**Personal Notes**

* Simple ML model doesn’t suffice specifically in the case of healthcare and financial field.
* Finance – different stakeholders (Data Scientist, Business Stakeholders, Regulatory bodies)
* AI regulations to make sure model is not biased and model comply with the regulations.
* When individuals understand the reasoning behind an AI-driven decision, they are more likely to trust and accept the outcomes.
* Deep checks to find segments in the model where the ML model doesn’t work or faulty (Beneficial for stakeholders).
* Model working on global level (How some change of value in one feature impact our prediction via SHAP)
* What are the rules that if the data follows will always have the same predictions. E.g. when my age is this and gender is this the output is always this.
* Counterfactuals – if my prediction is default what can I change to get the model as not default.

Loan information: user\_id , Loan category, Amount, interest rate, Tenure(years)

Employment: user\_id, Employment, Tier of Employment, industry, Role, Work Experience, Total Income

Personal Information: user\_id, Gender, Married, Dependents, Home, Pincode, Social Profile, is\_verified

Other\_information : user\_id, Deling\_2yrs (how many times the payment has been delayed), Total Payment, Received principal, Interest Received, Number of loans, Defaulter

-Done with:

1) Missing values

2) Skewness

3) Encodings

4) Hande imbalance using ( from imblearn.over\_sampling import RandomOverSampler )(SMOTE – revise)

 oversampler = RandomOverSampler(random\_state=42)

  X, y = oversampler.fit\_resample(X, y)

Two types of oversampling:

1. Random Over Sampling
2. SMOTE (Synthetic minority oversampling techniques) : A drawing of a triangle with dots and lines on a piece of paper

   Description automatically generated

Here, 3 is represented as the K nearest neighbors. First split the data and apply smot on training data

MODELING:

1. Neptune: Keep track of the models
2. Hyperpot: Best parameters