparation: I prepared the data by converting the categorical variables into numerical variables.
9. I split the data set into train and test part.
10. I train the model using supervised ml algorithms as it classification problem
11. I validate the model apply cross-validation technique.
12. I test the model on testing data.
13. I evaluate the model by creating different evaluation metrics to check if the model performs well.
14. I create the api end point of model to develop in web using flask.
15. I hosted the api end point in docker by creating docker image.
16. I push the files including docker file into my github public repo name Nahakul-prasad-jaishi-mk-task.
17. I did commits and created branches of github.

These are the almost steps I did for this project. For more details about these steps you can go to the related files.