Assumed Role

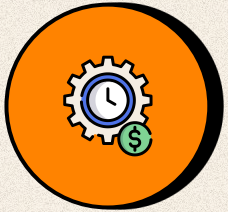
Data Scientist working at Global Finance Company

Background

The management wants to create an intelligent system to categorize individuals into credit score brackets to help the company make smarter, faster, and more profitable decisions.

QUICK QUESTIONS

Would you agree that more accurate credit scoring classification can help reduce lending risk and improve profitability?



Do you think automating credit scoring classification could help reduce delays in offering financial services to customers?

Credit Score Classification

Machine Learning Solution

Table of contents

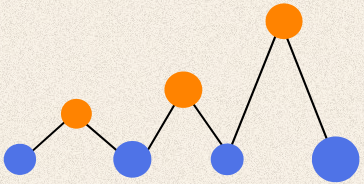
01

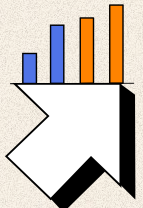
PROJECT GOAL

WHY ??

PROJECT GOAL

The goal is to develop a machine learning model that categorizes individuals into credit score brackets (Good, Standard, and Poor), helping the company make more informed lending decisions.





02

PROJECT BENEFITS

WHAT FOR ??

PROJECT BENEFITS

Loan Approval and Risk Management

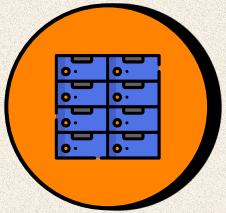
The machine learning model helps our company evaluate the risk of lending to individuals by categorizing them based on their credit scores.

Personalized Financial Offerings

The model lets our company offer customized financial products, like loans or credit cards, with terms based on each customer's credit score. Higher scores lead to better offers, while lower scores may require higher rates or extra security.

Interest Rates

Individuals with higher credit scores get better loan terms, while those with lower scores are seen as riskier and may face higher interest rates.



Profitability

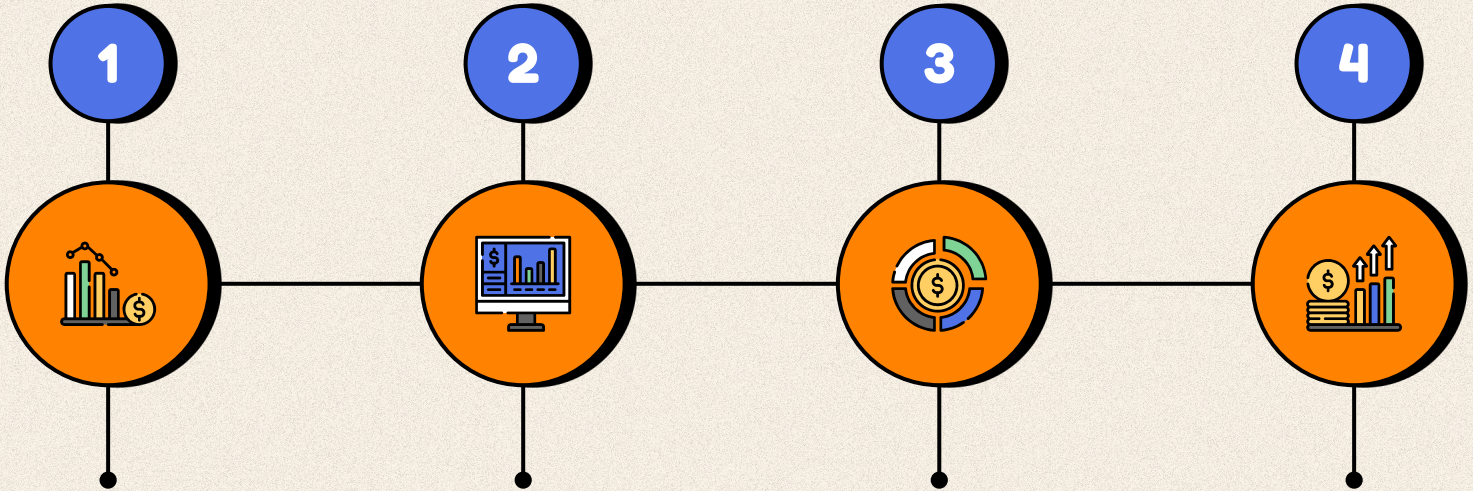
By categorizing customers, our company can manage portfolio more effectively. We can focus on offering high-value services to customers with good credit, while also managing the risk associated with those who have poor credit.

03

PROJECT FLOW

HOW DOES IT WORK??

PROJECT FLOW



STAGE 1

EXPLORE OUR DATASET

STAGE 2

DATA PREPARATION

STAGE 3

DATA TRAINING AND EVALUATION

STAGE 4

MODEL DEPLOYMENT

04

ALGORITHM EVALUATION

WHICH ONE IS THE BEST??

ALGORITHM EVALUATION

	KNN	SVC	Decision Tree	Random Forest	XGBOOST	ADABOOST
Train F1-Score	0.678, 0.745, 0.809	0.593, 0.727, 0.769	1.0, 1.0, 1.0	1.0, 1.0, 1.0	0.675, 0.730, 0.780	0.0, 0.031, 0.693
Test F1-Score	0.514, 0.601, 0.698	0.579, 0.698, 0.748	0.618, 0.704, 0.741	0.709, 0.777, 0.802	0.644, 0.703, 0.762	0.0, 0.026, 0.693
Cross Validation	0.601	0.667	0.685	0.754	0.705	-
Standard Deviation	0.003	0.005	0.005	0.005	0.006	-
Train Time	~5 Seconds	~5 Minutes	15 Seconds	~1 Minutes	~110 Seconds	~3 Hours

XGBOOST WITH HYPER PARAMETER TUNING

	XGBOOST WITH HYPER PARAMETER TUNING
Train F1-Score	0.680, 0.735, 0.784
Test F1-Score	0.639, 0.705, 0.760
Cross Validation	0.706
Standard Deviation	0.005
Train time	~3 minutes

05

MODEL DEPLOYMENT TEST

HUGGINGFACE

06

FUTURE IMPROVEMENT

HOW TO IMPROVE??

FUTURE IMPROVEMENT

Addressing Overfitting:
While tree-based algorithms generally outperform KNN and SVC in our evaluation, they tend to overfit. This can be mitigated by reducing the data dimensions through better feature selection and the application of Principal Component Analysis (PCA).

Balancing Data Distribution:
To reduce bias in the output, we plan to balance the data distribution, ensuring that the model learns from a more representative sample of the data.

Further Enhancing Model Performance:
There's still room to improve the model's performance. We will continue to experiment with hyperparameter tuning, additional feature engineering, and potentially incorporating other machine learning techniques.

Thank You!