Credit Score Analysis

Arpan Banerjee Saurabh Kankekar Suyash Mohta

Business Case

Predictive Model Development



- Forecasts credit scores using historical financial data
- Helps stakeholders predict customer creditworthiness.

Categorization of Credit Scores



- Segments into Bad, Average, and Good
- Simplifies risk evaluation for lenders

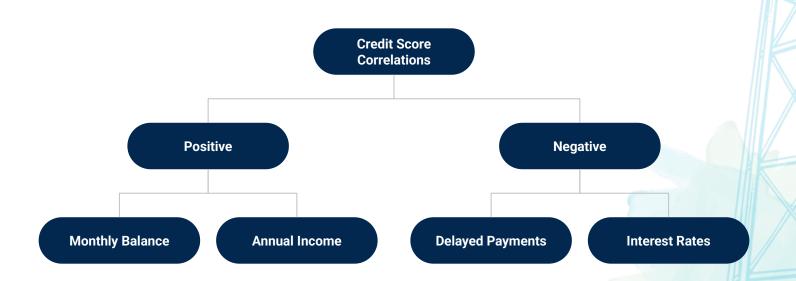
Value Proposition for Stakeholders



- Supports data-driven decisions to mitigate risk
- Empowers lenders with actionable insights.



Key Findings from EDA





Data Cleaning and Transforming

Imputation of missing values

8 rows of data per customer aggregated into single rows

Used Average for numerical columns

Used Mode for categorical columns



Correlation Matrix

- monthly_inhand_salary (correlation = 0.86 with annual_income)
- amount_invested_monthl
 y (correlation = 0.03 with
 credit_score)

Correlation Matrix of Financial Features with Credit Score

- 0.50

- 0.25

-0.25

-0.50

age	1.00	-0.00	-0.00	0.01	-0.01	0.03	0.00	0.01	-0.02	-0.02	0.01	-0.04	-0.00	-0.02	0.02	-0.00	0.00
annual_income	-0.00	1.00	0.99	0.00	-0.01	-0.01	0.01	0.02	0.02	0.01	-0.03	-0.02	-0.06	-0.05	-0.59	-0.07	0.86
monthly_inhand_salary	-0.00	0.99	1.00	0.01	-0.01	-0.01	0.01	0.02	0.03	0.02	-0.03	-0.03	-0.07	-0.06		-0.07	0.86
credit_history_age	0.01	0.00	0.01	1.00	0.01	-0.00	-0.04	0.01	-0.01	-0.07	0.03	0.03	0.01	0.02	0.01	0.01	-0.01
total_emi_per_month	-0.01	-0.01	-0.01	0.01	1.00	-0.02	0.02	0.02	0.03	0.01	-0.03	0.04	-0.02	-0.04	-0.02	0.01	-0.04
num_bank_accounts	0.03	-0.01	-0.01	-0.00	-0.02	1.00	0.02	0.02	-0.02	-0.04	0.03	0.07	0.01	-0.02	-0.02	0.04	0.00
num_credit_card	0.00	0.01	0.01	-0.04	0.02	0.02	1.00	-0.04	0.02	-0.00	0.01	0.03	-0.03	-0.03	-0.03	-0.01	0.02
interest_rate	0.01	0.02	0.02	0.01	0.02	0.02	-0.04	1.00	0.04	-0.02	-0.02	-0.03	0.02	0.01	-0.01	0.00	0.01
num_of_loan	-0.02	0.02	0.03	-0.01	0.03	-0.02	0.02	0.04	1.00	-0.03	-0.02	-0.04	-0.04	-0.00	-0.02	0.01	0.01
delay_from_due_date	-0.02	0.01	0.02	-0.07	0.01	-0.04	-0.00	-0.02	-0.03	1.00	-0.02	-0.03	-0.01	-0.04	-0.04	-0.03	0.03
num_of_delayed_payment	0.01	-0.03	-0.03	0.03	-0.03	0.03	0.01	-0.02	-0.02	-0.02	1.00	-0.01	0.02	0.02	0.03	-0.01	-0.03
changed_credit_limit	-0.04	-0.02	-0.03	0.03	0.04	0.07	0.03	-0.03	-0.04	-0.03	-0.01	1.00	-0.01	0.02	0.01	0.03	-0.03
num_credit_inquiries	-0.00	-0.06	-0.07	0.01	-0.02	0.01	-0.03	0.02	-0.04	-0.01	0.02	-0.01	1.00	-0.04	0.02	0.03	-0.03
outstanding_debt	-0.02	-0.05	-0.06	0.02	-0.04	-0.02	-0.03	0.01	-0.00	-0.04	0.02	0.02	-0.04	1.00	0.73	-0.02	-0.55
credit_utilization_ratio	0.02	-0.59	-0.59	0.01	-0.02	-0.02	-0.03	-0.01	-0.02	-0.04	0.03	0.01	0.02	0.73	1.00	0.05	-0.86
amount_invested_monthly	-0.00	-0.07	-0.07	0.01	0.01	0.04	-0.01	0.00	0.01	-0.03	-0.01	0.03	0.03	-0.02	0.05	1.00	-0.05
monthly_balance	0.00	0.86	0.86	-0.01	-0.04	0.00	0.02	0.01	0.01	0.03	-0.03	-0.03	-0.03	-0.55	-0.86	-0.05	1.00

ster of the contraction of the state of the contraction of the contrac

Metrics



→ Weighted Cohen's Kappa

Measures inter-rater agreement while accounting for the degree of disagreement using weights.

→ F1 score

A balance between precision and recall, useful for evaluating classification models, especially on imbalanced data.

Cohen's Weighted Kappa (κ_W):

$$\kappa_W = 1 - rac{\sum w_{ij} \cdot fo_{ij}}{\sum w_{ij} \cdot fe_{ij}}$$

F1-Score:

$$F1 = 2 imes rac{ ext{Precision} imes ext{Recall}}{ ext{Precision} + ext{Recall}}$$



Model Selection & Evaluation

Preprocessing

- Dropping irrelevant columns
- One-hot encoding the categorical
- Scaling the numerical features

Fitting Models

- Hyperparameter Tuning
- Cross Validation

Selecting Best Model

 Based on overall balance of F1-score & Weighted Kappa Score

Model Inference

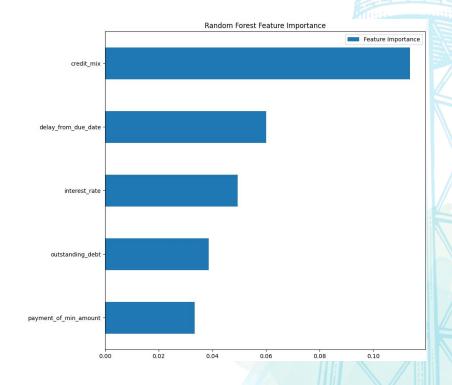
- Permutation importance
- PDP & ICE Plots

	F1 Score	Weighted Kappa
Logistic Regression	0.709565	0.597992
Random Forest	0.767050	0.683618
XGBoost	0.758447	0.670593
CatBoost	0.762795	0.661908

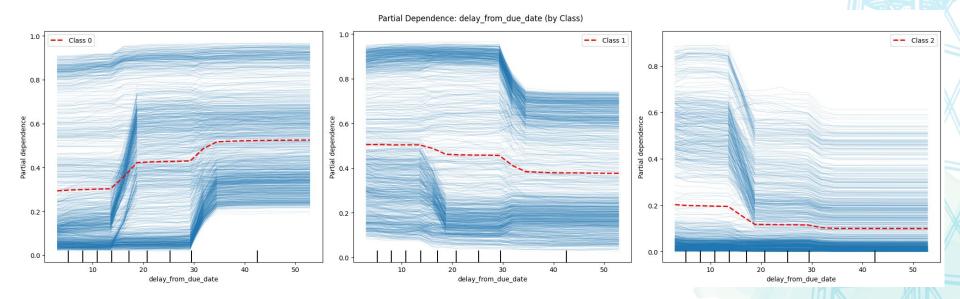
Permutation Importance

Feature's impact on model performance by randomly shuffling its values and observing the performance drop

- Credit_mix (Bad, Standard, Good)
- Delay_from_due_date
- Interest rate
- Outstanding_debt
- Payment_of_min_amount (Yes, No)



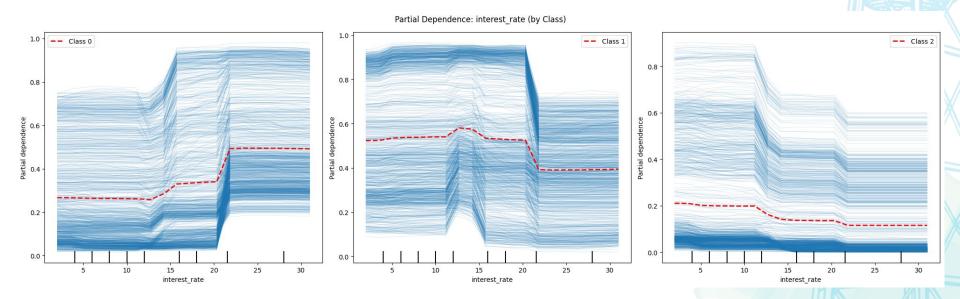
- Partial Dependence Plots
- Individual Conditional Expectation Plots



More the delay, more probability of having lower credit score



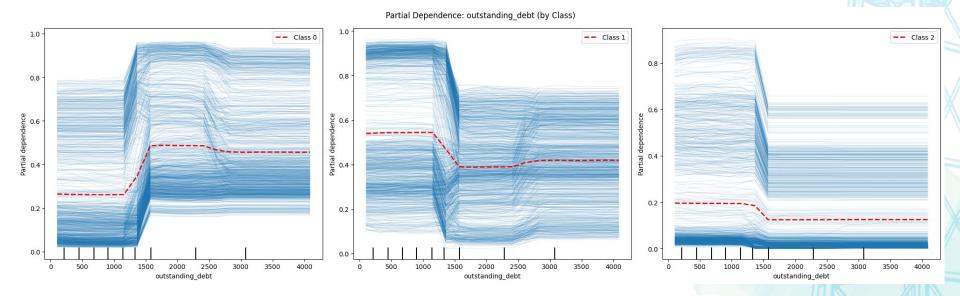
- Partial Dependence Plots
- Individual Conditional Expectation Plots



Lower the interest rate, more probability of having high credit score



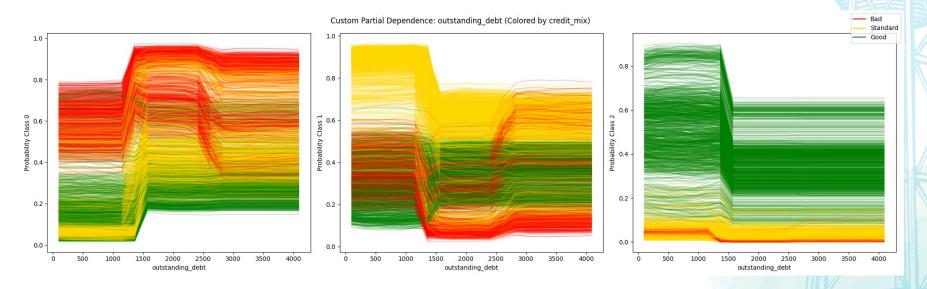
- Partial Dependence Plots
- Individual Conditional Expectation Plots



More the outstanding debt, more probability of having lower credit score



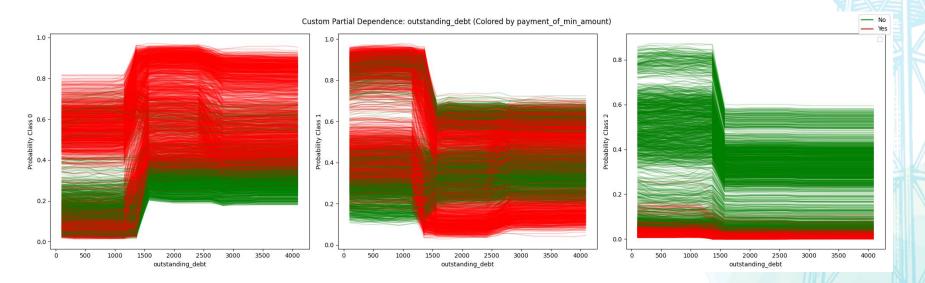
Custom Individual Conditional Expectation Plots



Having a Good Credit Mix is critical, however large outstanding debt weighs down probability of good credit score



Custom Individual Conditional Expectation Plots



Paying more than min_amount_due is important, however large outstanding debt weighs down probability of good credit score



Limitations & Future Work

 Causal Inference: Exploring causal inference techniques to understand the direct drivers of creditworthiness and improve model interpretations.

 Model Generalizability: Testing & Training the model on diverse data to ensure it generalizes well to new populations and situations.

 Ethical Considerations: As the model influences credit decisions, it's important to address potential biases and ensure fairness

