ML Methods of Classifying Credit Risks

Flatirons Consulting Team

Presented by Chan Lee

Date
December 2024

Table of Contents

Business goals

- Reduce human bias and error in classifying loan requests as good or bad credit risks
- Find previously unnoticed indicators
  - Employees currently using about 12 attributes to make decisions
  - Organization is capturing up to a 100 attributes
- Explore automated solutions as the first gate in decisions

Introducing the dataset

Credit-G Dataset

- Published in 1994 by Dr. Hans Hofmann at UC Irvine
- Classifies individual loan requests as good or bad credit risks
- Contains 1000 data points with 20 features

personal_status	other_parties	...	property_magnitude	age	other_payment_plans	housing	existing_credits	job	num_dependents	own_telephone	foreign_worker	class
b'male single'	b'none'	...	b'real estate'	67.0	b'none'	b'own'	2.0	b'skilled'	1.0	b'yes'	b'yes'	b'good'
b'female div/dep/mar'	b'none'	...	b'real estate'	22.0	b'none'	b'own'	1.0	b'skilled'	1.0	b'none'	b'yes'	b'bad'
b'male single'	b'none'	...	b'real estate'	49.0	b'none'	b'own'	1.0	b'unskilled resident'	2.0	b'none'	b'yes'	b'good'
b'male single'	b'guarantor'	...	b'life insurance'	45.0	b'none'	b'for free'	1.0	b'skilled'	2.0	b'none'	b'yes'	b'good'
b'male single'	b'none'	...	b'no known property'	53.0	b'none'	b'for free'	2.0	b'skilled'	2.0	b'none'	b'yes'	b'bad'

Introducing the dataset (cont.)

The Credit-G dataset comes with a cost matrix:

	Good (predicted)	Bad (predicted)
Good (actual)	0	1
Bad (actual)	5	0

i.e. it is far worse to classify a "bad" as "good" than vice-versa

Our Objective:

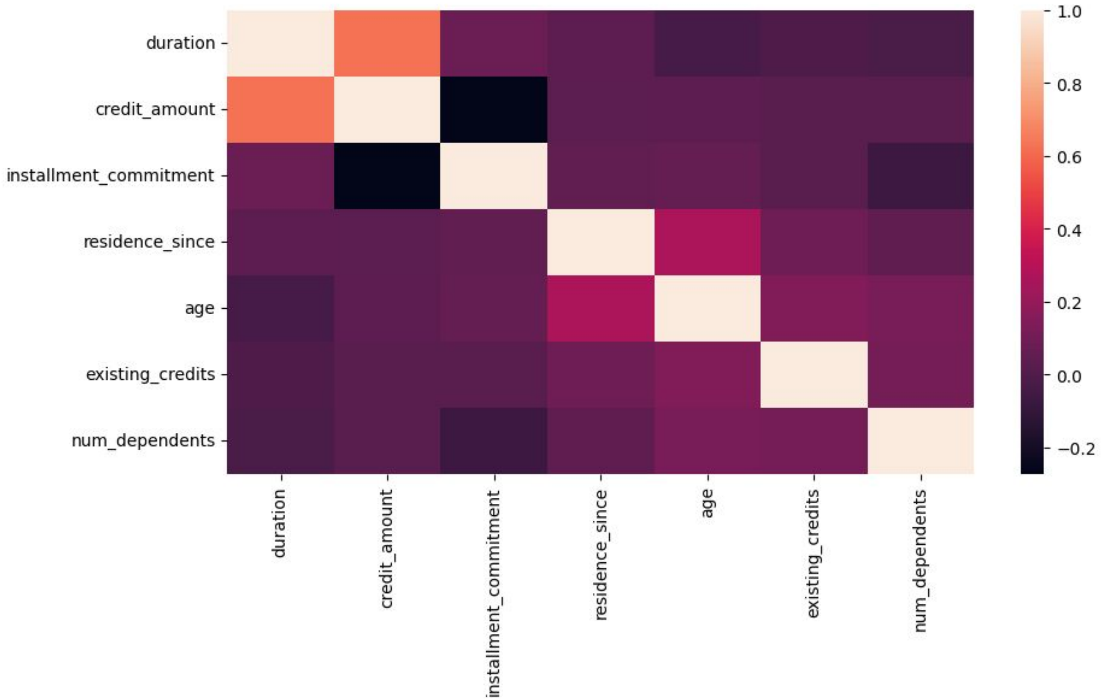
Can we train a model to predict good and bad credit risks accurately while minimizing cost?

Understanding the data

- A correlation heatmap shows which features are collinear, i.e. which features are related

- Credit amount and duration are highly correlated in the dataset
  - Intuitively, larger credit amounts are typically paid off over a longer period of time

- We correct for this by creating two new features
  - Monthly payment amount
  - Is loan short term (< 18 months*)?

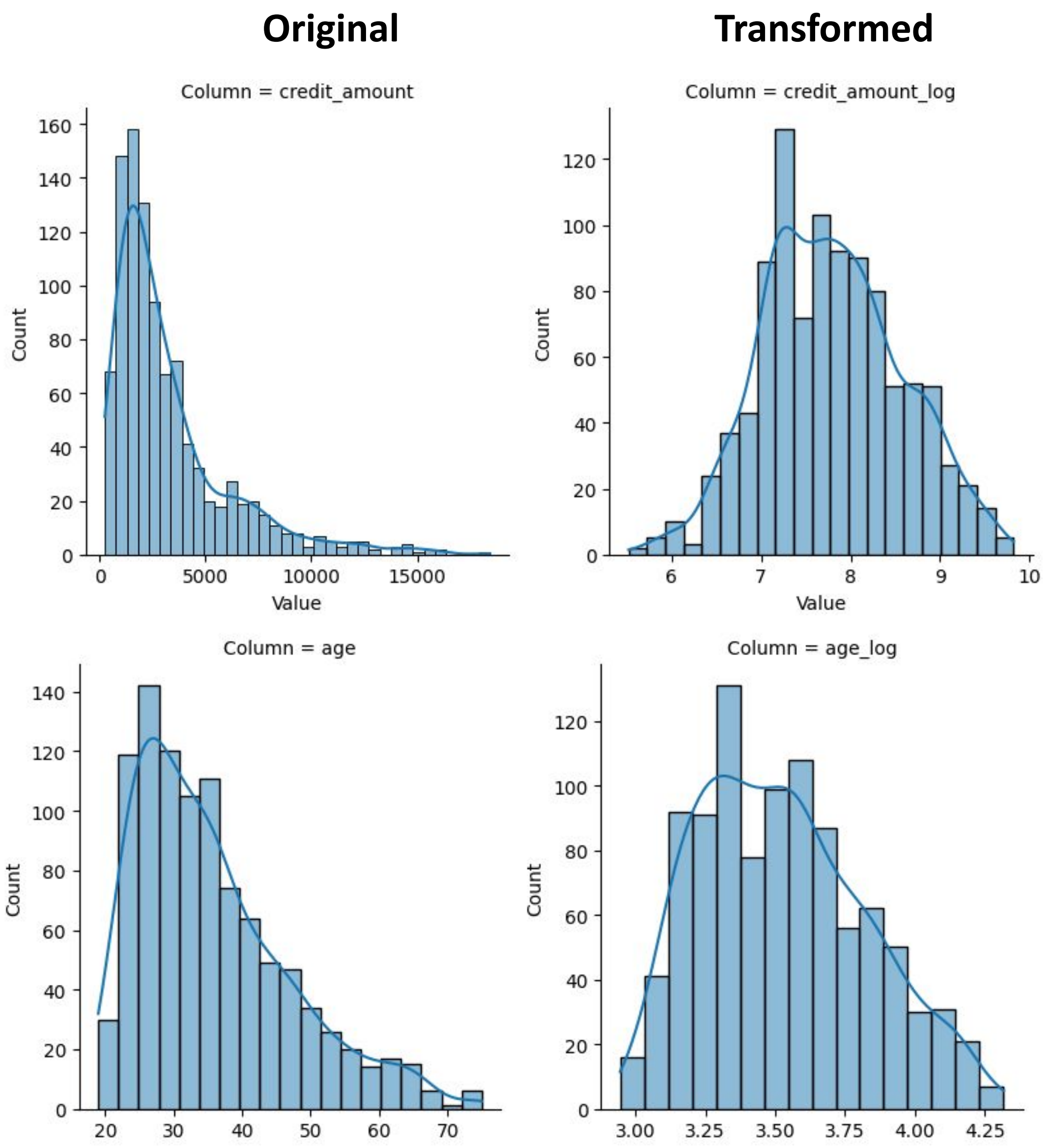


** as per https://corporatefinanceinstitute.com/resources/accounting/short-term-loan/*

Understanding the data

- Some numeric features are highly skewed and some models will benefit from normalization

- Models such as Random Forest do not require normalization and may negatively impact interpretation.



Understanding the data

- T-test can indicate to us if a numeric feature is a good predictor

	feature	t_stat	p
2	duration_log	-7.070214	4.406112e-12
3	credit_amount_log	-3.280379	1.109388e-03
0	age_log	3.238357	1.273879e-03
1	monthly_payment_amount_log	1.932800	5.383323e-02

- All features appear to be statistically significant with a p-value of <=0.05

Understanding the data

- For categorical and ordinal features, a Chi-square test indicates whether or not a feature will be a good predictor.
- Most features in the dataset appear to be good predictors.
- We will use all features with a p-value <= 0.05

	feature	chi2	p	dof
0	checking_status	123.720944	1.218902e-26	3
7	credit_history	61.691397	1.279187e-12	4
1	savings_status	36.098928	2.761214e-07	4
17	is_short_term	18.820055	1.436487e-05	1
11	property_magnitude	23.719551	2.858442e-05	3
13	housing	18.199842	1.116747e-04	2
8	purpose	33.356447	1.157491e-04	9
2	employment	18.368274	1.045452e-03	4
12	other_payment_plans	12.839188	1.629318e-03	2
16	foreign_worker	5.821576	1.583075e-02	1
9	personal_status	9.605214	2.223801e-02	3
10	other_parties	6.645367	3.605595e-02	2
3	installment_commitment	5.476792	1.400333e-01	3
15	own_telephone	1.172559	2.788762e-01	1
5	existing_credits	2.671198	4.451441e-01	3
14	job	1.885156	5.965816e-01	3
4	residence_since	0.749296	8.615521e-01	3
6	num_dependents	0.000000	1.000000e+00	1

Models

- We selected 4 models for training
  - Random Forest
  - Gradient Boosted Trees
  - Extra Trees
  - Convoluted Neural Networks

- 20% of the training data was reserved as test data
- All models were trained with cross validation with either the K-fold or train test split methodology
- All models were tuned with hyperparameters to get the best result

Models

Random Forest	Gradient Boosted Trees	Extra Trees	CNN
Accuracy: 0.77	Accuracy: 0.75	Accuracy: 0.73	Accuracy: 0.67

Initial round of training shows that Random Forest has the most promise.

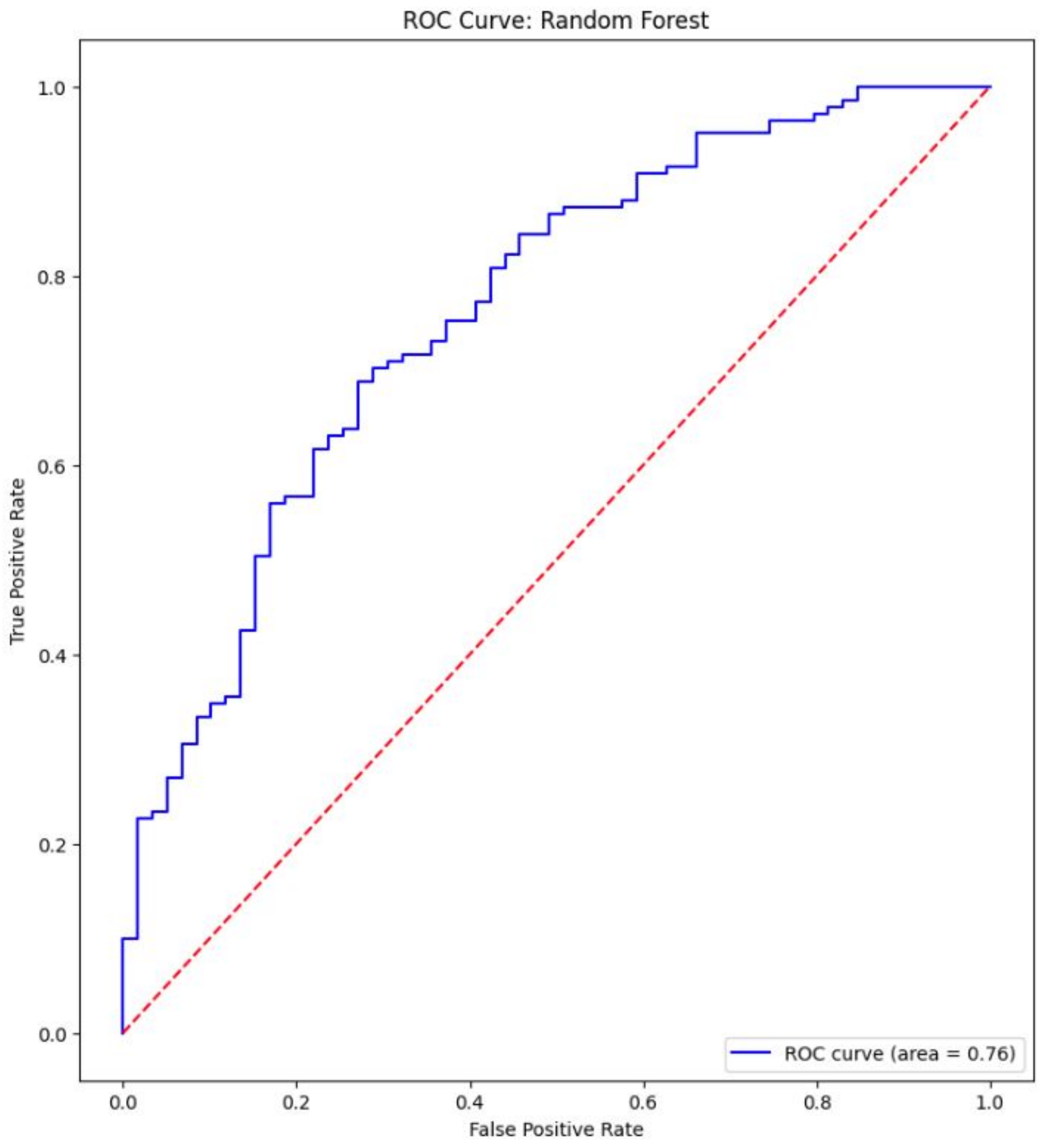
Let’s analyze its performance on the reserved test data

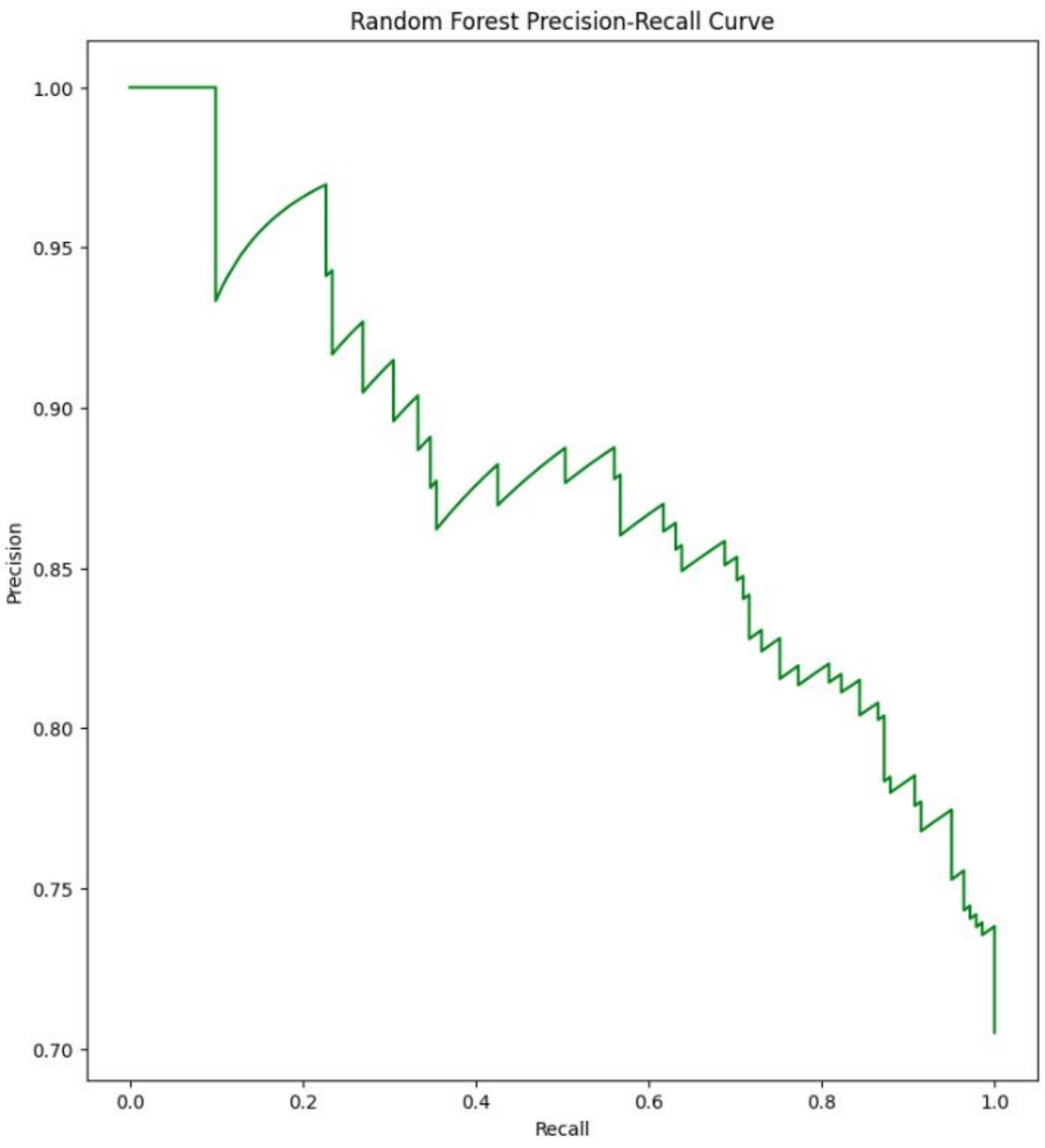
Model analysis on test data

- Accuracy: 0.77
  - Of all the predictions made, 77% were correct
- Precision: 0.774
  - When the model predicts a positive result, it is correct 77.4% of the time
- True positive rate (TPR): 0.95
  - Of all “good” credit risks, this model identified 95% correctly
- False positive rate (FPR): 0.661
  - Of all “bad” credit risks, the model incorrectly identified 66.1% of them as positive

In summary, the model is reliably predicting good credit risks but has a high rate of false “alarms”

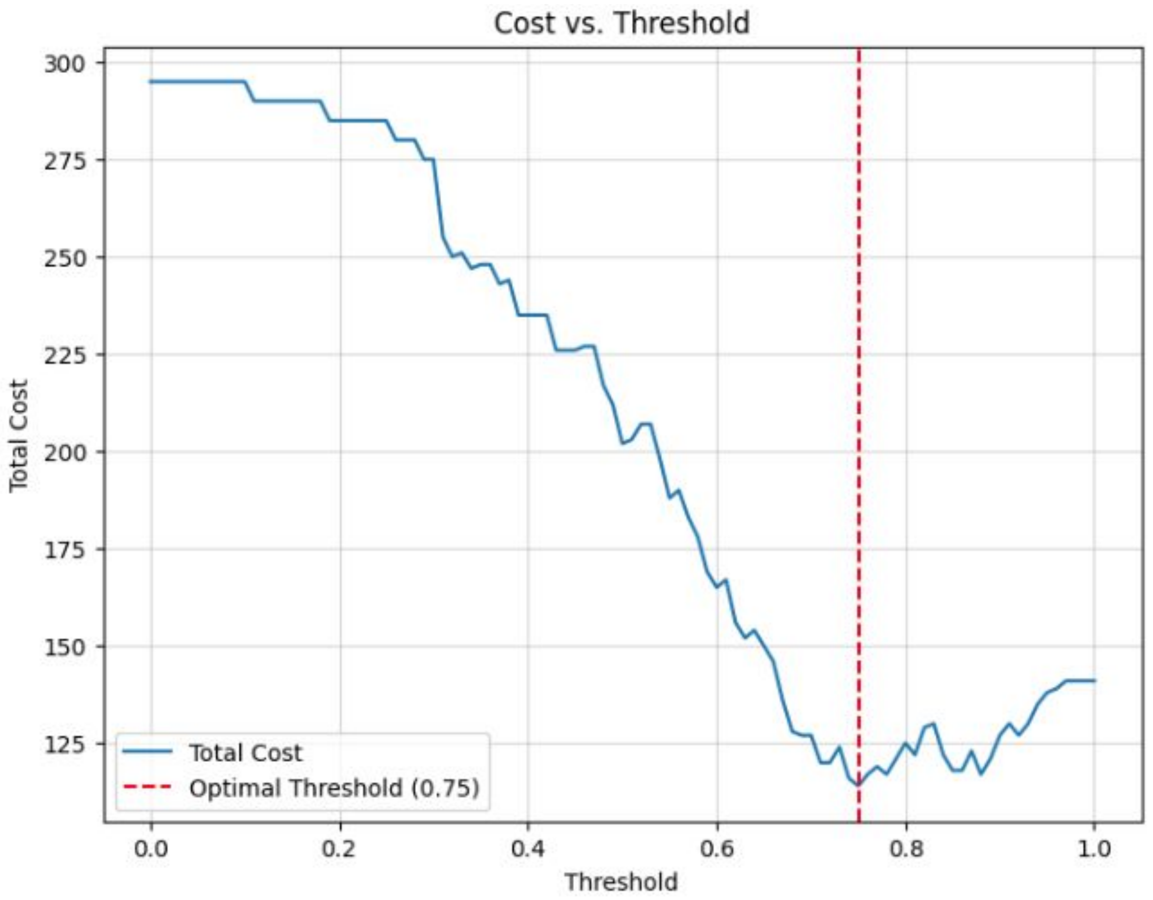
Model analysis on test data





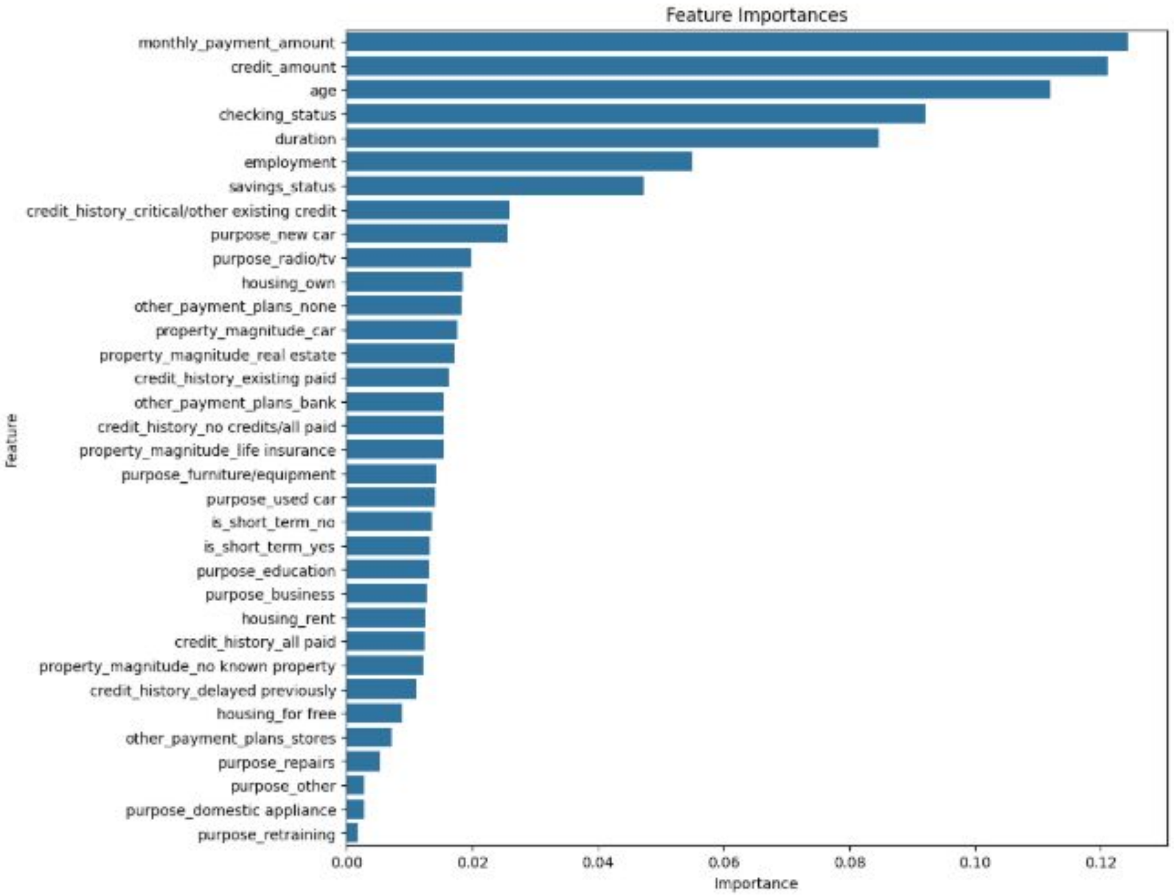
Model analysis on test data

- Cost matrix allows us to tune the threshold to minimize cost
- A threshold of 0.75 provides us
  - Accuracy: 0.63
  - Precision: 0.88
  - TPR: 0.54
  - FPR: 0.17
- In other words, we can lower our FPR at the cost of general accuracy



Model analysis on test data

Random Forest allows analysis extraction and “decision tracing”. Black box models such as CNN do not allow such analysis.



Feature importance as ranked by the trained model

```
Decision path for the first tree:
Node 0: Feature other_payment_plans_none > 0.5
Node 92: Feature savings_status <= 2.5
Node 93: Feature is_short_term_yes <= 0.5
Node 94: Feature monthly_payment_amount > 50.0
Node 96: Feature duration > 29.0
Node 146: Feature age > 25.5
Node 154: Feature credit_history_critical/other existing credit <= 0.5
Node 155: Feature credit_history_no credits/all paid <= 0.5
Node 156: Feature age > 55.5
Leaf Node 182: Class Distribution = [1. 0.], Predicted Class = 0
```

Decision path taken for a single data point

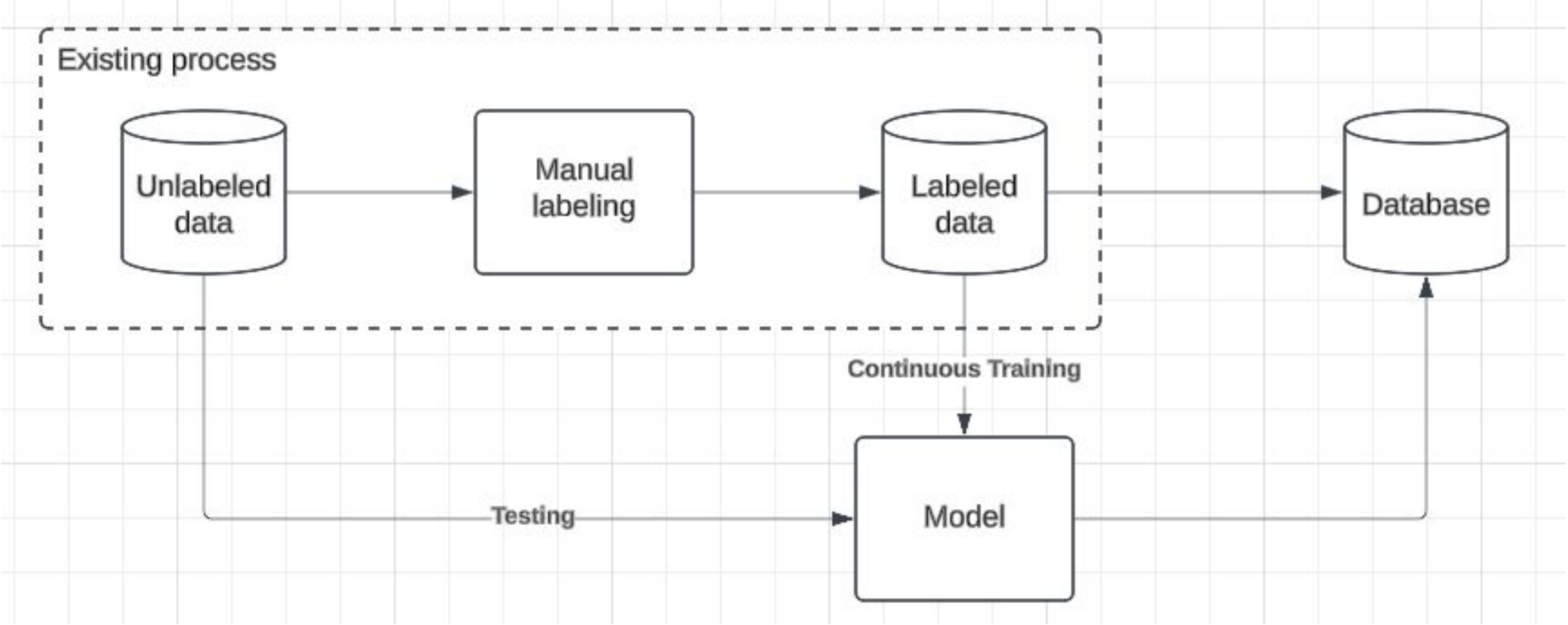
Model analysis on test data

In summary,

- We have proved it is possible to reliably predict good and bad credit risks
- Using Random Forest, we are able to audit how it makes decisions
- We can tune the trained model further using a cost matrix

Next steps

- More data and features will allow us to perform similar analysis to improve model performance
- We can begin operationalizing this model into your organization for continuous training using the CRISP-DM framework



FLATIRON CONSULTING