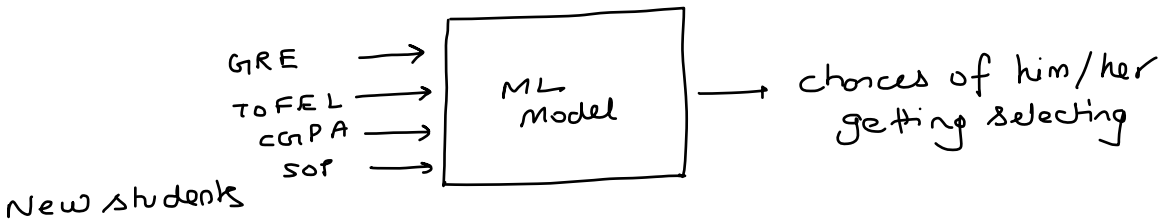
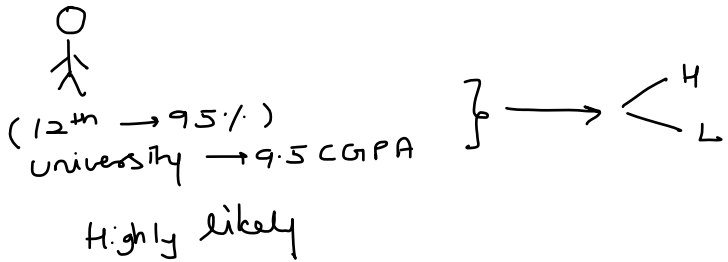
-3 ——— +3
0 ——— 1

Jamboree →



Target → continuous → Linear Regression



multicollinearity didn't impact performance of the model

But

Removing multicollinear features would reduce the complexity of the model.

Assumption

$y = mx + c$

$y = m_1 x_1 + m_2 x_2 + \cdots m_n x_n + c$

$m_1 \quad (x_1 + x_2 + \cdots x_{500}) \longrightarrow 0.91$

$m_2 \quad (x_1 + x_2 + \cdots x_{10}) \longrightarrow 0.85$

Is multicollinearity a problem?

Rock band

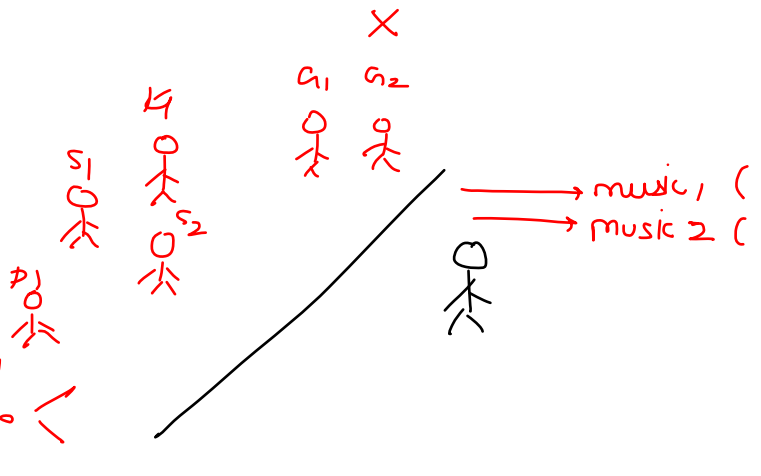
x_1 1 keyboard

x_2 2 Singers < male, female

x_3 1 Drum

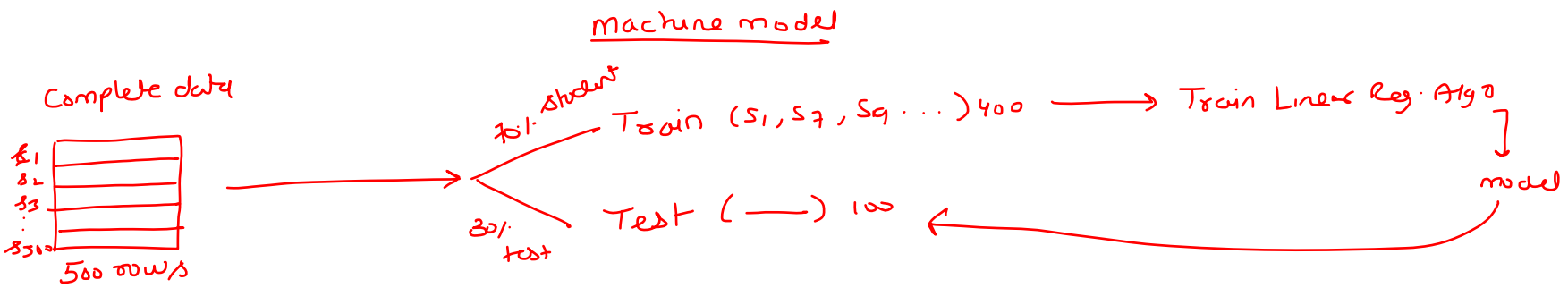
x_4 2 guiter — playing at speed, playing same tone, playing at some pitch

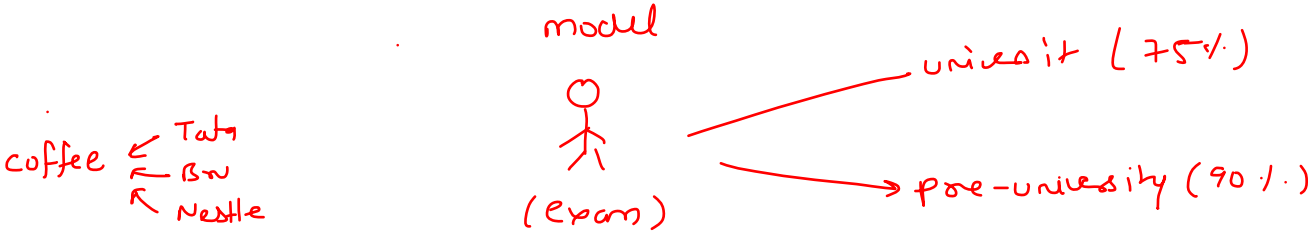
x_5



$x_1 + x_2 + x_3 \longrightarrow R^2 = 0.85$

$x_1 + x_2 + x_3 + x_4 + x_5 \quad R^2 = 0.86$





model
- train → 0.83
- test → $R^2 = 0.85$

} good model

Insignificant var.

$y = mx + c$

m → $-$, $+$

$H_0: m = 0$ (Independent var is not influencing y)

$H_a: m \neq 0$ (Independent var is either +vely or -vely impacting target y.

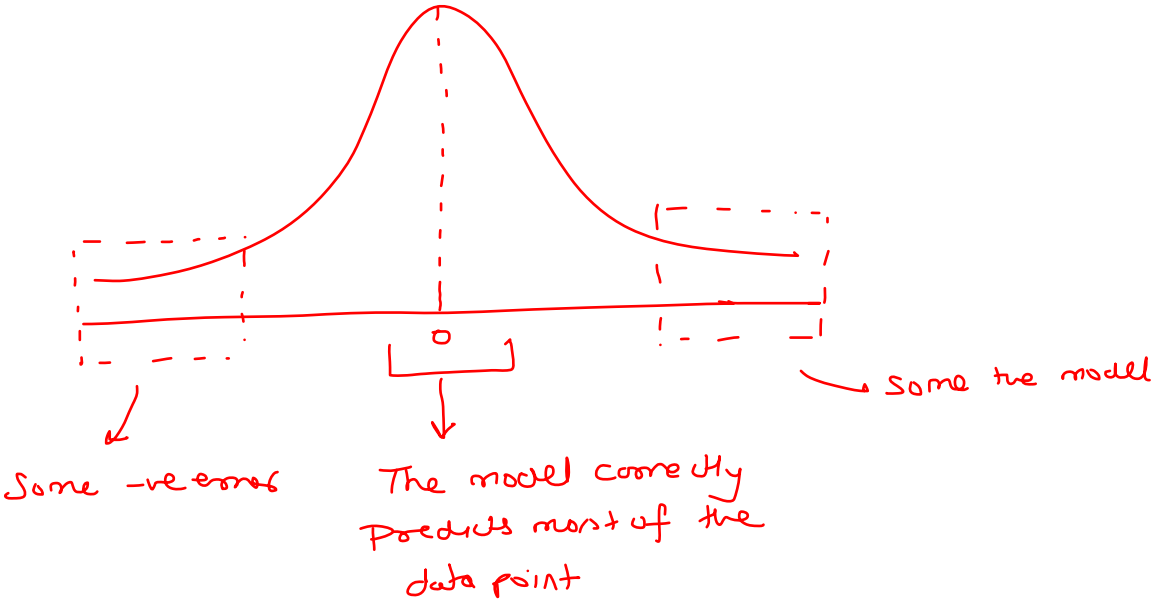
Test (T-Test)

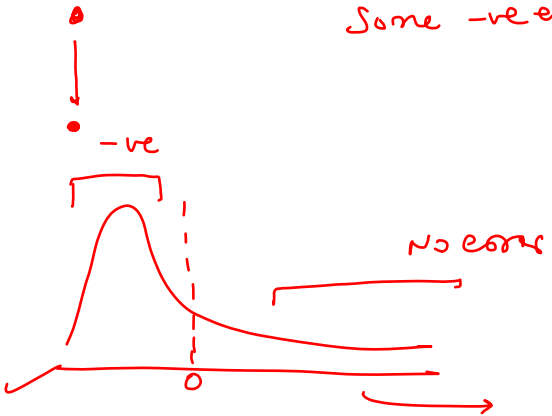
$\alpha = 0.05$

$p < 0.05 \longrightarrow$ Reject null

$p > 0.05 \longrightarrow$ fail to reject null

residual = (y - actual) - (y - pred)





Introduction

When applying for a bank loan, there is a lengthy process involving the submission of various documents and details such as a good credit score, a no dues certificate, and sometimes even a no crime certificate. However, not everyone is eligible for these loans and able to meet all the documentation requirements and profile evaluations.

Loan Tap, as a company, aims to provide loans to individuals and small-sized firms (MSMEs) who may not be able to fulfill all the necessary requirements. They specifically target high-risk profiles as their potential customers, but within this category, there are three sub-profiles that loan applicants can fall into.

The first category is the "White Collar" customers, who are considered high risk but are expected to repay their loans. The second category is the "Grey Collar" customers, who have a mixed likelihood of repaying the loan. Lastly, the "Black Collar" customers are those who are unlikely to be able to repay the loan.

Loan Tap focuses on the White Collar and Grey Collar customers only. However, they cannot approve loans for all Grey Collar customers. This is where data scientists come in. They are needed to properly profile the Grey Collar customers so that Loan Tap can have a balanced mix of high-risk and low-risk customers, allowing them to potentially make a profit even if a few customers default on their loans.

“Derogatory” is seen as negative to lenders, and can include late payments, collection accounts, bankruptcy, charge-offs and other negative marks on your credit report.

The debt-to-income (DTI) ratio measures the amount of income a person or organization generates in order to service a debt. A DTI of 43% is typically the highest ratio a borrower can have and still get qualified for a mortgage, but lenders generally seek ratios of no more than 36%.

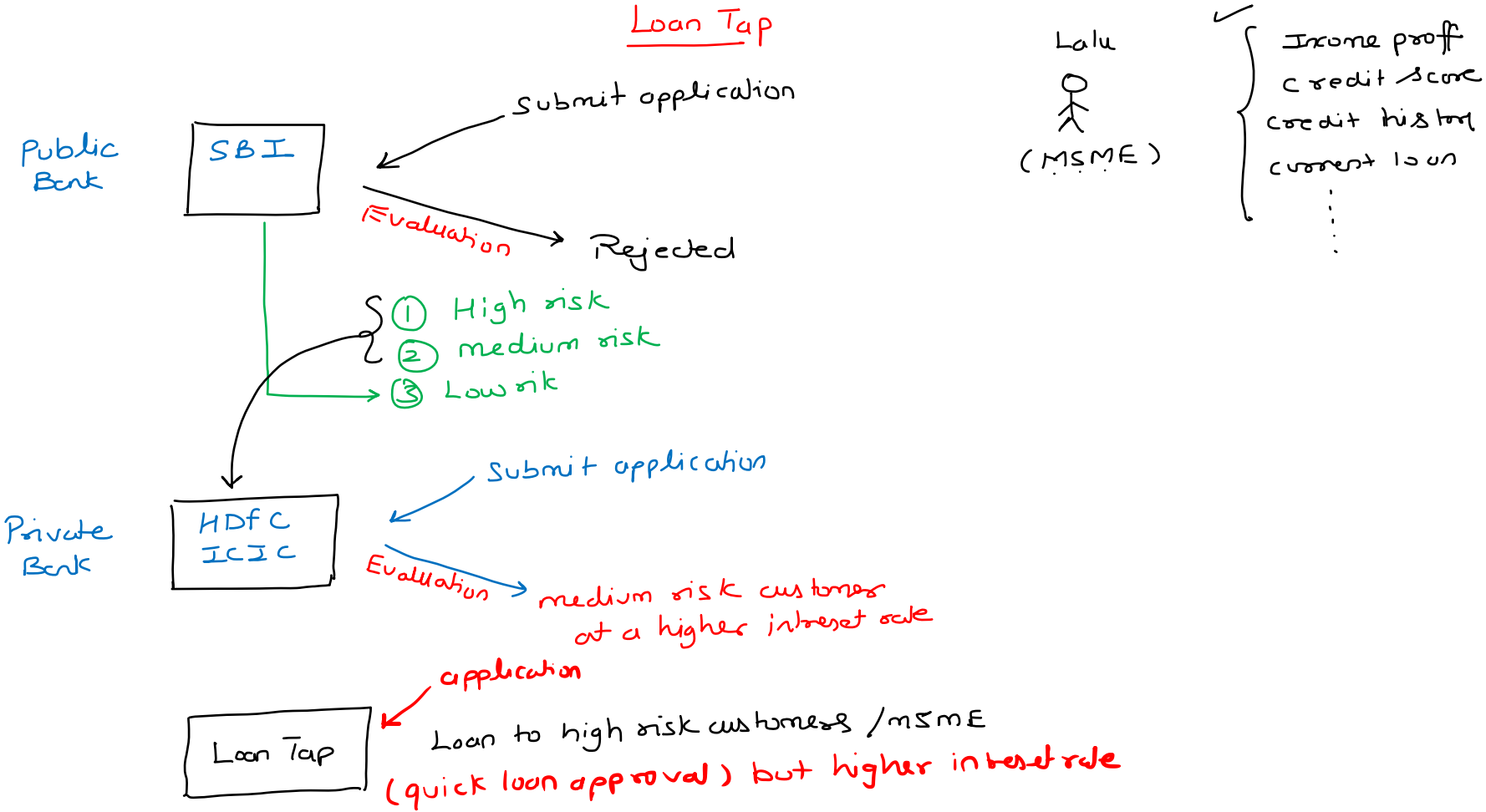
To calculate your DTI, you add up all your monthly debt payments and divide them by your gross monthly income. Your gross monthly income is generally the amount of money you have earned before your taxes and other deductions are taken out.

What is a revolving balance? BALANCE? With revolving credit, a consumer has a line of credit they can keep using and repaying over and over. The balance that carries over from one month to the next is the revolving balance on that loan.

Bankruptcy is a legal proceeding initiated when a person or business is unable to repay outstanding debts or obligations. It offers a fresh start for people who can no longer afford to pay their bills.

The variable initial_list_status is available in the public data and identifies whether a loan was initially listed in the whole (W) or fractional (F) market. Loans listed “whole” become available for fractional funding (and vice versa) if there are no buyers within a certain time frame.

Loan Tap


Public Bank
SBI
Submit application
Evaluation
Rejected
1 High risk
2 medium risk
3 Low risk
Private Bank
HDfC
ICIC
Submit application
Evaluation
medium risk customers at a higher interest rate
application
Loan Tap
Loan to high risk customers / MSME
(quick loan approval) but higher interest rate
Lalu
(MSME)
Income proff
credit score
credit history
current loan

High risk customers

1. White collar → 100% capable of paying back loan
2. Grey collar → 50-50% ← either pay default
3. Black collar → clearly not going to pay back

help loan tap to come up with a classification that can help them to approve or reject loan of these customers.

$$P(\text{defaulting} \mid X)$$

↳ i/p variable

Rbi

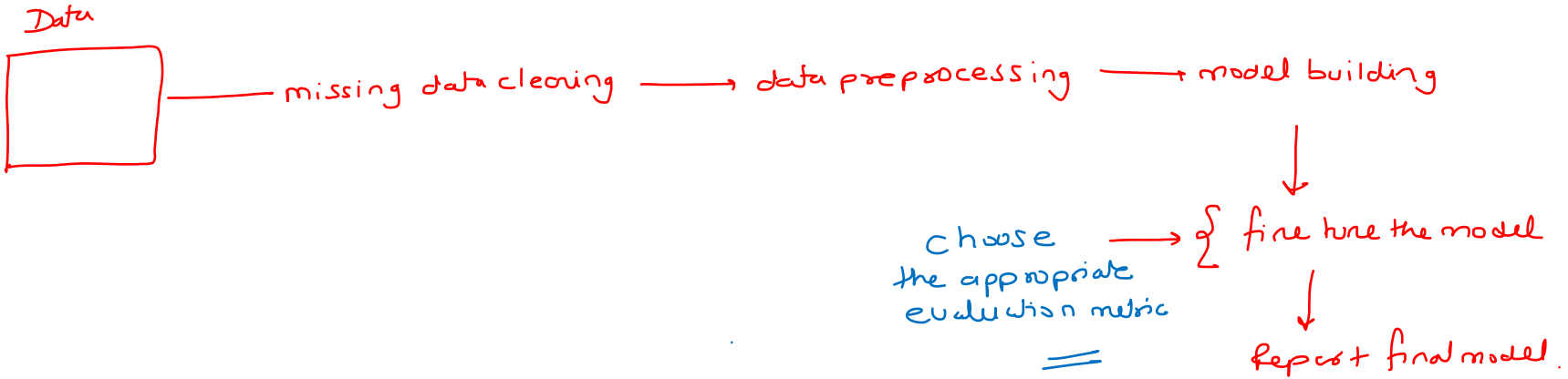
$$dti = \overline{60\%} = \frac{\overline{\text{total debt}}}{\text{total income}}$$

↑ ↓

[Uni
Slice] they are not credit cards

→ Credit line

→ 20,000 → 100%


Data
missing data cleaning
data preprocessing
model building
choose the appropriate evaluation metric
fine tune the model
Report final model.