CREDIT RISK ANALYSIS

Presented by:
Mohamed islem Beggari

INTRODUCTION:

What is Credit Analysis?

Credit analysis is the systematic process of evaluating the creditworthiness of individuals, businesses, or entities seeking to borrow funds from financial institutions or lenders

The process of credit analysis involves

What is credit risk analysis ?

Credit risk analysis is the process of evaluating the potential risk of a borrower defaulting on their financial obligations. It involves assessing various factors, including the borrower's financial history, creditworthiness, income, collateral, and economic conditions,

Scoring Technics ?

A credit score is a 3-digit number that reflects the likelihood that a consumer will repay his debts.
This score is a numerical representation of the borrower's creditworthiness and risk profile.

DATA OVERVIEW

32,581 entities

12 columns

3 columns of FLOAT

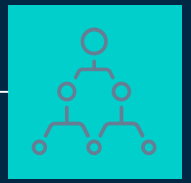
5 columns of INT

4 columns of OBJECT.

DATASET

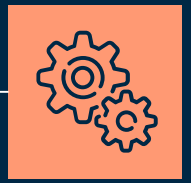
feature	description
person_age	The person's age in years
person_income	The person's annual income.
person_home_ownership	The type of home ownership (RENT, OWN, MORTGAGE, OTHER)
person_emp_length	the person's employment length in years.
loan_intent	the person's intent for the loan (PERSONAL, EDUCATION, MEDICAL, VENTURE, HOMEEMPROVEMENT, DEBTCONSOLIDATION).
loan_grade	the of risk on the loan(A,B,C,D,E,F,G)(A-> not riscky | G-> very riscky
loan_amnt	the loan amount.
loan_int_rate	the loan interest rate (between 6% and 21%)
loan_status	Shows wether the loan is currently in default with 1 being default and 0 being non-default.
loan_percent_income	The percentage of person's income dedicated for the mortgage.
cb_person_default_on_file	If the person has a default history (YES , NO).
cb_person_cred_hist_length	The person's credit history.

DATA CLEANING PLAN



01

Checking / removing duplicates

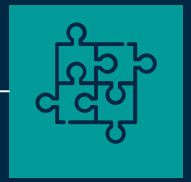


02

Feature Selection

03

Checking/dealing with Missing Data



04

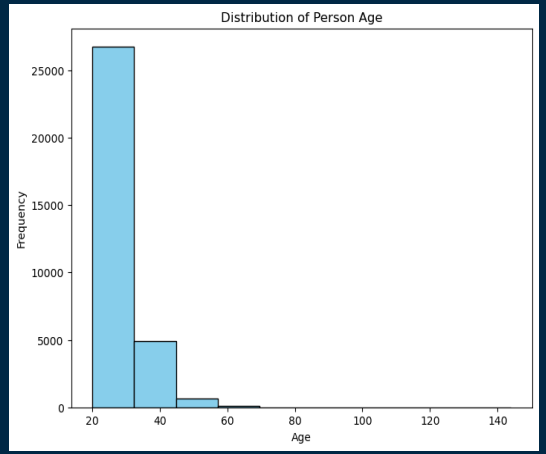
Detecting / Removing outliers

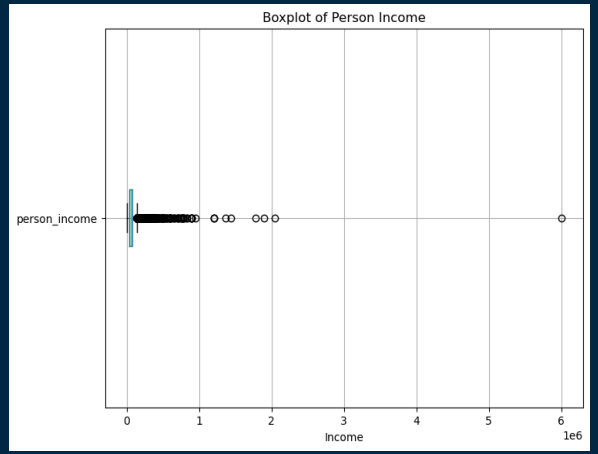
- Checking / Removing duplicates

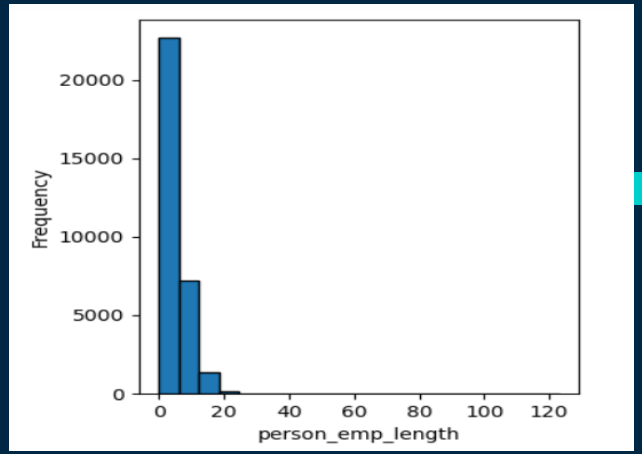
```
In [7]: ## Checking for Duplicates
        dups = df.duplicated()
        dups.value_counts()

Out[7]: False    32416
        True       165
        dtype: int64
```

- Feature Selection
- Detecting / Removing Outliers:

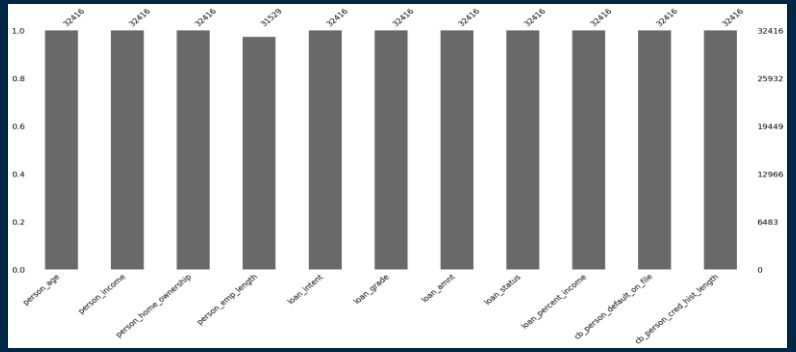






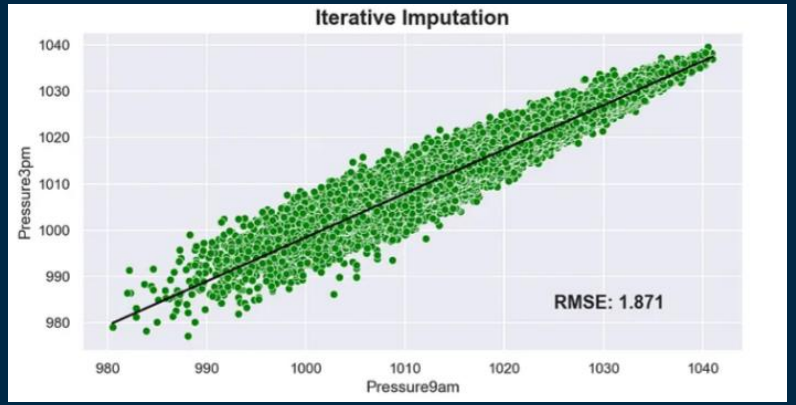
- # Checking for Missing DATA:

MSNO Library :



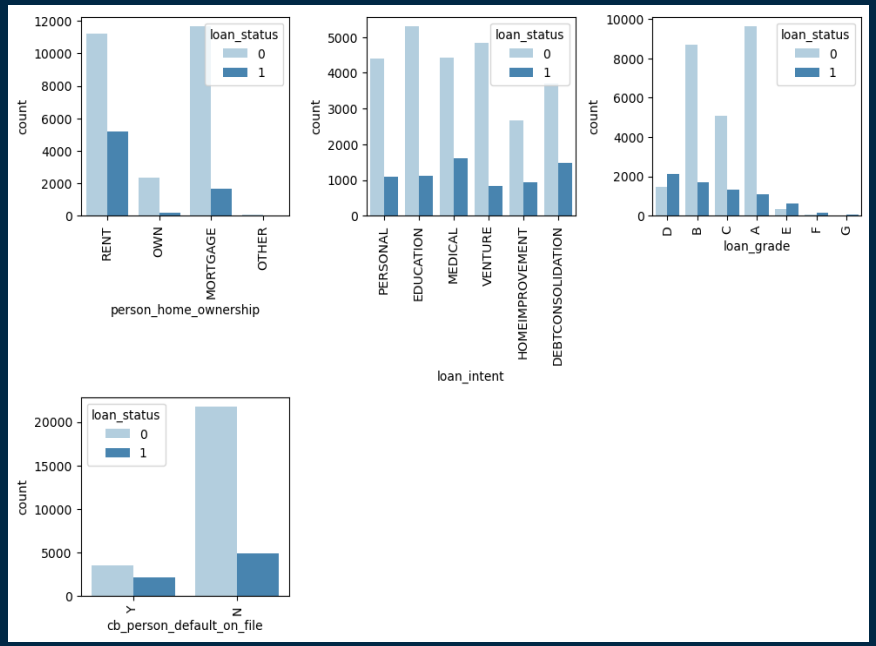
- # Dealing with Missing DATA:

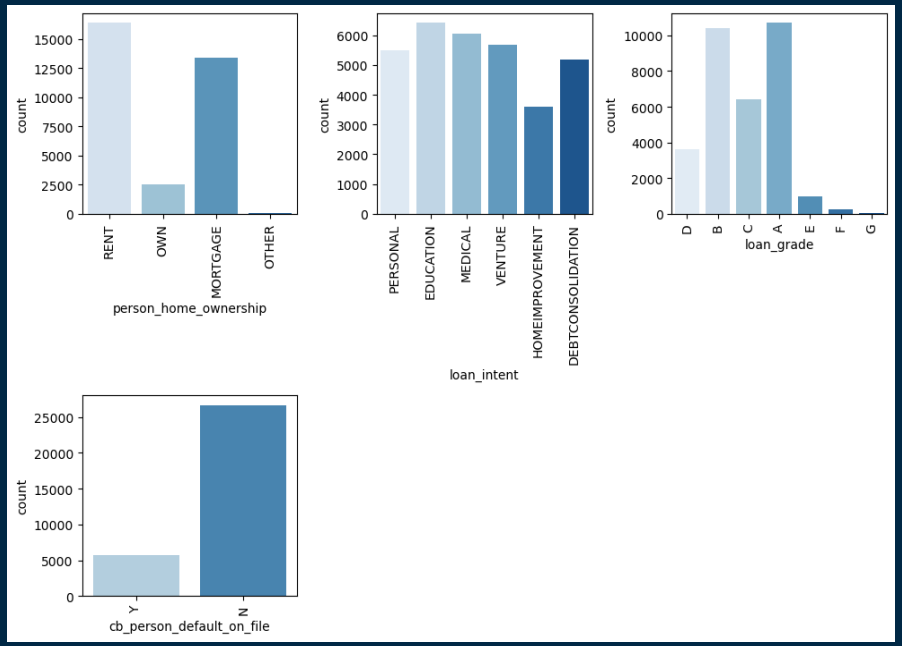
Iterative Imputer :



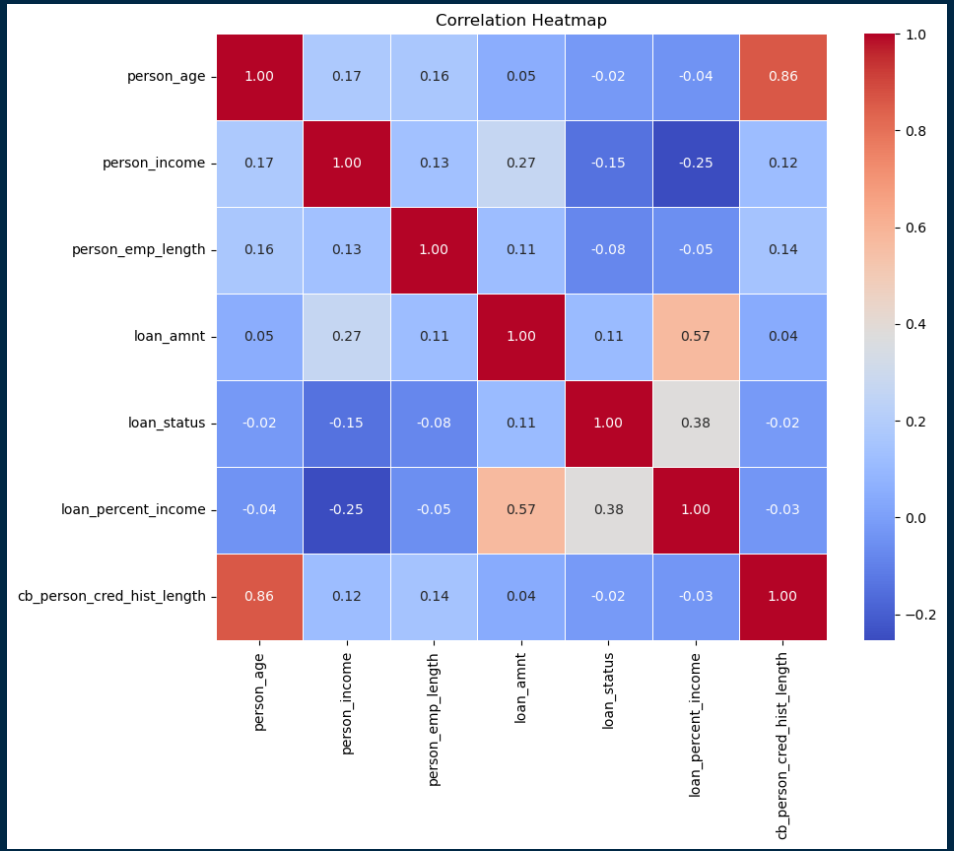
DATA VISUALIZATION & EXPLORATION

Categorical features

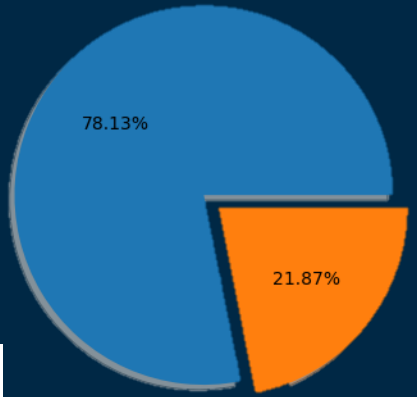


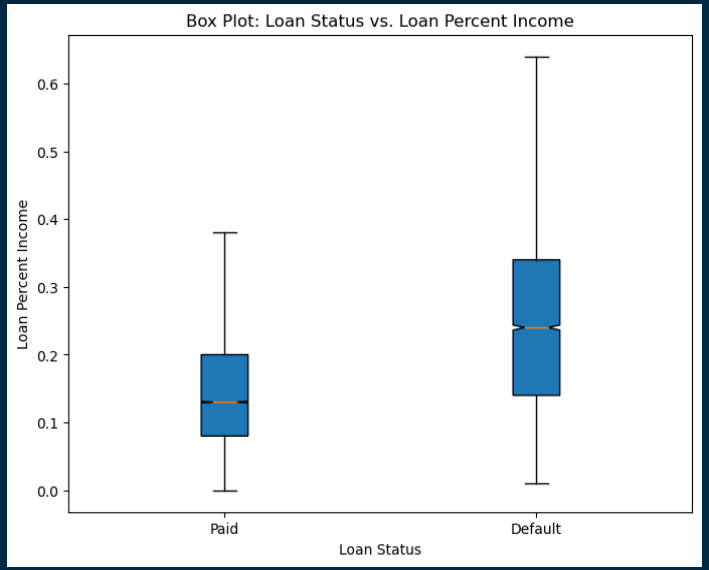


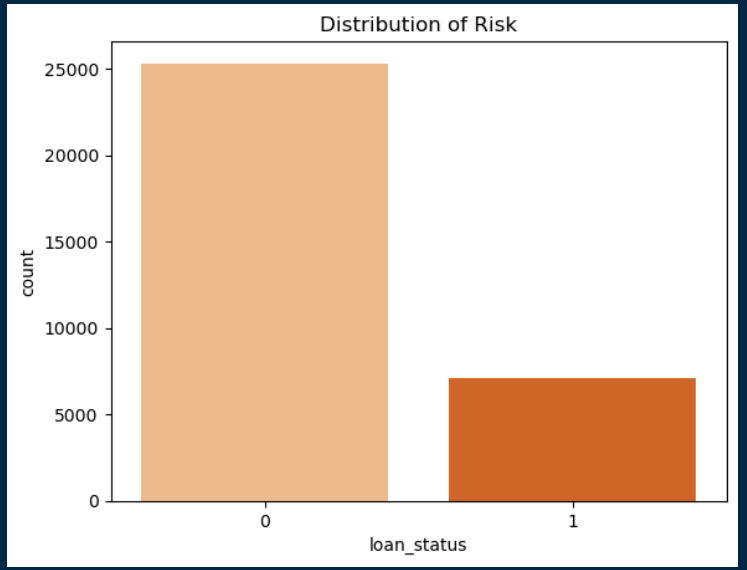
Numerical features



Target feature

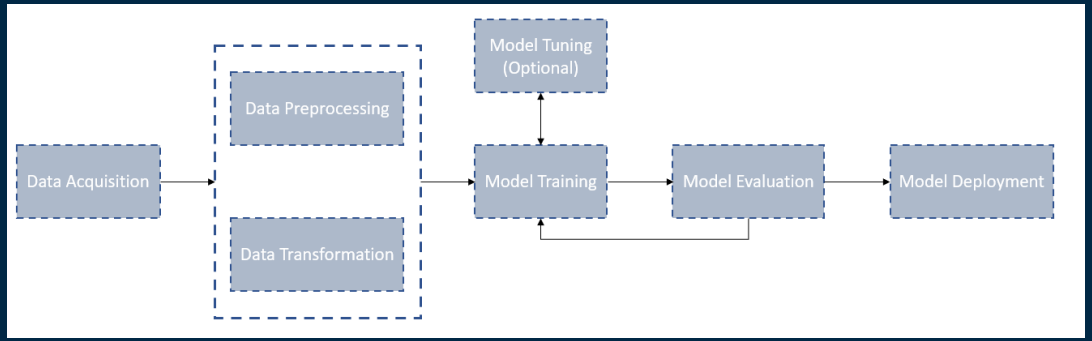






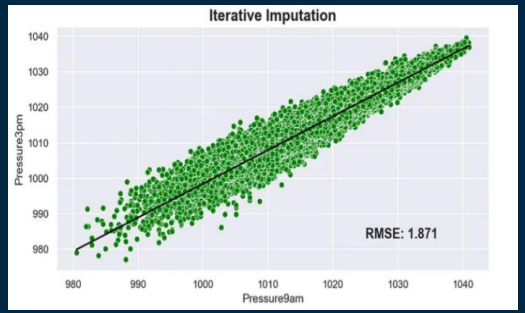
DATA PREPROCESSING :

Pipeline:

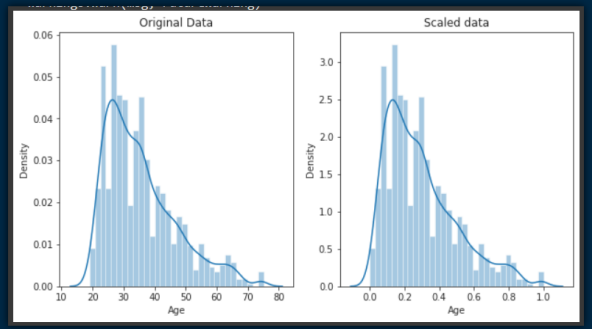


Preprocessing techniques

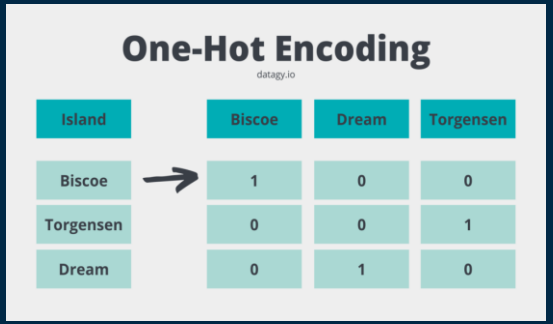
Iterative Imputer



Scaling

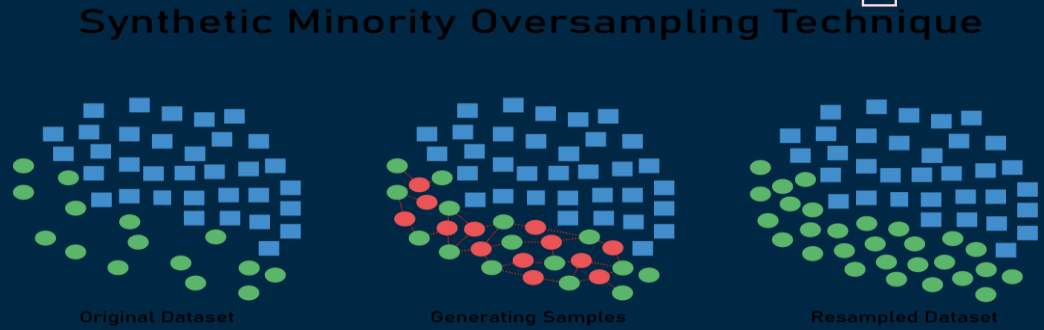


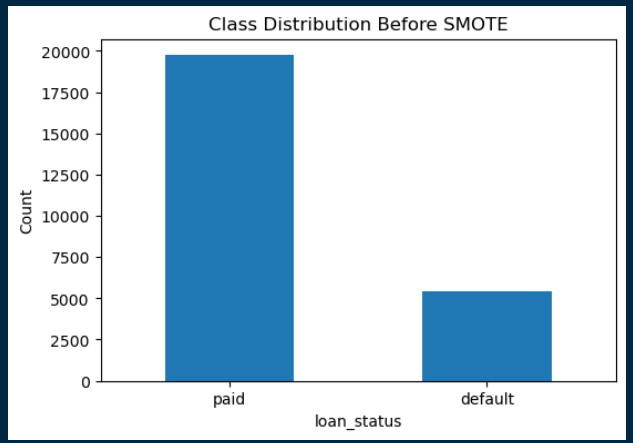
One-Hot Encoder



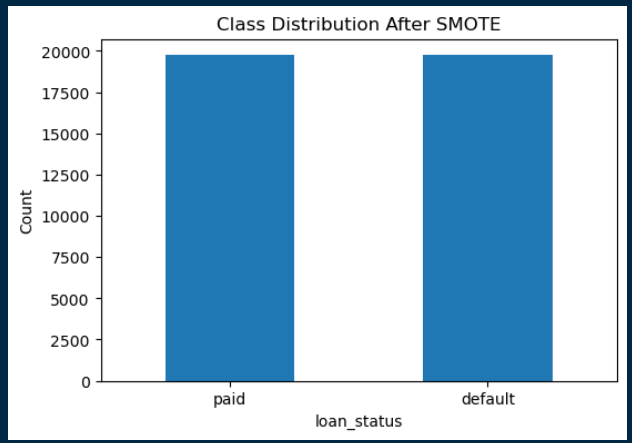
Handling Data Imbalance :

SMOTE is an oversampling technique designed to address the issues caused by data imbalance.
It generates synthetic samples for the minority class by creating new instances that are similar to existing minority class instances.





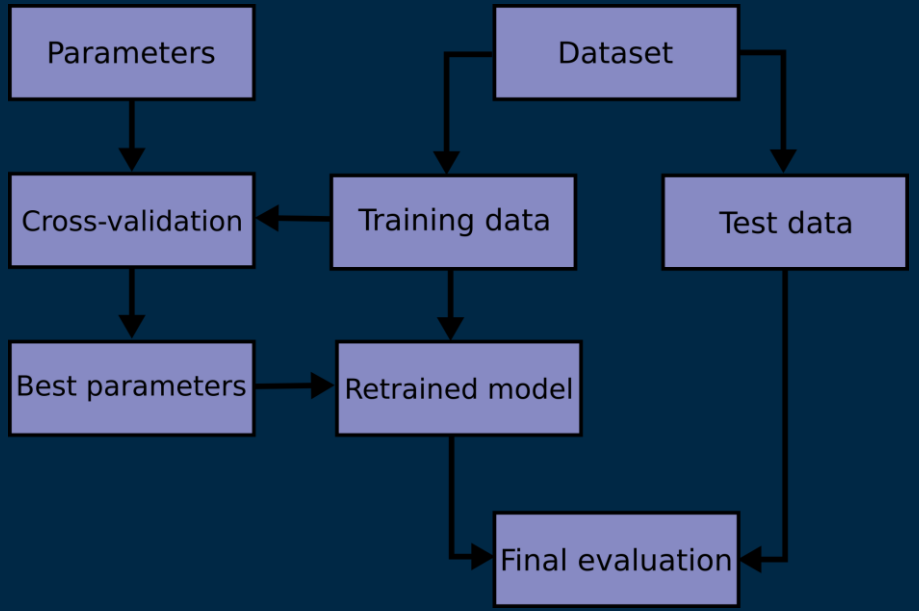
SMOTE



Model Selection and Hyper-parameter Tuning:

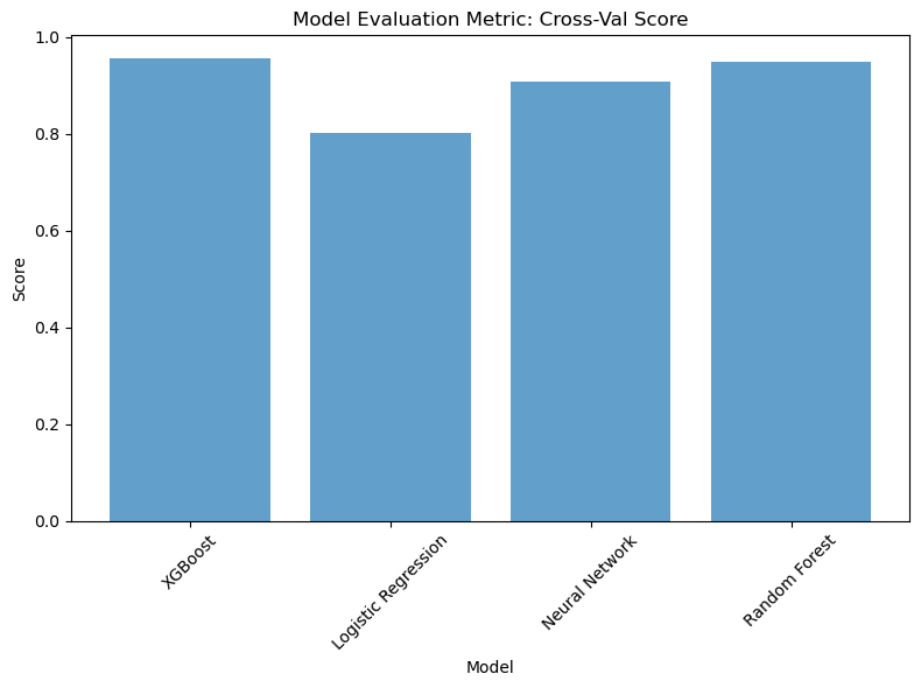
- Logistic Regression
- Random Forests
- Gradient Boosting Algorithms
- Support Vector Machines
- Neural Networks

Cross-Validation

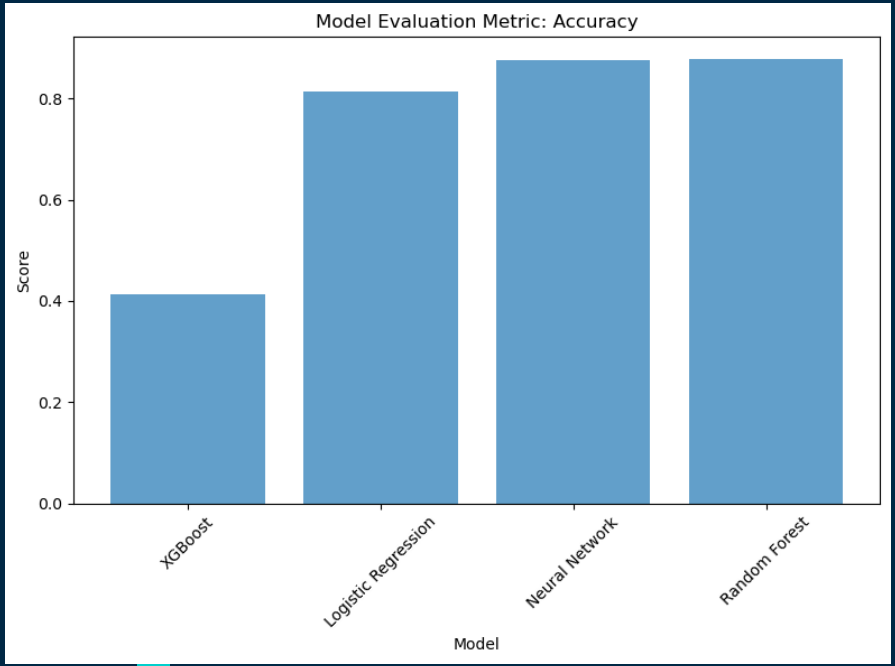


Model Evaluation

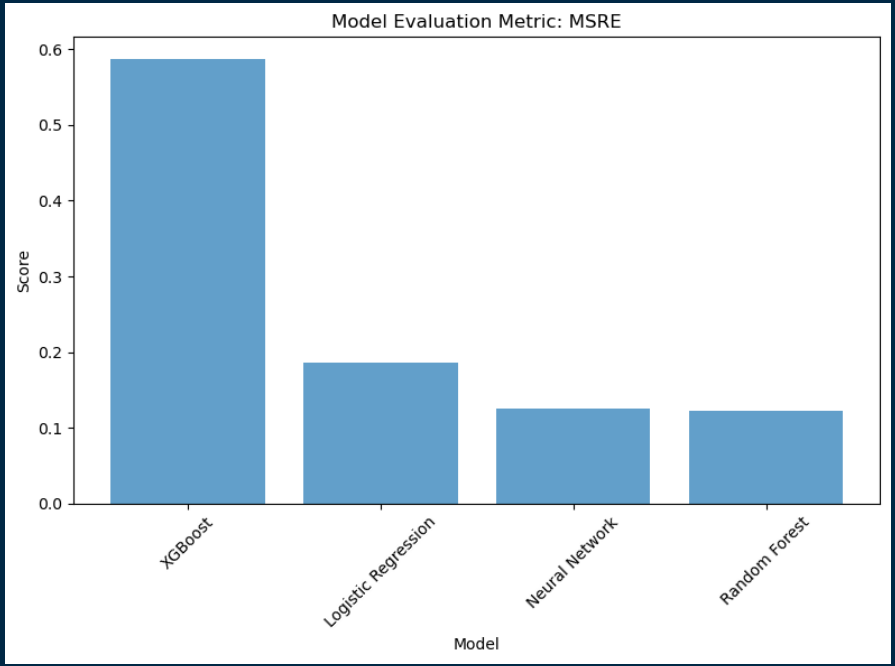
Cross-Validation Score :



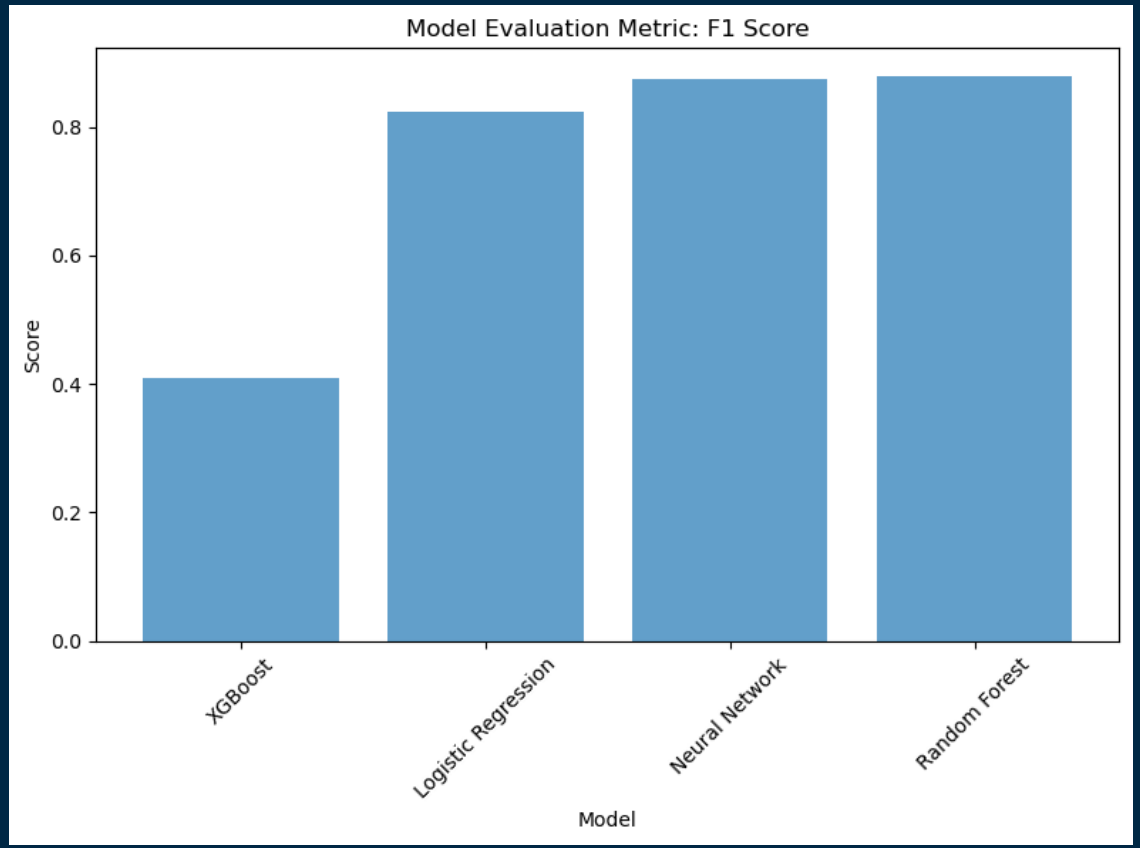
Accuracy



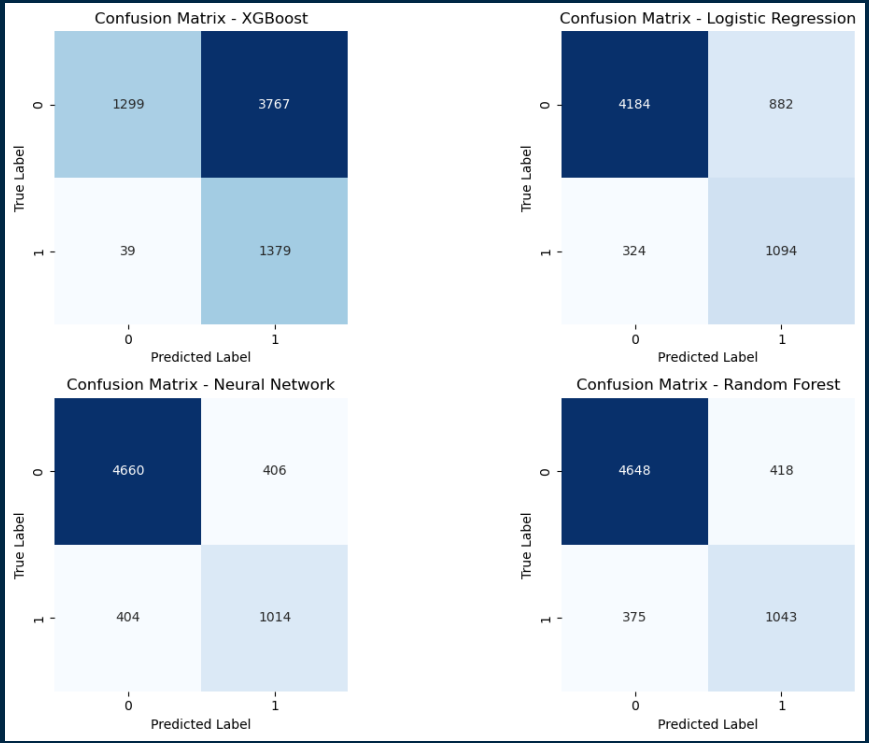
MSRE



F1-SCORE


Model Evaluation Metric: F1 Score
Score
0.0
0.2
0.4
0.6
0.8
XGBoost
Logistic Regression
Neural Network
Random Forest
Model

The confusion matrix



True Positive (TP): The model correctly predicted a positive class instance as positive.

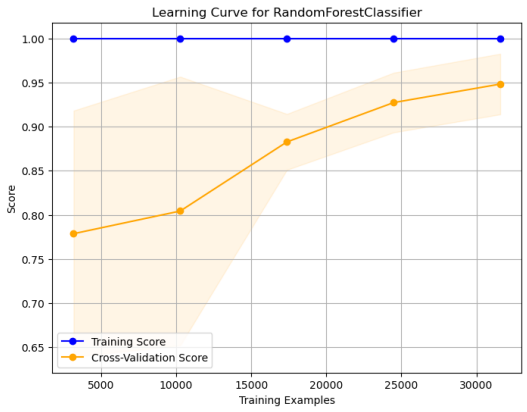
False Positive (FP): The model incorrectly predicted a negative class instance as positive.

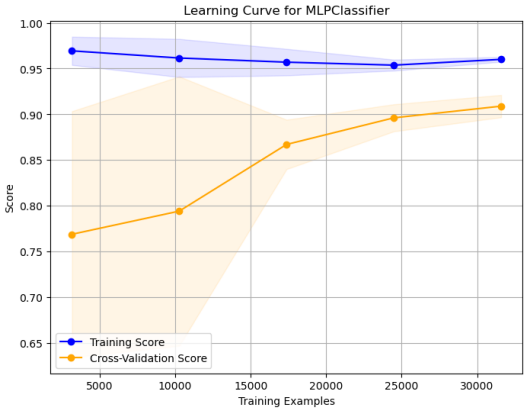
True Negative (TN): The model correctly predicted a negative class instance as negative.

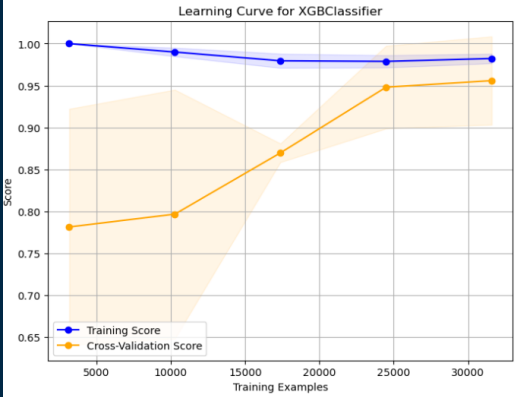
False Negative (FN): The model incorrectly predicted a positive class instance as negative.

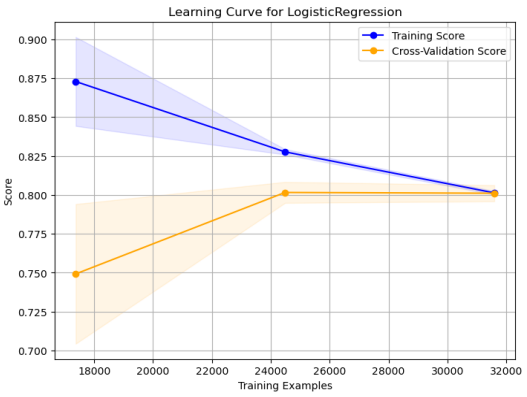
	Correct classifications	Wrong classifications
XGBoost	2678	3806
Random forest	5691	793
Logistic regression	5278	1206
Neural Network	5674	810

Learning Curve

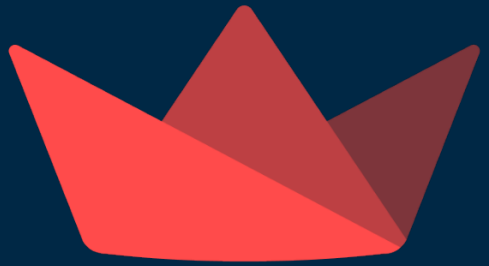








DEMONSTRATION



Streamlit

CONCLUSION

In this research we discovered the world of credit risk analysis, presenting a long learning journey through various stages of data preprocessing, model development, and evaluation.

By using a range of machine learning techniques and methods, this study has demonstrated the potential of predictive models in assessing creditworthiness accurately.

In this project, we have been exposed to many concepts like:

--> Building a pipeline
--> Hyper-parameter tuning
--> Evaluating models
--> Building our first Streamlit application
--> Deploying it.

Any questions?

mohamed.islem.beggari@etu.univ-st-etienne.fr

THANK YOU FOR YOUR ATTENTION