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.

Evaluation-metric

(Banking-application) → F1-Score

① Cancer Detection model
- cancerous — 1
- non-cancerous — 0

FN: Actual: cancerous (1) → Predicted: non-cancerous (0) → can lead to death because we may miss out a cancerous patient

Fp: non-cancerous (0) → cancerous (1) → can be detected in next round of manual testing

$$\text{Recall/Sensitivity/TPR} = \frac{TP}{TP+FN} \qquad \uparrow R \propto \frac{1}{FN\downarrow}$$

② Email Spam-ham Detection model
- Spam — 1
- not-spam(ham) — 0

FN: Actually Spam → Predicted: not-Spam → manually ignored by the user

Fp: Ham → Spam → email goes to spam folder without notification and hence user may miss out some important news/notice

$$\text{Precision} = \frac{TP}{TP+FP} \qquad \uparrow P \propto \frac{1}{FP\downarrow}$$

③ fraud transaction Detection model — fraud – 1, valid – 0

	Actual		Predicted	
FN:	fraud	→	valid	] loss of the bank
FP:	valid	→	fraud	] good customer getting bad banking experience Can lead to a customer loss

Reduce both FN & FP

Reduce → FN → Improve Recall
Reduce → FP → Improve Precision
} Improved simultaneously

$$f1\text{-}score = \frac{2 \times P \times R}{P + R} \quad \text{(Harmonic avg)}$$

Class - Imbalance problem

- Random oversampling
- Random undersampling
- SMOTE

} Something done to the data

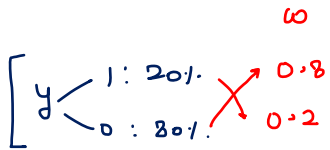
Data (y) < 1 : 20%, 0 : 80% } → (Algorithm) → ML-model

The algorithm is exposed to more no. of examples from negative class and very few from negative

Very good for class - 0 but will fail for class - 1

weight of class (nothing done to the data but to the algorithm)

$$\mathcal{L}(\beta) = -\left[y_i \log(y_{(x)}) + (1-y_i) \log(1-y_{(x)}) \right]$$



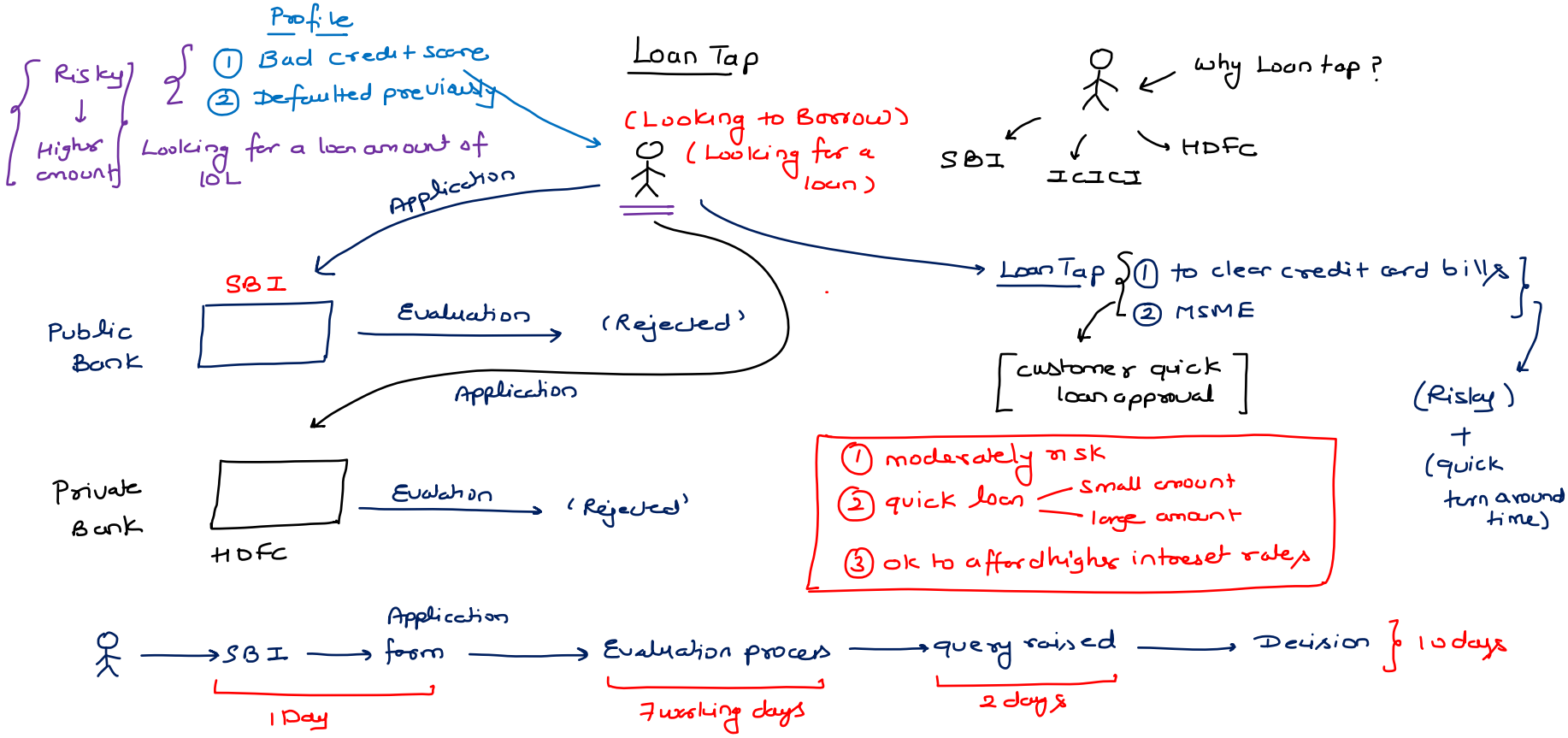
$$\mathcal{L}(\beta) = \begin{cases} -\log(y_{(x)}) & \text{if } y_i = 1 \\ -\log(1-y_{(x)}) & \text{if } y_i = 0 \end{cases}$$

$$\mathcal{L}(\beta) = \begin{cases} -0.8 \log(y_{(x)}) & \text{if } y_i = 1 \\ -0.2 \log(1-y_{(x)}) & \text{if } y_i = 0 \end{cases}$$

✓

Logistic Regression (w class_weight = {0 : 0.2, 1 : 0.8})

class_weight = 'balanced'


Profile
Risky
Higher amount
1. Bad credit score
2. Defaulted previously
Looking for a loan amount of 10L
Loan Tap
(Looking to Borrow)
(Looking for a loan)
Why Loan tap ?
SBI
ICICI
HDFC
Application
SBI
Public Bank
Evaluation
(Rejected)
Application
Private Bank
Evalution
(Rejected)
HDFC
Loan Tap
1. to clear credit card bills
2. MSME
[customer quick loan approval]
(Risky)
+
(quick turn around time)
1. moderately risk
2. quick loan — small amount / large amount
3. ok to afford higher interest rates
SBI
Application form
1 Day
Evaluation process
7 working days
query raised
2 days
Decision
10 days

Profile of the customer

public Bank / private

1. white collar → capable of paying back the loan ← good credit profile, good customer
2. Grey collar → Not sure about these types of customers ← moderate credit profile, low credit profile
3. Black collar → clearly not capable of paying back the loan → credit profile very bad, Previously defaulted

problem : Loan tap wants to cater to Grey collar customers but the problem is they want to be very sure if a customer will pay back the loan or not?

Grey collar ← capable of paying back (80%)
Grey collar ← not at all capable of paying back the loan (20%)

(quick / fast / efficient)

Data → ML-model → (probability value) → Loan approved / Rejected

$P(y = \text{Default} \mid \text{profile of the customer})$

→ fill the application → Loan approved / Rejected } overall loan approval time ≤ 45 mins

E-KYC (30 mins almost)