Loan Default Prediction

Problem Statement

Given information on the loan and borrower, what is the probability of default?

Motivations

- Financial stability and profitability for banks
- High-risk borrowers lead to potential losses and non-performing loans

Goals:

- Avoid loss
- Better allocation of resources

Dataset

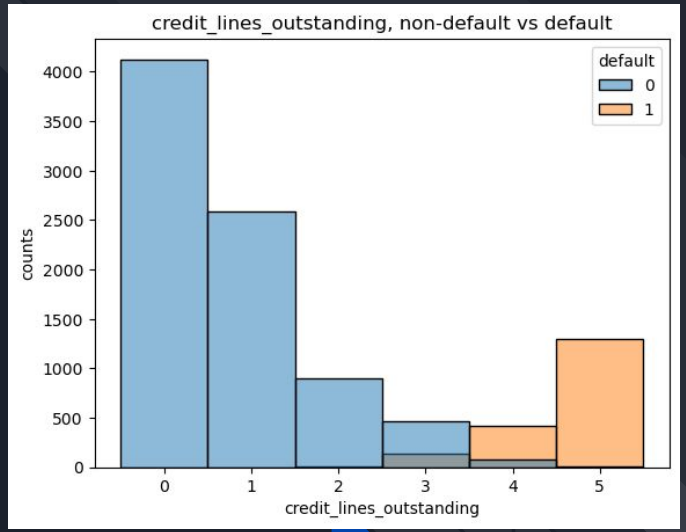
	customer_id	Credit_lines outstanding	Loan_amt outstanding	Total_debt outstanding	income	Years employed	Fico score	default
0	8153374	0	5221.545193	3915.471226	78039.38546	5	605	0
1	7442532	5	1958.928726	8228.752520	26648.43525	2	572	1
2	2256073	0	3363.009259	2027.830850	65866.71246	4	602	0
3	4885975	0	4766.648001	2501.730397	74356.88347	5	612	0
4	4700614	1	1345.827718	1768.826187	23448.32631	6	631	0

Exploratory Data Analysis / Feature Importance

	Feature Id	Importances
0	credit_lines_outstanding	41.642504
1	years_employed	27.133336
2	total_debt_outstanding	10.848737
3	fico_score	10.598897
4	income	6.910185
5	loan_amt_outstanding	2.866341

Credit lines outstanding is a key indicator on default probability!

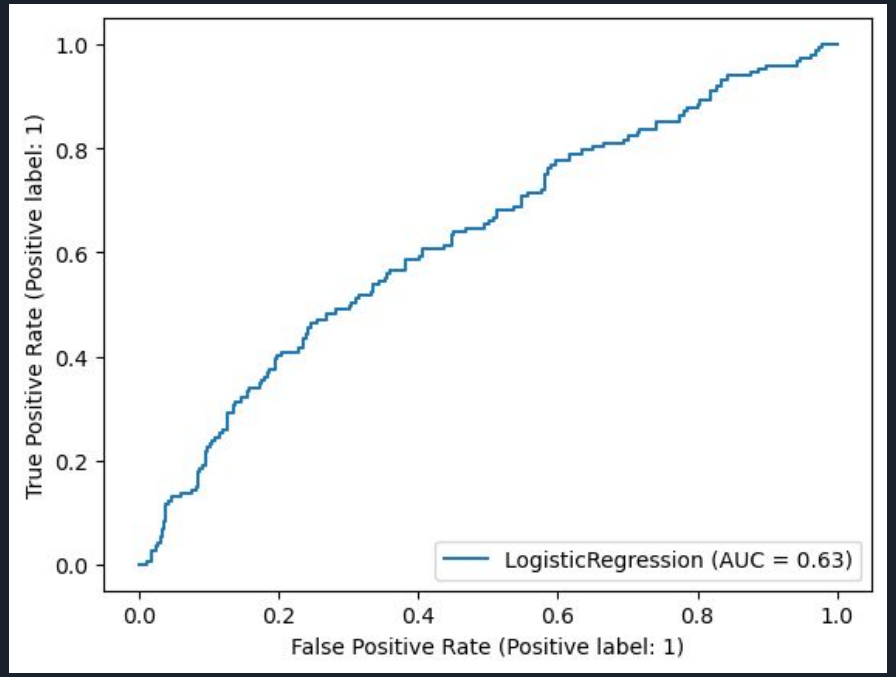
The default vs. non-default populations differ significantly in this attribute...



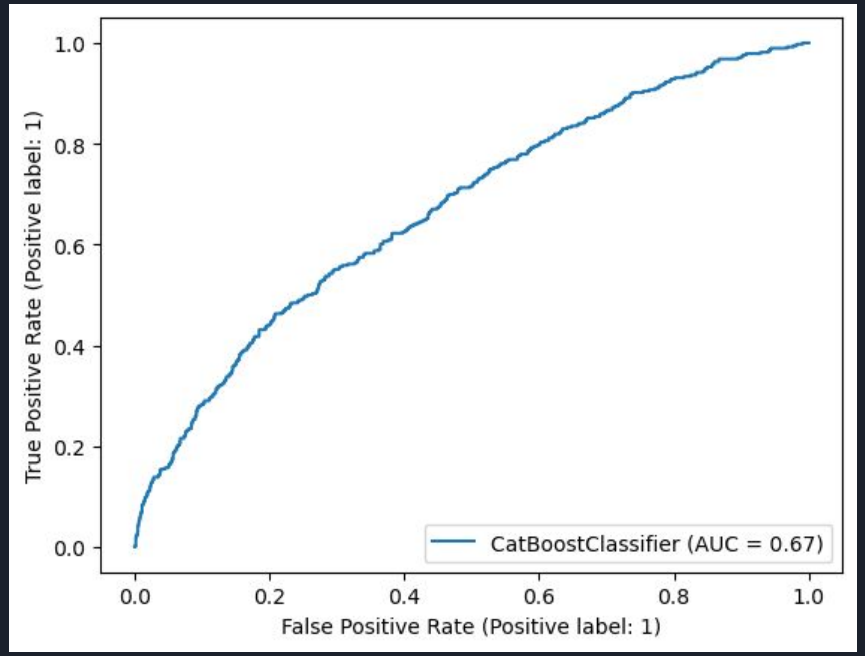
But what if some of the information is not available?

- Consider removing some features and train a model on just income, fico_score and years_employed
- These are commonly asked for by banks for entry-level credit cards

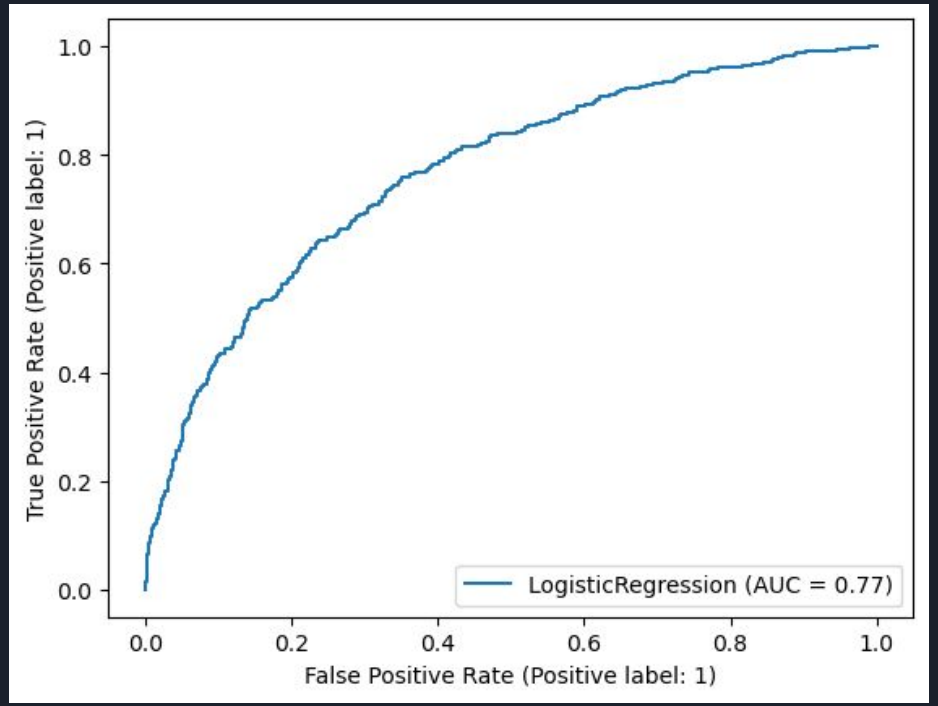
Model Comparisons -
Default Logistic Regression Classifier



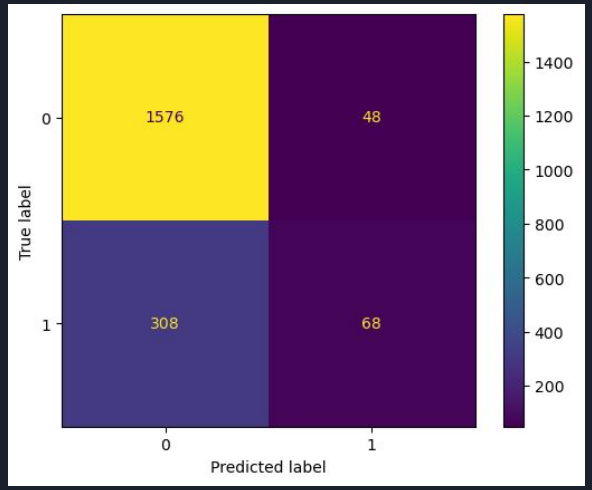
Model Comparisons - Tuned CatBoost Classifier



Model Comparisons -
Tuned Logistic Regression Classifier



Final Model Metrics



	precision	recall	f1-score	support
0(no default)	0.84	0.97	0.90	1624
1(default)	0.59	0.18	0.28	376
accuracy			0.82	2000
macro avg	0.71	0.58	0.59	2000
weighted avg	0.79	0.82	0.78	2000

Brier score loss: 0.12757509999134695
log loss: 0.4087386060695741
ROC_AUC score: 0.7668041216853578

Further Considerations

- Building a feature which further penalizes mis-identification of default class
- Weigh the samples with higher loan amount more to mitigate overall risk in the portfolio (i.e. worse mistake if bigger default was not caught)
- Weigh the default class more depending on management needs