

CREDIT RISK ANALYSIS

NESMA DEHILI

DAFT NOV 2021 - 11 FEB 2022

WHAT IS CREDIT ANALYSIS AND WHAT IS ITS PURPOSE?

The main purpose of credit risk analysis is to quantify the level of credit risk that the borrower presents to the lender.

Credit analysis is the process of determining the ability of a company or person to repay their debt obligations.

In other words, it is a process that determines a potential borrower's default risk. It incorporates both qualitative and quantitative factors.

DATASET

BANK INFORMATION ABOUT
CLIENTS

19 Columns

110867 Rows



CLIENT RELEVANT INFO

- Customer ID,
- Credit Score,
- Annual Income,
- Years in current job,
- Home Ownership, Purpose,
- Years of Credit History,
- Months since the last delinquent,
- Number of Open Accounts,
- Number of Credit Problems,
- Maximum Open Credit, Bankruptcies, Tax Liens



LOAN DETAILS

- Loan ID
- Loan Status
- Current Loan Amount
- Term
- Monthly Debt
- Current Credit Balance

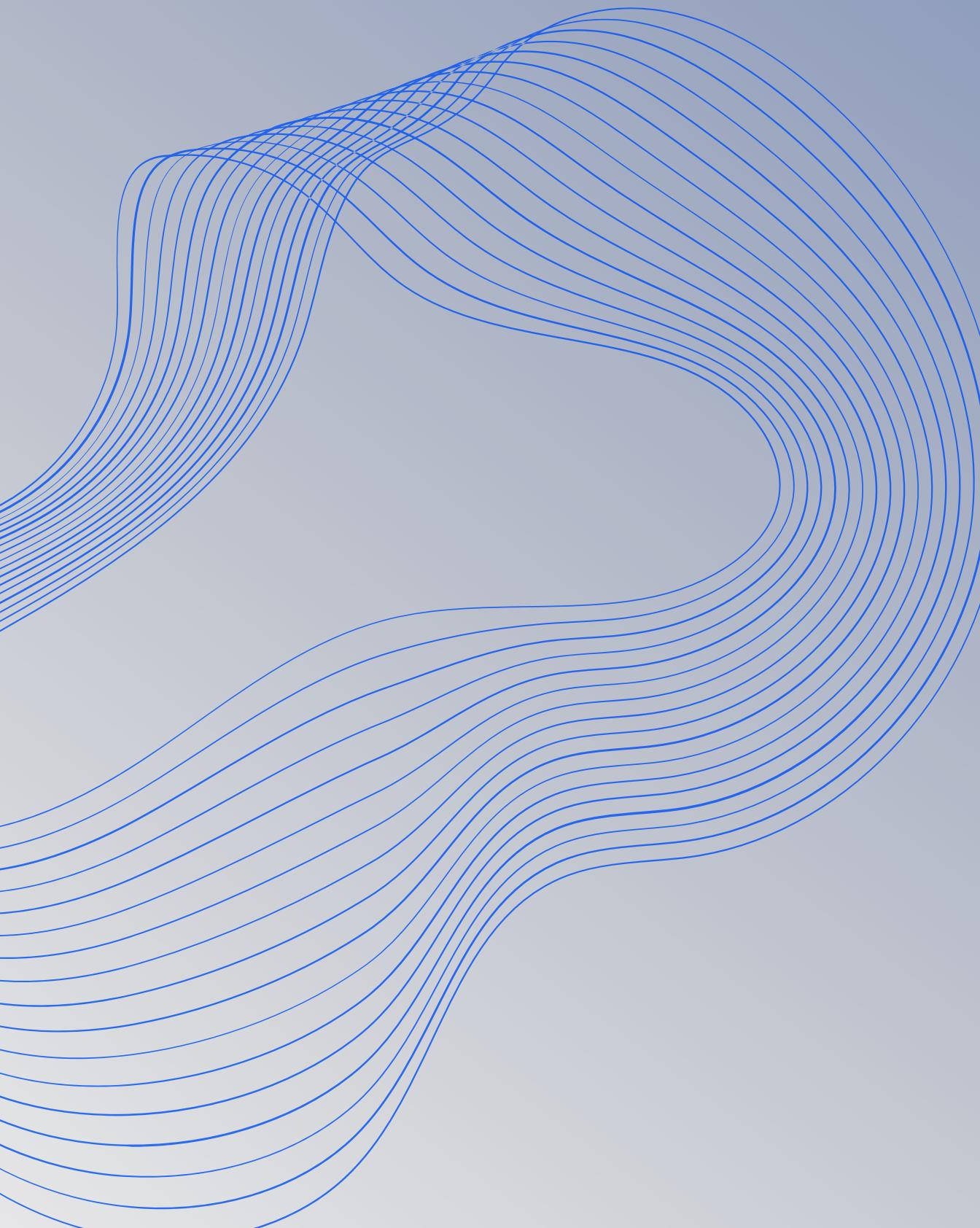
PLANNING

DATA CLEANING

MODELING

STRATEGISING





DATA CLEANING



- Overall Data description
- Deleting 11081 duplicates
- Looking for outliers column by column
- Dropping columns with more than 50% nan values
- Dropping NaN values from the target column
- Interpolating empty cells



Github link:

<https://github.com/Nesmad/Credit-Risk-Analysis/blob/main/Credit%20Risk%20Analysis%20.twb>

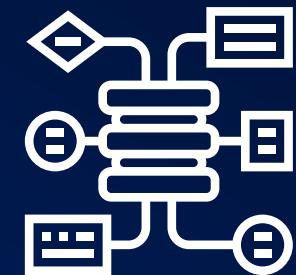
MODELS USED

 LOGISTIC
REGRESSION

 GRADIENT
BOOSTING
CLASSIFIER

 RANDOM FOREST
CLASSIFIER

 TPOT (RFC)



PIPELINE FOR MODELING

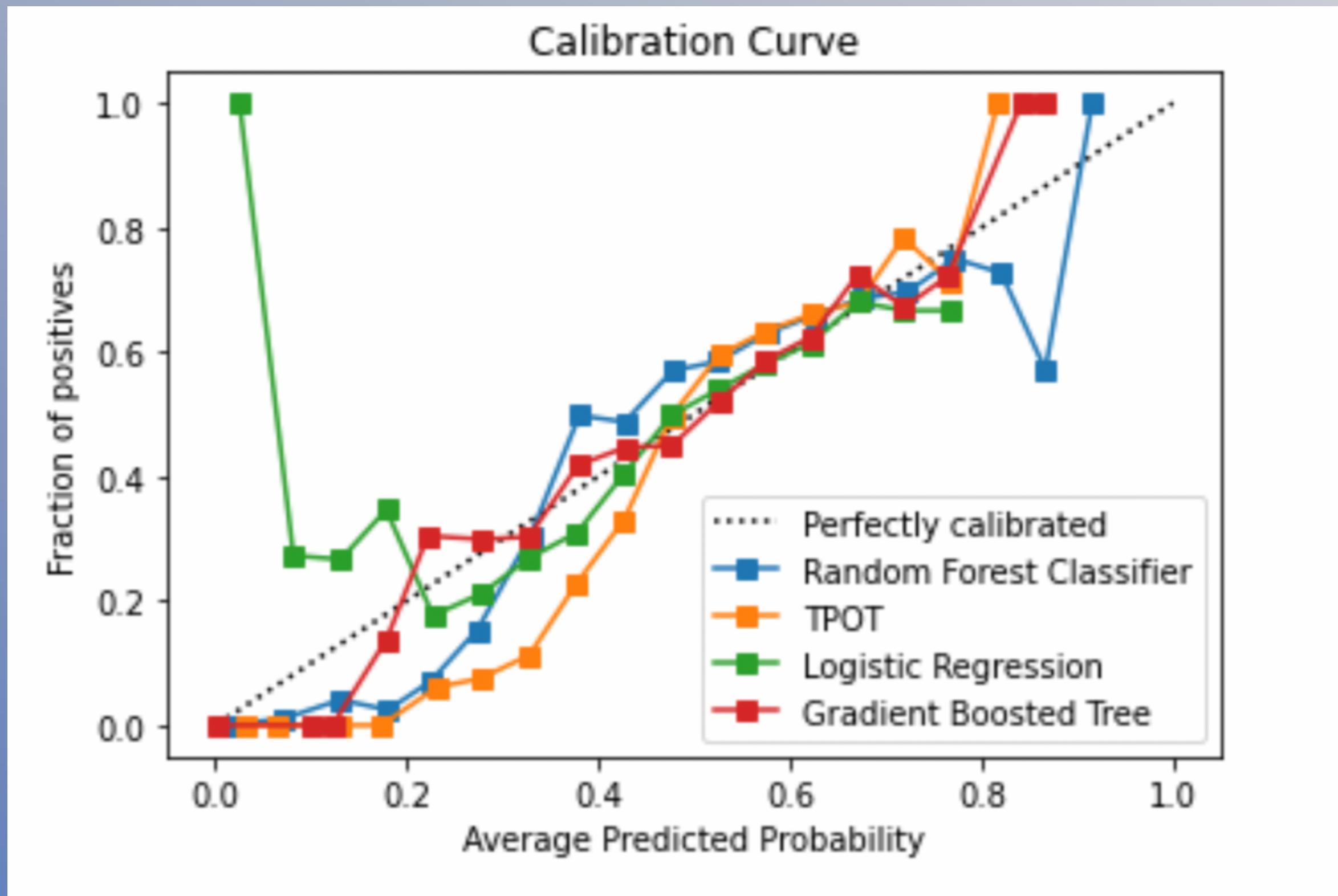
- Encoding
- Feature Selection
- Hyperparameter tuning
- Models
- Compared accuracy
- AUC-ROC Curve
- Checking Calibration



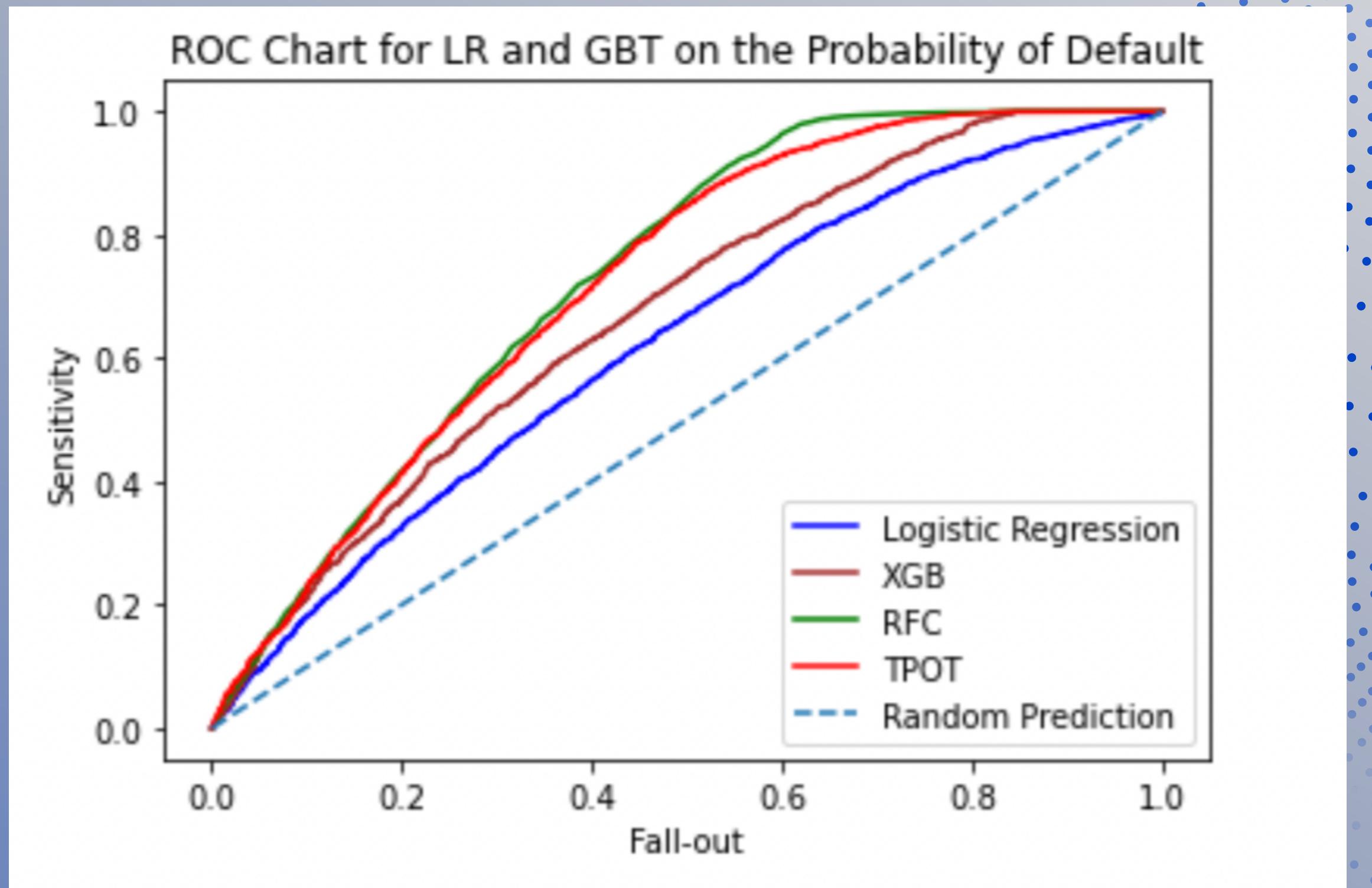
MODELS COMPARISON

Models	Accuracy	AUC
Logistic Regression	58,49%	62,13%
Gardian boosted tree	61,91%	67,00%
Random Forest	66,72%	73,05%
TPOT	67,02%	72,02%

CALIBRATION CURVE



AUC MODELS COMPARISON





Business Goal:
**MAXIMIZING
VALUE**

Banks need to define a strategy by defining the percentage of loans granted with the goal of keeping defaults below a certain rate

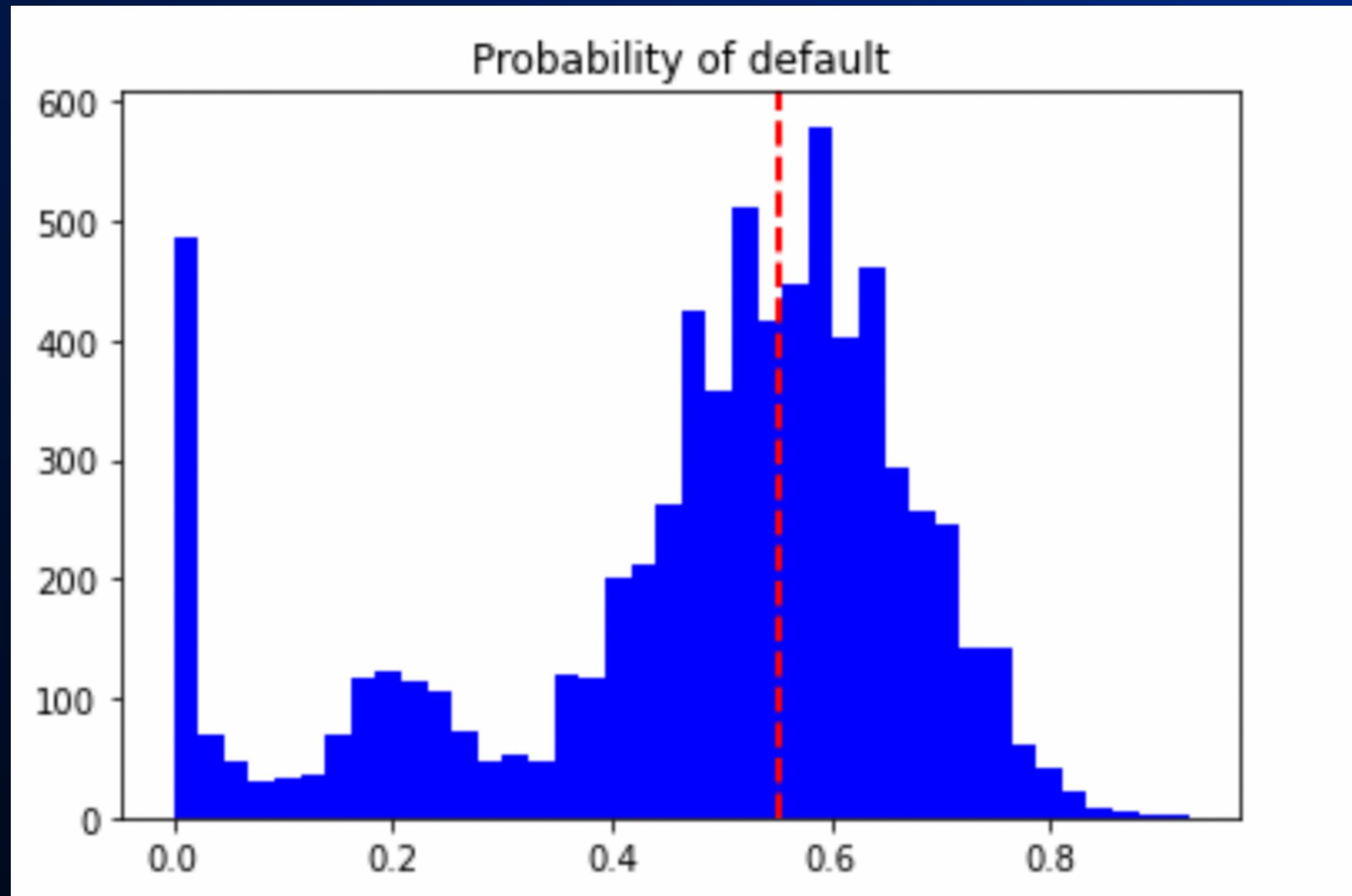


STRATEGISING

- Defining acceptance rate
- Defining limit probability
- Assessing error rate
- Calculating accepted loans
- Estimating the portfolio value
- Finding the BEST combination

Acceptance Rate	Thresholds	Error rate	Num of Accepted Loans	Avg Loan Amount	Estimated Value
1,000	0,916	0,502	7168	295295,56	-9152885,452
0,950	0,721	0,502	6793	295295,56	-8674044,486
0,900	0,684	0,490	6423	295295,56	38251968,697
0,850	0,658	0,502	6049	295295,56	-7724024,010
0,800	0,637	0,490	5720	295295,56	34065274,942
0,750	0,616	0,477	5332	295295,56	71825185,485
0,700	0,600	0,502	4990	295295,56	-6371777,122
0,650	0,584	0,490	4638	295295,56	27621458,947
0,600	0,568	0,477	4280	295295,56	57654124,883
0,550	0,553	0,469	3936	295295,56	72583484,955
0,500	0,532	0,502	3545	295295,56	-4526643,266
0,450	0,516	0,490	3209	295295,56	19111095,680
0,400	0,495	0,477	2799	295295,56	37704181,203
0,350	0,474	0,469	2467	295295,56	45493764,579
→ 0,300	0,447	0,456	2113	295295,56	54861875,084
0,250	0,411	0,502	1769	295295,56	-2258852,450
0,200	0,342	0,490	1427	295295,56	8498452,332
0,150	0,226	0,477	1070	295295,56	14413531,221
0,100	0,147	0,469	705	295295,56	13000852,869
0,050	0,016	0,456	337	295295,56	8749858,922

RESULTS EXPLAINED



RECOMMENDATION

Acceptance Rate	Thresholds	Error rate	Num of Accepted Loans	Avg Loan Amount	Estimated Value
55,00%	55,30%	46,88%	3936	\$295,295,56	\$72,583,484

Predicted probability



ESTIMATED VALUE ON THE PORTFOLIO

We need to keep in mind that this value could be even maximized if the data provided more insights on both the customer, such as age and field of work and so on, or even vital information about the interest rate and length of the credit.

ANALYSING CREDIT RISK FOR BANKS

Questions?

Nesma Dehili