- NIRAJAN RIJAL

ML lifecycle:

1. Frame the problem
2. Gathering the data/importing the data from various sources
3. Data Preprocessing
4. Exploratory Data Analysis
5. Feature Engineering and Selection
6. Model Training and Evaluation
7. Model Deployment
8. Testing
9. Optimize

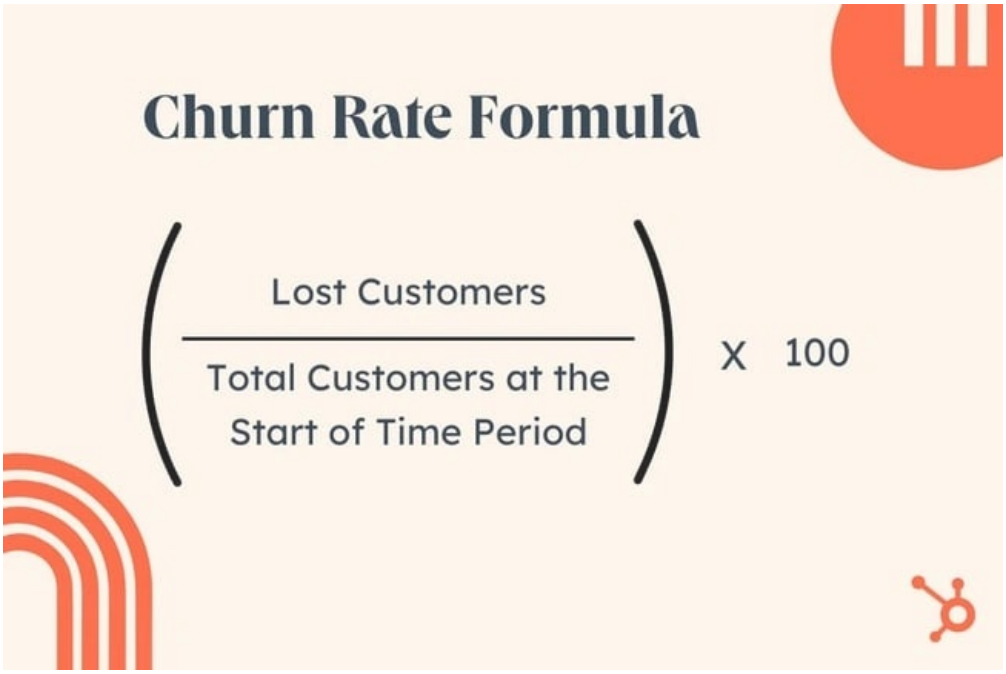
Today we will be focusing on first point: Frame the problem

Suppose we work Data scientist in amazon prime, and now CEO is telling us, we have to increase revenue of the company. There may be 3 solution.

- Increase price of subscription/membership
- Strategy to increase new customer
- or, the customer which may be leaving amazon prime, stop them or retention

So, now we decide third solution can be better option for now.

- First step is look for the churn rate. Churn rate: The annual percentage rate at which customers stop subscribing to a service or employees leave a job.



So, this is real news of 2023: Over the past five years, member retention after one year has ranged from just under 90% to over 95% in the past two years of Amazon Prime.

Means churn rate can be like 4-5%, so in meeting they decide it to bring down by 1-2% atleast.

So, in first step of framing problem is - identify the main problem and try to convert normal problem to mathematical or ML problem

Second Step: Type of Problem

Look the problem, is it supervised or unsupervised, if it is supervised - it is regression or classification. Try to think end goal - Decreasing chunk rate - means we have to know - who may leave subscription in coming month or year

Like: First of all lets predict, in supervised ML (in classification) - This customer may leave the Amazon Prime - YES/NO, Secondly, look and do regression model for again to them with YES - what is the probability that this customer may leave Amzon prime.

Third Step: Solution (Think Current solution, it may not be Final)

Talk with team, how has predict chunk rate and give your solution too in meeting like - if with regression, you are getting different probability % for different customer and try to mix the solution.

Fourth step: Getting the Data

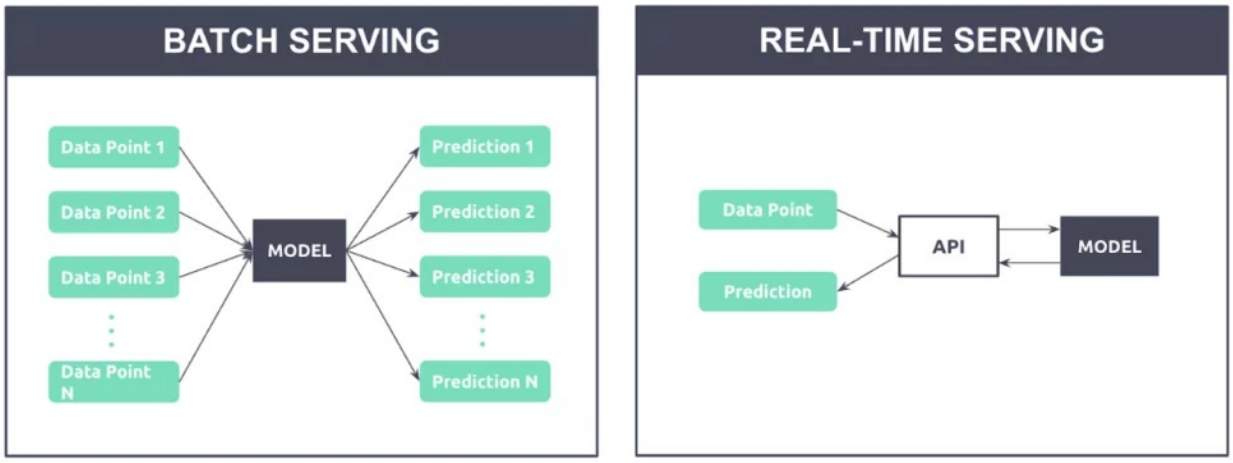
- Watch Time
- Search movie but did not find in our platform
- Content left in the middle, did not finished watching it
- Click on recommendation

Fifth Step: Metrices to measure

Just look if we are in good direction or not: look the actual chunk rate and preiction one and actual leaving customer data. Compare it for first month when we are working on solution, if percentage difference is low then we are going towards right direction. There may be other way too according to business houses and problem.

Sixth Step: Online VS Batch

Online Learning will be better in this case:



Seventh Step: Check Assumption