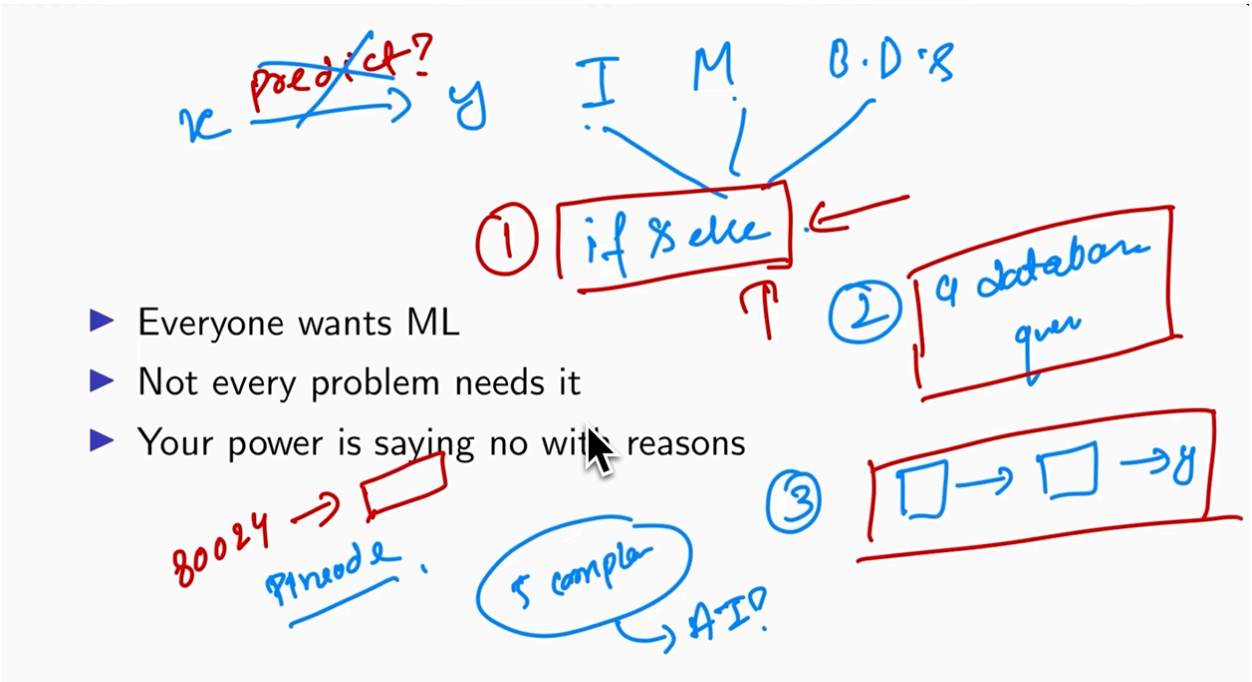


Let us say you are given a problem :

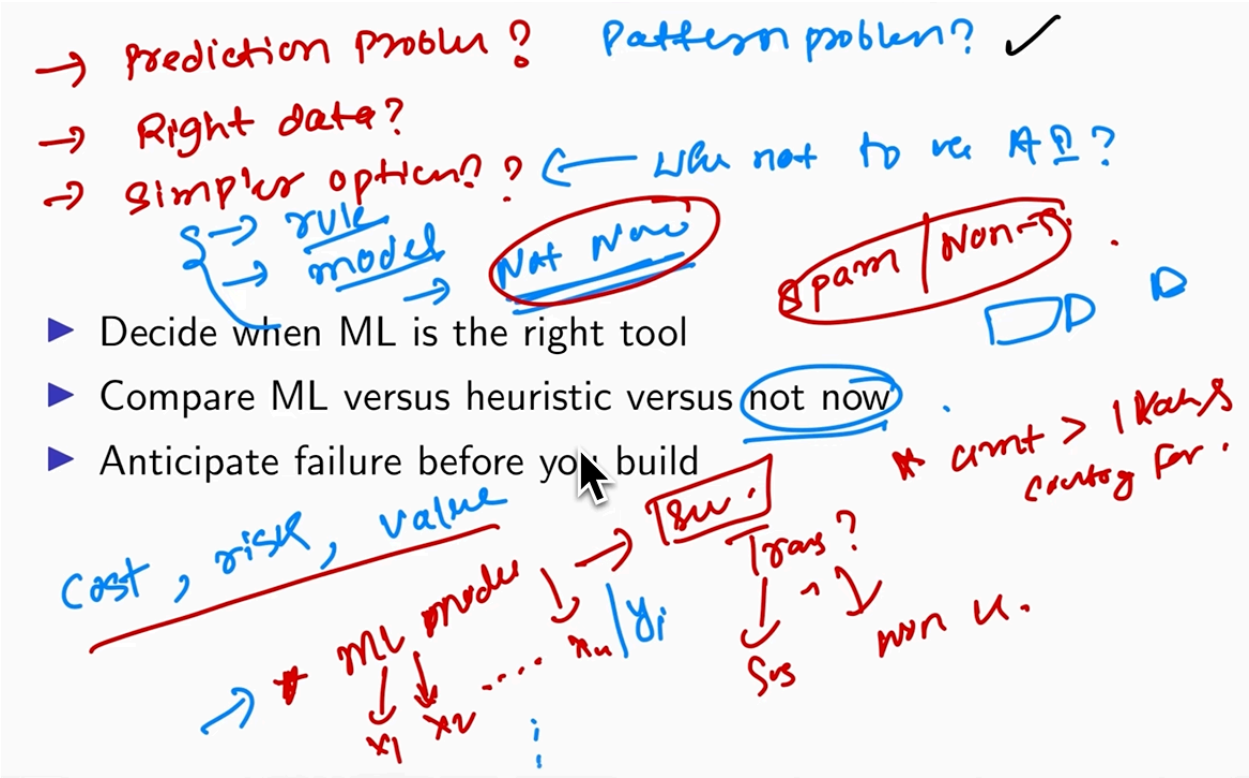
1. Convert pincode to city via an ml model, but why do u need ml for that ? Just look up the database
2. First see if it can be solved by rule based systems like if else
3. See if database querying (sql query) can be done ?
4. Or see if a simple process can be carried out ?

Let us say the business tell we get 5 customer complaints a month, we need ai to handle that . But why ? its just 5 complaints, use ai only when u get around 500-1000 complaints a day. Even then classify it, if there should be human intervention or not.

Not every problem is a prediction problem.
Every ai output is not 100% accurate, so if u want accurate results like pincode to city, just do a database lookup ? Simple.

This is what I have, do we have a sustainable business outcome if i build it ?
Is there a revenue outcome if i build it ?

AI cannot take business decisions.
We should be able to tell we dont need LLM's for this, we can just train a simple ML model, which gives a pickle file in kb's
Cost and complexity reduced.

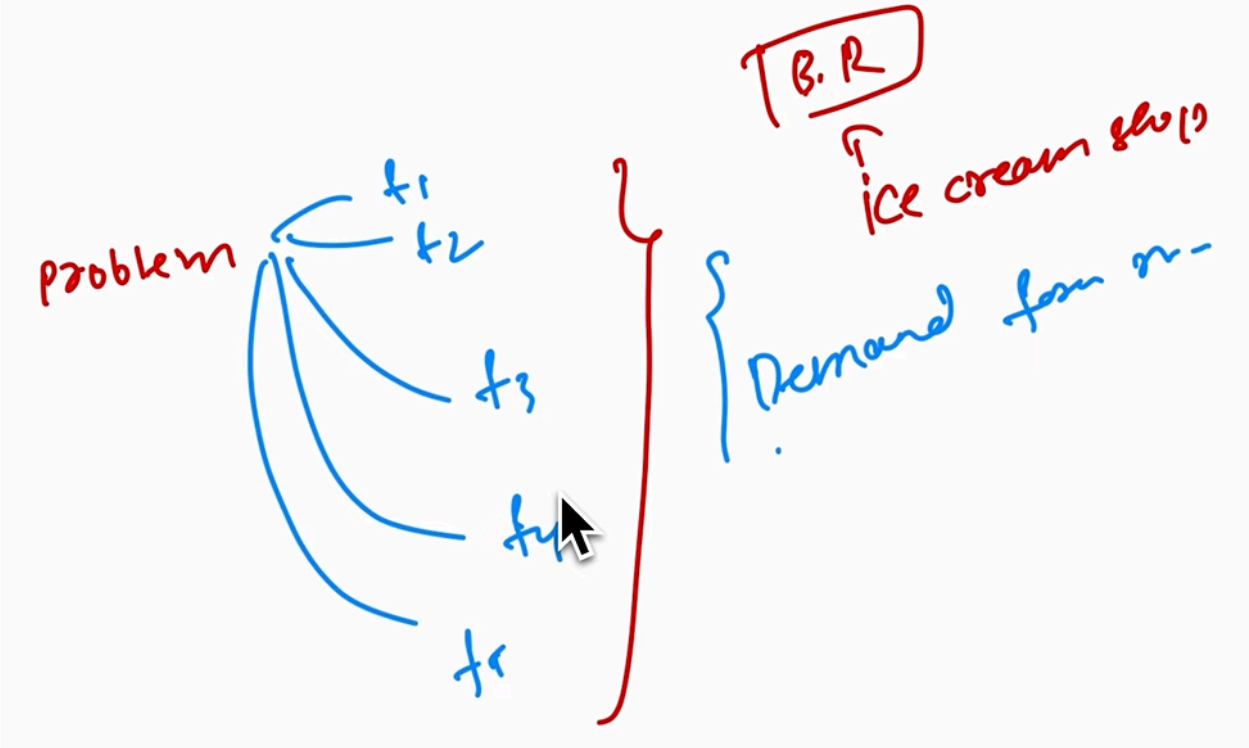


1) Is ask if it can be done using a rule based system
2) Train a simple ml model
3) Not now -> is we dont have enough data, so first with the help of human reviewers collect the labelled data and then do step 2

Before writing the code, identify where the model can fail ?

1) Can the labelling made by human reviewers be wrong ?
2) Is there data drift ?
3) How can people misuse the system ?
4) How can people identify patterns which my model would have learnt ?
5) Etc…

Think about failures early.



Demand Forecasting model -> Anticipating failures :

1) What if there is a sudden heatwave
2) What if a new competitor opens next door
3) Model trained on last year data, but in this year several icecream shops have opened nearby
4) Sudden cold -> demand less

Why ml projects fail ?

Talk to users, domain experts, operations.

1. How long the model takes to respond ?
2. Do we have the budget limits ?
3. Do we have the regulations ? GDPR, HIPAA.
4. Are they compliant ?
5. What is the latency ?
6. How the data flows in and out ?
7. What are the limits ?
8. Etc….
9. Store data in my region only ?

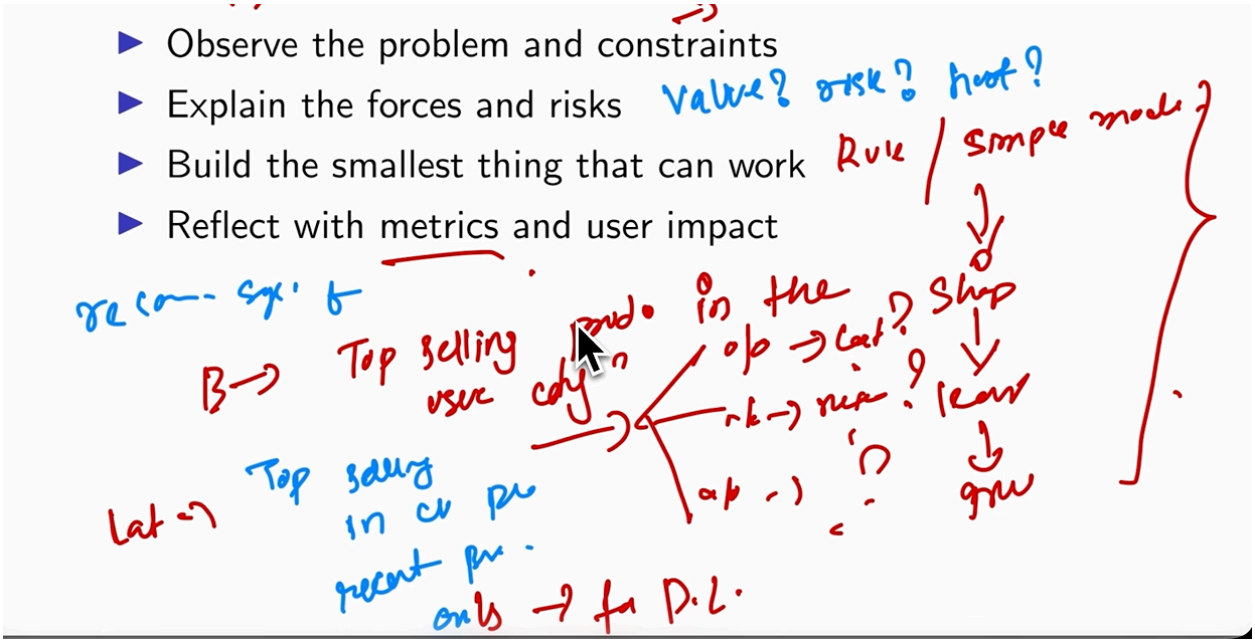
Your response must be under milliseconds - for ex : recommendations
Let us say you have to deploy your model in China, then you have to pick China Datacenter.
We have datacentres in europe, USA,etc.

So you have to take the instances of AWS of these data centres

We have to be able to tell :

1. What value are we chasing ?
2. What costs and risks are we taking ?
3. Who will get hurt if the model is wrong ?

Ex : Loan approval model :
Value : faster decisions and less manual work that anybody who wants a loan can quickly come in and get it
Risk : unfair denial to good customers
So we need to care about interpretability and bias, not just accuracy



1) Rule/simple model -> ship -> learn -> grow (Baseline)
   For ex : recommender system
   Baseline : Top selling products in the user's city (a simple database query)
   And then analyse the cost, revenue and metrics, if it is okay proceed with
   Later : Top selling products in city + recent browsing history of user
   Again analyse cost,revenue, metrics
   Then only go for fancy deep learning models.