



Transparent Credit Scoring Using Explainable Machine Learning

Chong Tin Tak | IEDA4500 Engineering Foundation of FinTech

Table of contents

01

Introduction

What is credit scoring?

02

Data Preprocessing

Feature engineering

03

Modeling Approaches

Logistic regression
Random forest
XGBoost

04

Model Evaluation Metrics

Accuracy
Precision
Recall

05

ROC-AUC Curve

What is ROC?

06

SHAP

Top important features

Table of contents



07

Final Model

Which model did I choose?

08

Challenges

What challenges did I face?

09

Conclusion

Possible improvements

A low-angle, black and white photograph of several skyscrapers reaching towards a cloudy sky. The perspective creates a sense of height and scale. A white rectangular frame is centered over the image.

01 Introduction

Introduction

- What is credit scoring?
- Why is it important in FinTech?
- Why is transparency essential?

Goal — Predict creditworthiness using machine learning in a transparent and fair way.



Data Preprocessing 02

Dataset Overview

Types of features (Age, income, balance, outstanding debt, etc.)

Target variables – Credit_Score (Good, Standard, Poor)

Reference:

<https://www.kaggle.com/datasets/parisrohan/credit-score-classification/data>

Data Preprocessing

Handling missing value

Feature Engineering

Encoding categorical variables

Scaling



03

Modeling Approaches

Modeling Approaches

Logistic Regression



- ✓ Interpretable
- ✓ Fast
- ✓ Decent when features have linear relationships

Random Forest



- ✓ Handles non-linearity
- ✓ Robust to outliers
- ✓ Ignore irrelevant features

XGBoost



- ✓ Efficient
- ✓ Handles missing data internally
- ✓ Prevent overfitting

An aerial, high-angle photograph of a city street grid, showing a dense pattern of buildings and streets. A white rectangular frame is superimposed over the center of the image.

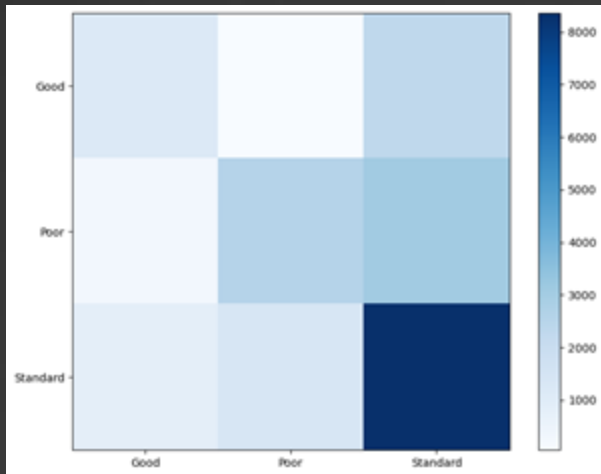
Model Evaluation Metrics

04

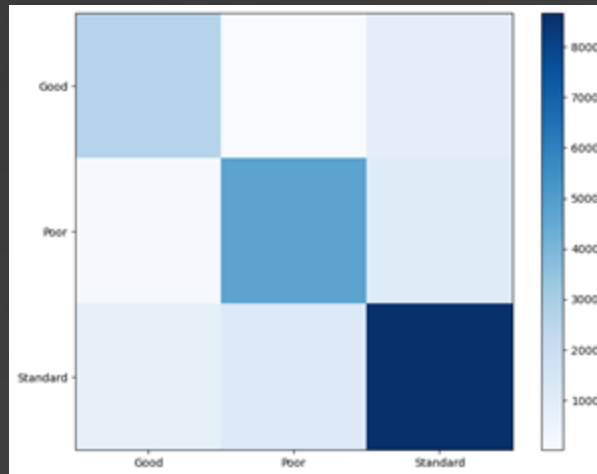
Model Evaluation Metrics

	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.6049	0.59	0.52	0.54
Random Forest	0.8013	0.79	0.79	0.79
XGBoost	0.7599	0.75	0.74	0.74

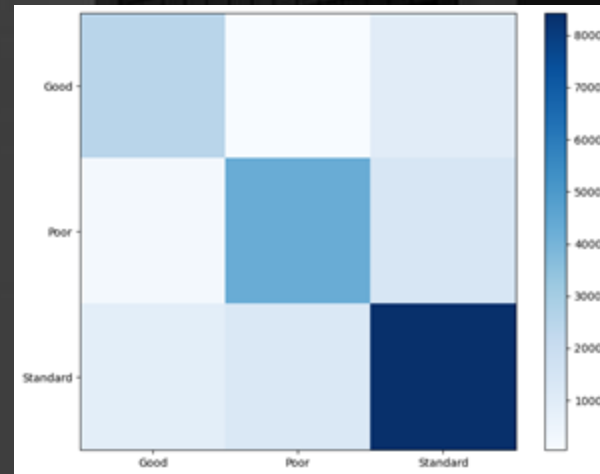
Confusion Matrix



Logistic
Regression



Random Forest



XGBoost



05

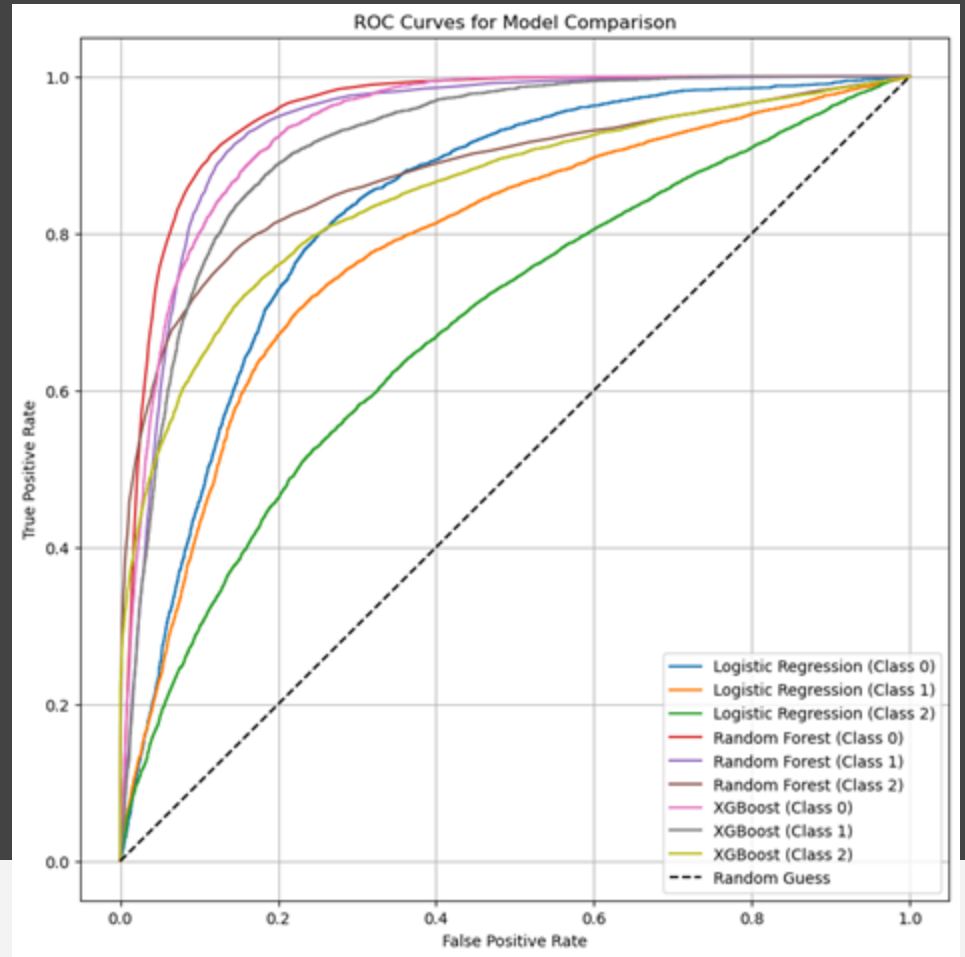
ROC-AUC
Curve

ROC Curves

ROC = Receiver Operating Characteristic

AUC = Area Under the Curve

Logistic Regression	0.7661
Random Forest	0.9228
XGBoost	0.9008



A low-angle, black and white photograph of several tall skyscrapers reaching towards a cloudy sky. The perspective is from the ground looking up, creating a sense of height and scale. The buildings are modern, with many windows visible. The sky is filled with soft, grey clouds.

SHAP

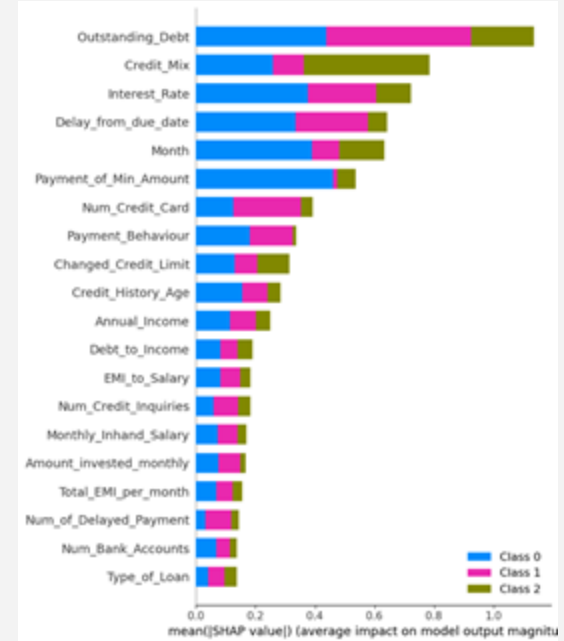
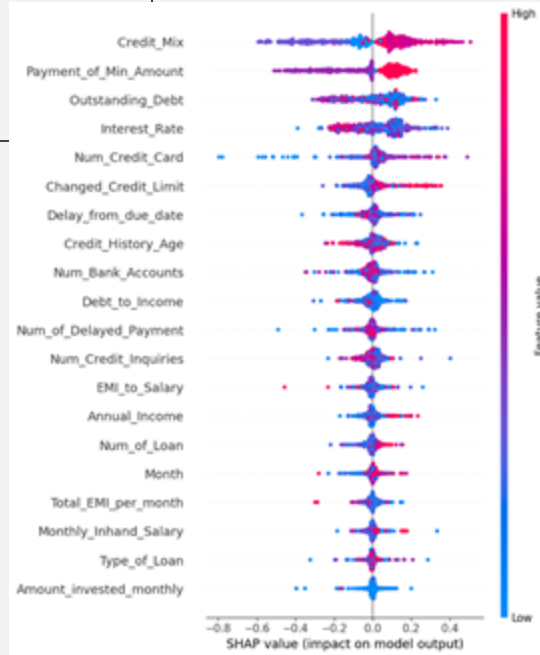
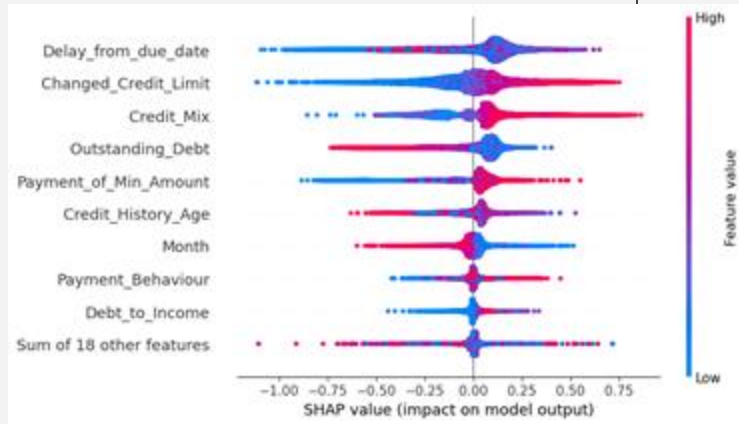
06

SHAP (SHapley Additive exPlanations)

Random Forest

XGBoost

Logistic Regression





07

Final Model

Final Model — Random Forest

High accuracy

Achieved 80.1 % accuracy

Balanced metrics

Precision, recall, F1-score
are highest

Handles non-linearity

Captures complex relationships
and interactions

Robustness

Less sensitive to noise
and outliers

Explainability

SHAP works well with
tree based models



08

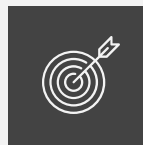
Challenges

Challenges



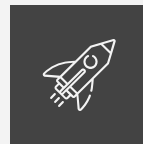
SHAP runtime

Sampling 500 data



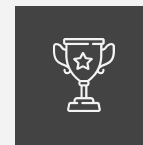
Model explainability

Learned how to use
SHAP for different model
types



Overfitting risk

Used validation set
performance and metrics
like macro F1 for
generalization



Feature availability

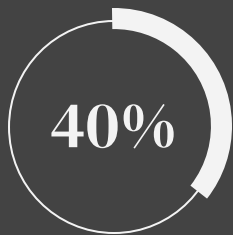
Implemented robust
error handling and
fallback imputation
methods



09

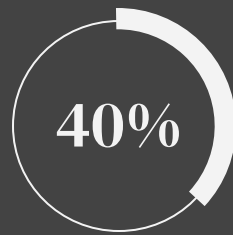
Conclusion

My Findings



Built a model

Transparent and high performing credit scoring model



SHAP and ROC

For interpretability and validation



AI in FinTech

Ethical and responsible



Thanks

Q&A

CREDITS: This presentation template was created by **Slidesgo**, and includes icons, infographics & images by **Freepik**

Please keep this slide for attribution