Framing the Problem in ML (Steps)

1. **Business Problem to ML Problem:**

First thing in framing the problem that you have to convert your problem into the mathematical form. Means you can speak of it, calculate it in terms of math. Like if you want to increase the revenue of e-commerce platform what you can do for it. Let’s assume you suggest to decrease the churn rate (Leaving of the customers from the platform) of customer.

1. **Type of the problem:**

You have to look into the big picture of the problem and also keep in the mind that what will be the end product of your problem. Let say we are trying to reduce the churn rate. For this we have to identify the potential customer which are going to leave our platform. From this it will become classification problem. Now what we can do is to give some discount to leaving customer so they can stay. Now what if we think that there must be some customer who are not surely decisive to leave. Means their probability of leave is not so high they why we give equal discount to all of them. Here we can decide to give the discount to our customer according to their probability of there leaving. Now here it will become the regression problem.

1. **Current Solution:**

You can take help if someone had already done some work on current. Let’s say someone has train the model to predict the churn rate.

1. **Getting Data:**

You have to spent of your time in this step. Here you will try to find out what type of data you want to gather for reducing the churn rate. You have to think all the factors for your data. Here data engineer may come into the scene.

1. **Metrices to measure:**

At this step you have define some measure matrix for your predictions to access whether you are moving in right direction or not. Means if have predicted 2% churn rate and churn rate becomes 2.3% then it means you moving in almost right direction etc.

1. **Online Vs Batch:**

You have to made decision to which type of learning you will use. Will it be online or batch learning?

1. **Check Assumptions:**

Check your assumptions if they are working or not.

**Date Gathering**

**Data Gathering:** CSV, JSON/SQL, From API, Web Scrapping

1. **Working with CSV Files:** see the Folder (Data Gathering)
2. **JSON / SQL:** It is universal form of object. Every programming language can understand this notation. When ever you have data on server you can fetch its data through API using JSON notation. See the folder for examples.
3. **Fetching from API:** APIs are used for communication between two applications. They are basically a data pipeline. See folder for examples.
4. **Web Scrapping:** If we can get data from API or any other direct source then we can scrap the web site. See the folder for example

**Understanding Your Data**

1. **Asking Basic Question About your Data**
2. **EDA (Univariant Analysis)**
3. **EDA (Multivariant Analysis)**
4. **Pandas Profiler**
5. **Basic Questions**: See the Folder
6. **Univariant**: Whenever you do the analysis of single variable that is called univariant analysis. When you done analysis on two variables is called bivariant analysis. Before going you have to know the data types. There are basically two types of datatypes numerical and categorical. For Example, see the folder.
7. **Multivariant / Bivariant**: See the folder