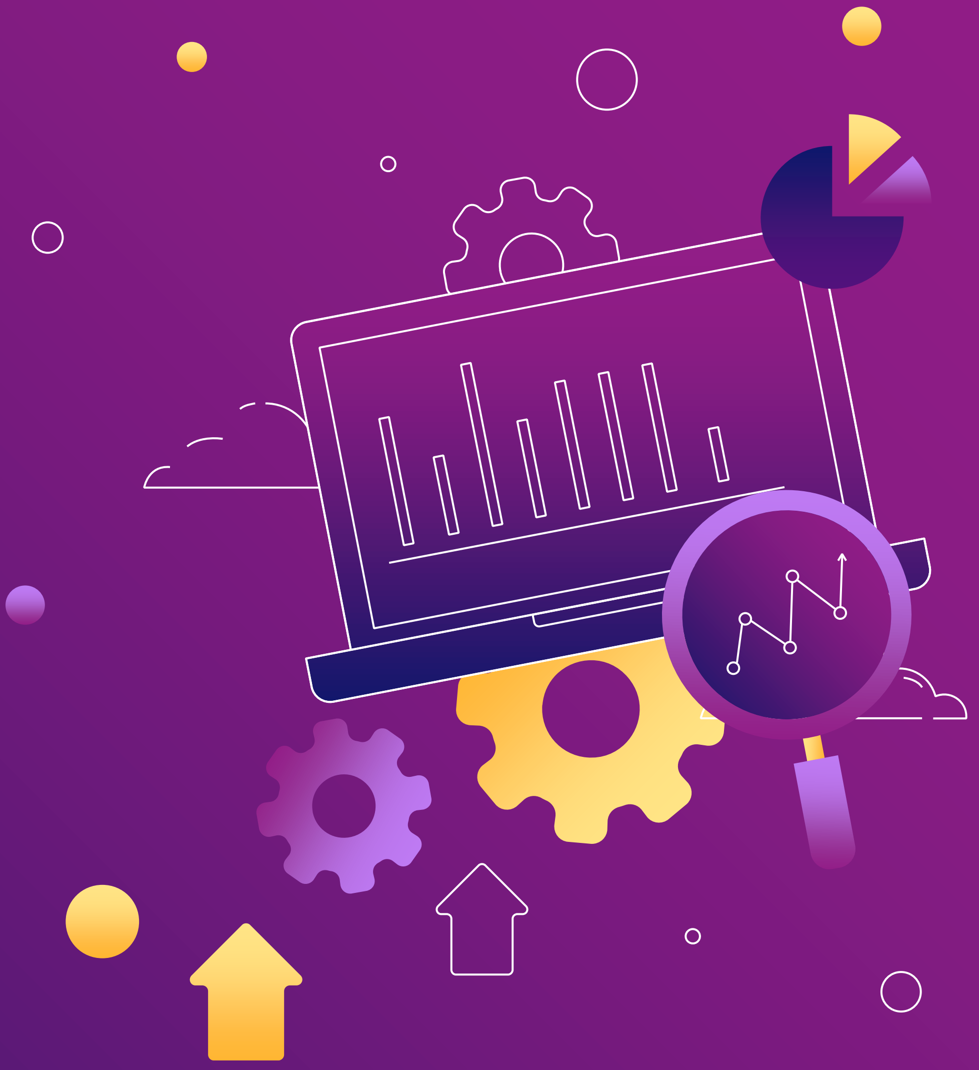
CREDIT SCORE FORECASTING

Malak Diab, Mohammed Hourani, Mayas Masalmeh

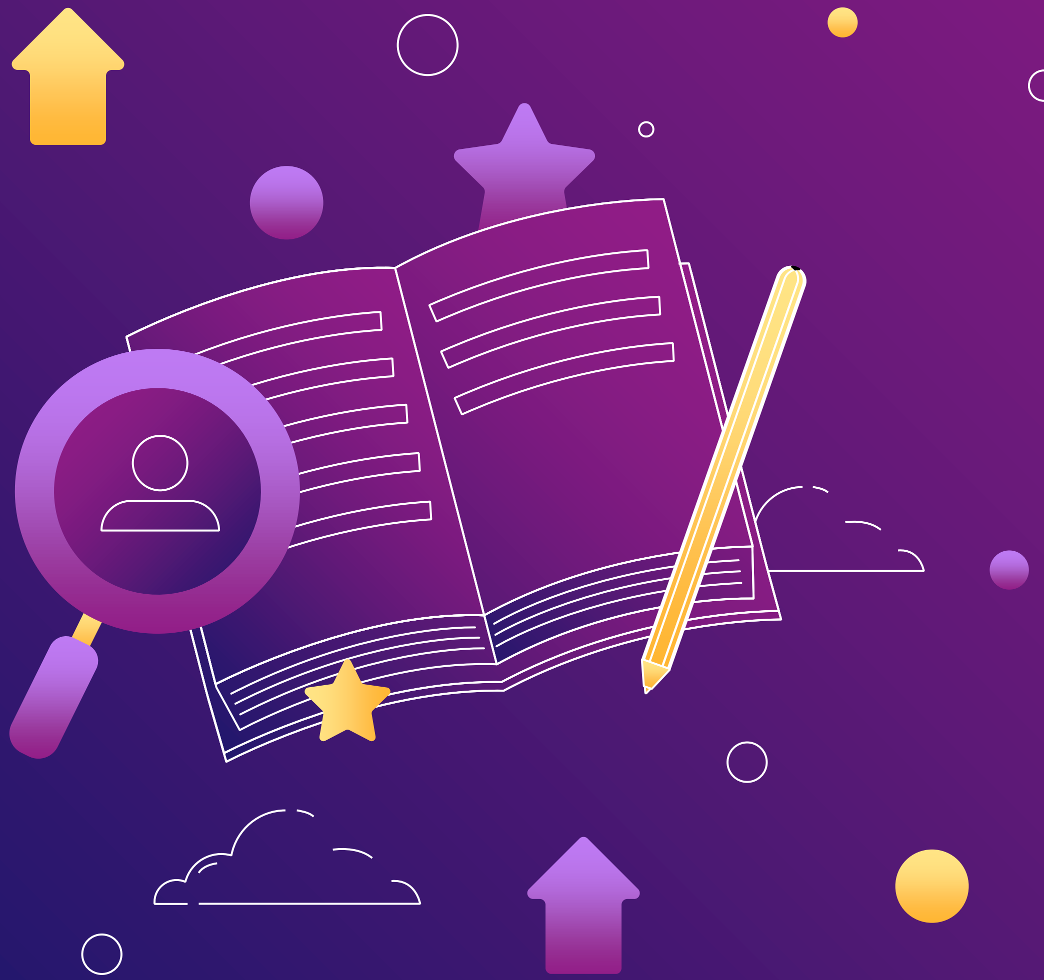


DATA SCIENCE BOOTCAMP GRADUATION PROJECT

TABLE OF CONTENT

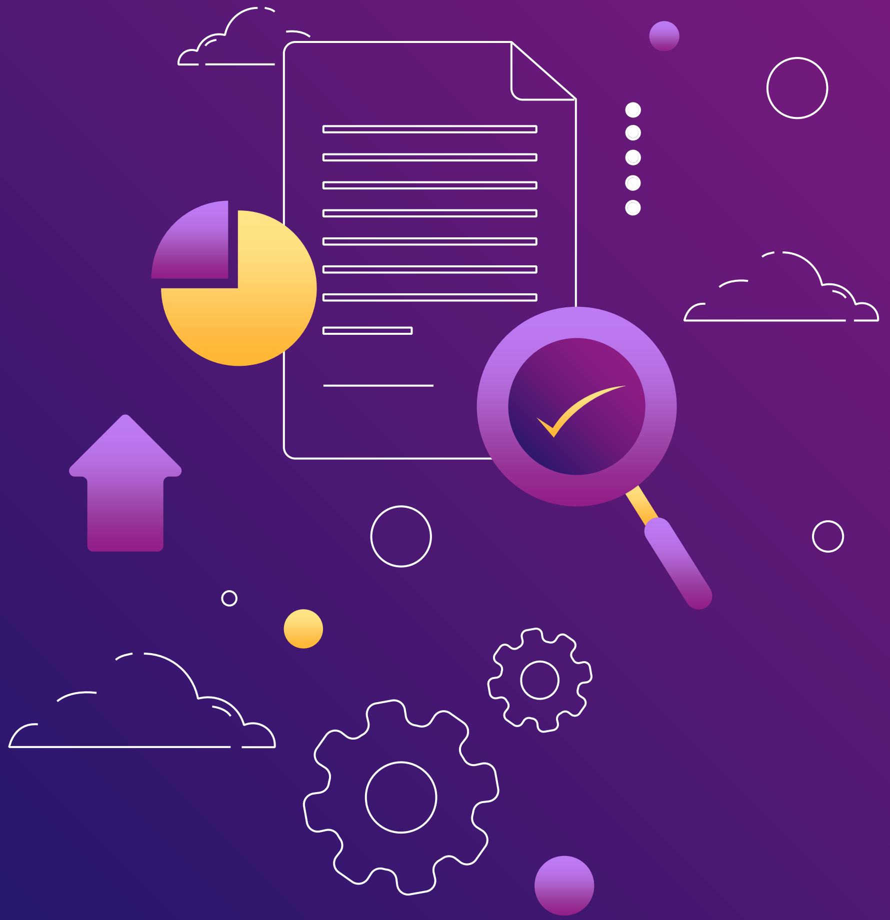
01 DATA INTRODUCTION

Problem Statement and the dataset



We’re working as a data scientists. Our dataset includes basic bank details with a lot of credit-related information. We’re looking forward building an intelligent system that predict customers credit score based on the main features that effects it.

PROBLEM STATEMENT



CREDIT SCORE (TARGET)

Represents the bracket of credit score (Poor, Standard, Good). It’s used to help the lender to decide either to lend the customer or not based on multiple features that affects the credit score.

CREDIT SCORE DISTRIBUTION

53%
Standard Credit Score

17%
Good Credit Score

28%
Poor Credit Score

02 APPROACH OVERVIEW

Model Steps & Algorithms

MODEL STEPS:

01 READING THE DATA
Reading the Data After the Analysis.

02 ENCODING
Get Dummmies & Ordirnal Encoder.

03 DROPPING & FEATURE SELECTION
Removing Irrelevant Columns and Selecting the Most Important Feature for the Model.

04 SPLITTING
Splitting the Data to Train and Test.

05 STANDARDIZATION
Re-scaling the Data.

06 MODEL BUILDING
Training the Model & Fine Tunning Hyperparameter

MODEL STEPS:

01 READING THE DATA
Reading the Data After the Analysis.

02 ENCODING
Get Dummmies & Ordirnal Encoder.

03 DROPPING & FEATURE SELECTION
Removing Irrelevant Columns and Selecting the Most Important Feature for the Model.

04 SPLITTING
Splitting the Data to Train and Test.

05 STANDARDIZATION
Re-scaling the Data.

06 MODEL BUILDING
Training the Model & Fine Tunning Hyperparameter

IN 8 MONTHS

12,500

Number of Customers

100,000

Number of Rows

28 ➡ 43

Number of Columns

01

COLUMNS DIGGER

TYPE OF LOAN
10 Columns

DELAY BY CUSTUMER MEAN
1 Column

MONTHLY BALANCE MEAN
1 Columns

NUMARIC -> CATEGORICAL
3 Column

02 ORDINAL ENCODING

CREDIT SCORE	MAPPINGS
Good	2
Standard	1
Poor	0

PAYMENT OF MIN AMOUNT	MAPPINGS
No	2
NM	1
Yes	0

CREDIT MIX	MAPPINGS
Good	2
Standard	1
Bad	0

GET DUMMIES ENCODING

OCCUPATION	ACCOUNTANT
Customer-1	1
Customer-2	1
Customer-3	0

PEYMENT BEHAVIOR	SMALL SPENT SMALL VALUE PAYMENT
Customer-1	0
Customer-2	1
Customer-3	0

03 COLUMN DROPPING

CUSTOMER ID

NAME

SSN

MONTHLY IN HAND SALARY

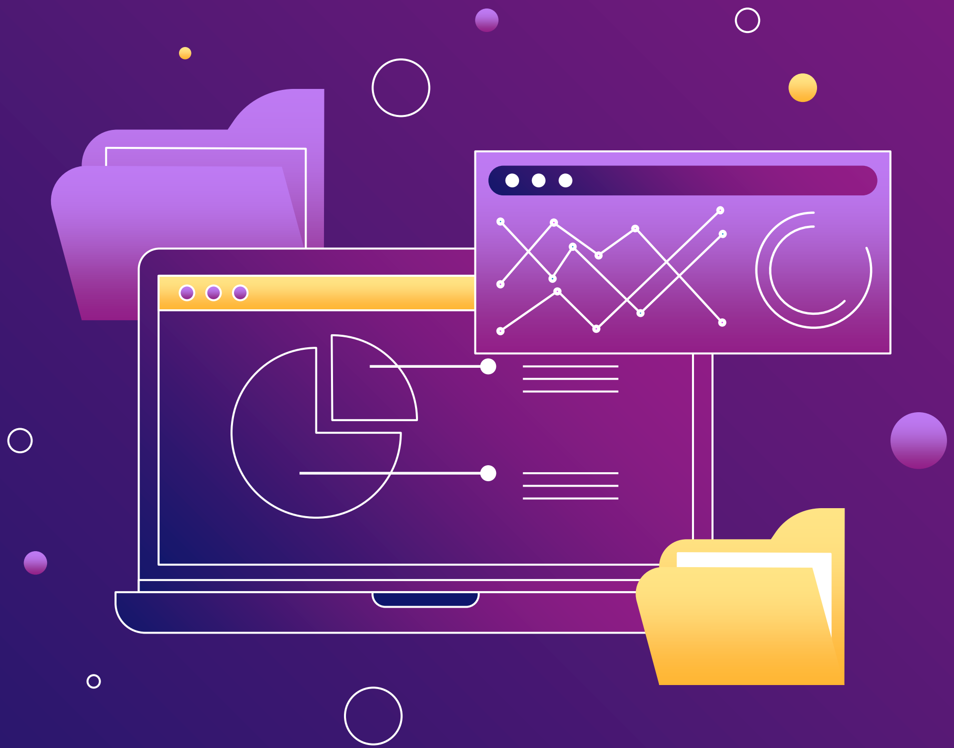
TYPE OF LOAN

ANNUAL CATEGORY

HISTORY AGE CATEGORY

AGE CATEGORY

FEATURE SELECTION:



Before feature selection, we've chosen one month for our model.

After encoding for categorical features, we ended up with 49 columns and used 16 of them.

We used **SelectKBest** with **Mutual Information.**

Mutual information is a measure between two (possibly multi-dimensional) random variables X and Y, that quantifies the amount of information obtained about one random variable.

04 DATA SPLITTING

Training	80%
Testing	20%

Stratify Split to Ensure the Same Distribution of Classes in Training and Testing Data.

ALGORITHMS

XGBOOST CLASSIFIER

First: Accuracy -> 77%
Then: Accuracy -> 74%

VS

GRADIENT BOOST

Accuracy -> 77%

03 CONCLUSION & RESULTS

Accuracy Before and After Fine Tunning

ACCURACIES

- RESULTS
  - TEST ACCURACY
    - BEFORE FINE TUNNING 77%
    - AFTER FINE TUNNING 76%
  - TRAIN ACCURACY
    - BEFORE FINE TUNNING 79%
    - AFTER FINE TUNNING 82%

04 MODEL DEPLOYMENT

User Interface for the Credit Score Model.

THANKS!

DO YOU HAVE ANY QUESTIONS?

Made with Lots of Love by the Most
Creative Clever People Ever.

MALAK DIAB

MOHAMMED HOURANI

MAYAS MASALMEH