

Capstone Project

Background and Objectives

CREDIT SCORING

Background

1. Background

- This data aims to analyze and mitigate credit default risks by examining various demographic and financial factors.
- Leveraging customer information such as age, gender, marital status, and financial indicators, the project seeks to understand the key drivers behind loan defaults.
- The use of credit scoring models is a standard practice in the UK banking and financial industry. These models assist lenders in making informed decisions regarding loan approvals, interest rates, credit limits, and other aspects related to lending while managing risks associated with extending credit.



Background

2. Primary Objectives

1. Analyze defaulters and understand the factors associated with it
2. Develop credit scorecard using statistical modeling
3. Compare scorecard using ML methods
4. Develop web application for estimating probability of default
5. Analyze customer feedback (NPS Survey)



Background

3. Data

The following datasets are available:

1. Customer Credit Details
2. Customer Info
3. Customer Profile
4. Customer feedback in text format



Data : Customer Credit Details

Content

The dataset includes customer information such as unique identifiers, debt-related metrics like debt-to-income ratio, credit card debt, other debt, and indicates whether customers are applying for a preloan.

custid	debtinc	creddebt	othdebt	preloan
1	12.34	13.26	5.88	2
2	18.65	2.12	5.13	1
3	7.22	3.31	3.65	1
4	6.15	2.95	2.34	1
5	20.64	2.67	4.07	2
6	12.44	3.06	2.57	2
7	34.21	6.51	17.62	2
8	8.27	2.38	3.09	1
9	26.67	3.41	4.22	2
10	19.81	2.81	4.06	2

Columns	Description	Type	Measurement	Possible values
custid	Customer Id	numeric	-	-
debtinc	Loan to Income Ratio	numeric	-	-
creddebt	Balance CC Debt	numeric	-	-
othdebt	Any other Debt	numeric	-	-
preloan	Any Pre Loan	Factor	-	2: Not the first loan

Data : Customer Info

Content

This data presents customer attributes, such as vehicle ownership, housing status, self-employment, account type, deposit status, employment duration, address duration, and details about the branch and reference

custid	veh	house	selfemp	account	deposit	emp	address	branch	ref
1	2	1	2	1	2	18	12	1	2
2	2	1	1	2	1	11	7	1	2
3	1	2	1	1	1	16	15	1	1
4	1	2	1	1	2	15	14	2	1
5	2	1	1	1	1	2	1	2	2

Columns	Description	Type	Possible values
custid	Customer Id	numeric	
veh	Owns Vehicle	Factor	2: Owns Vehicle
house	Owns house	Factor	2: Owns house
selfemp	Is customer self employed	Factor	2: Self-Employed
account	Customer having acc in the same bank	Factor	2: Acc in same bank (CASA)
deposit	Deposit with the same bank	Factor	2: Deposit with the same bank
emp	Years with current employer	numeric	
address	Years at current address	numeric	-
branch	Nearest Brank	Factor	2: nearest branch
ref	Any reference	Factor	2: Internal Reference



Data : Customer Profile

Content

The dataset includes customer's demographics

custid	age	gender	ms	child	zone	bad
1	2	1	2	2	7	1
2	2	2	1	1	5	0
3	1	1	2	1	15	0
4	2	1	1	1	3	0
5	2	2	1	1	20	0
6	1	2	1	1	5	0
7	2	1	2	1	3	0
8	1	2	1	1	1	0
9	2	1	2	1	8	0
10	2	1	2	1	17	0

Columns	Description	Type	Possible values
custid	Customer Id	numeric	
age	Age	Factor	1: <32, 2: 32-48, 3: >48
gender	Gender	Factor	2: Female
ms	Marital Status	Factor	2: married
child	No. of children	Factor	1: Zero or one, 2: More than 1
zone	Area code	numeric	
bad	Good/bad	Factor	0: Good, 1: Bad

Data :Customer Feedback

Content

The provided text file contains customer feedback related to personal loan experiences with a bank. Customers express varying sentiments, highlighting positive as well as negative aspects.

My experience of availing the personal loan was smooth.
I had a satisfactory experience with the Bank. The loan officer was helpful, and the approval process was quick. However, their interest rates were a bit higher than what I expected.
I had an excellent experience with the Bank while applying for a loan. The loan officer was efficient, and I got my loan approved within a few days.
I had to move on to another bank as it was taking a long time
I recently applied for a loan at the Bank, and the service was impeccable. They guided me through the entire process and explained all the terms clearly.
I got my loan approved within a few days, and the Interest rate was reasonable. Highly recommended!
I had a smooth experience with the Bank while availing a loan. The process was quick, and the staff was knowledgeable.
The loan officer was so helpful and provided me with various loan options. They were patient in answering all my questions.
It was surely disappointing experience.
I guess it was like any other good bank.Nothing special.
The willingness to explain loan products was missing.
It took weeks to get my loan approved, and their interest rates were too high. I won't be going back there again.
Bank's loan service was alright. The application process was straightforward, but it took a bit longer than I expected to get the approval
It was ok
It was quite good



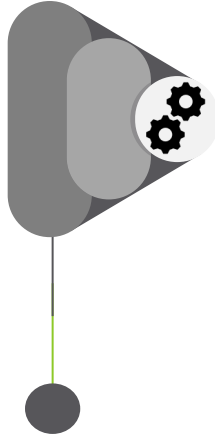
Next steps

Data management



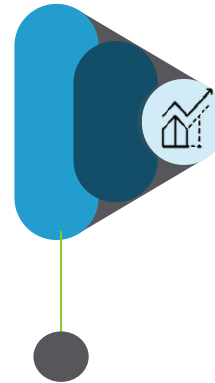
- Compile all 4 data files based on Customer ID
- Data cleaning , Handling missing values and completing Basic Data checks
- Check if any variables needed to be feature coded i.e made into groups or want to be left as continuous variables

Descriptive Statistics & Data visualization



- Understanding the data better like default rate.
- How can this data be presented better visually ?
- Once again post Data visualization check if any variable needs to be feature coded

Predictive modelling



- Answer the objectives a) To understand factors to pre-empt which employees who are likely to be a defaulter b) Parameter which leads to customer default
- Using different Predictive model techniques to find Significant variables
- Ensure you follow all steps like Train and test data , checking for Multicollinearity
- Check if any other ML technique fits better

Text mining



- Understand the customer feedback using various Text mining Techniques
- Answer the objective: To analyze customer feedback.

