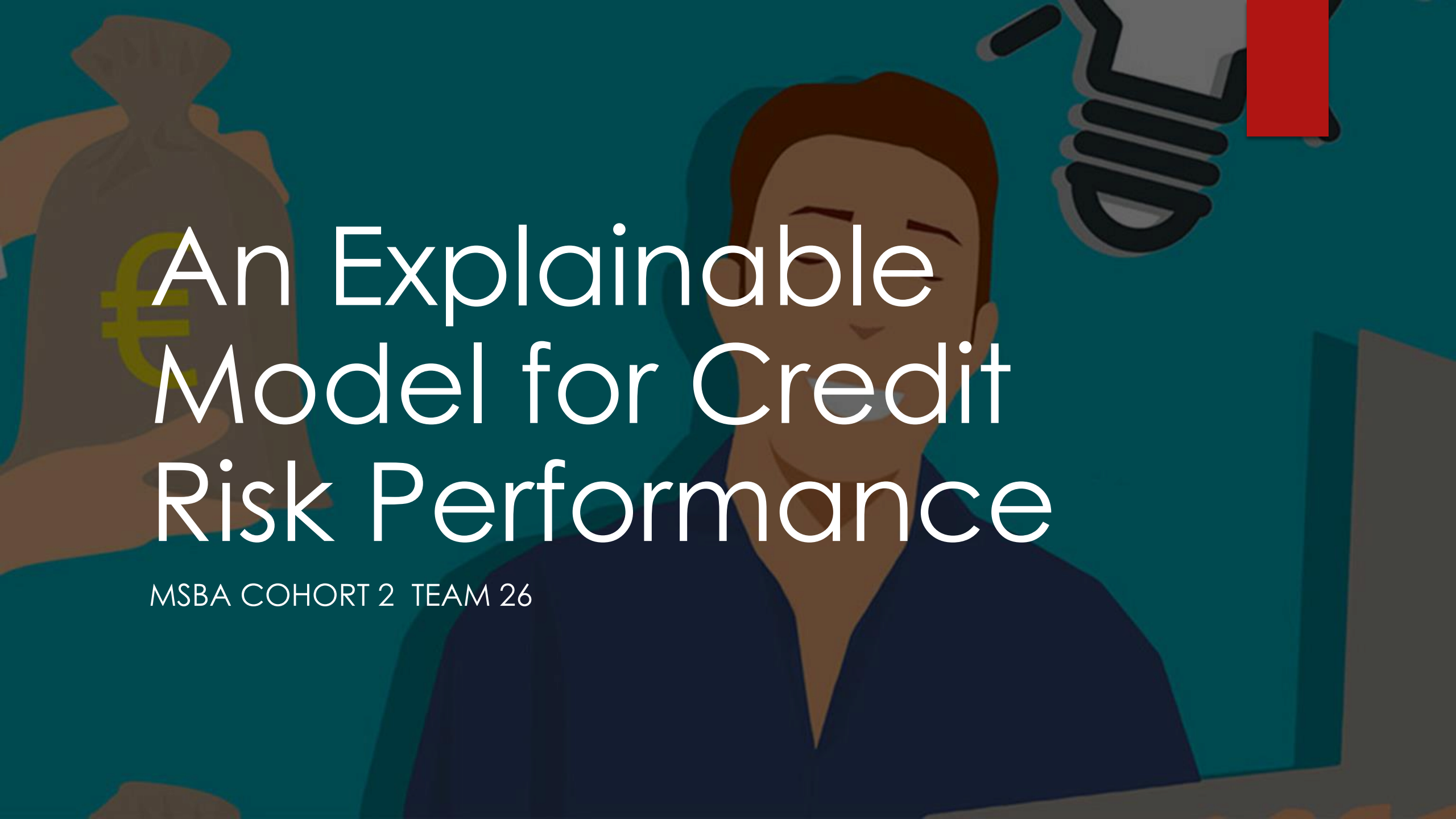
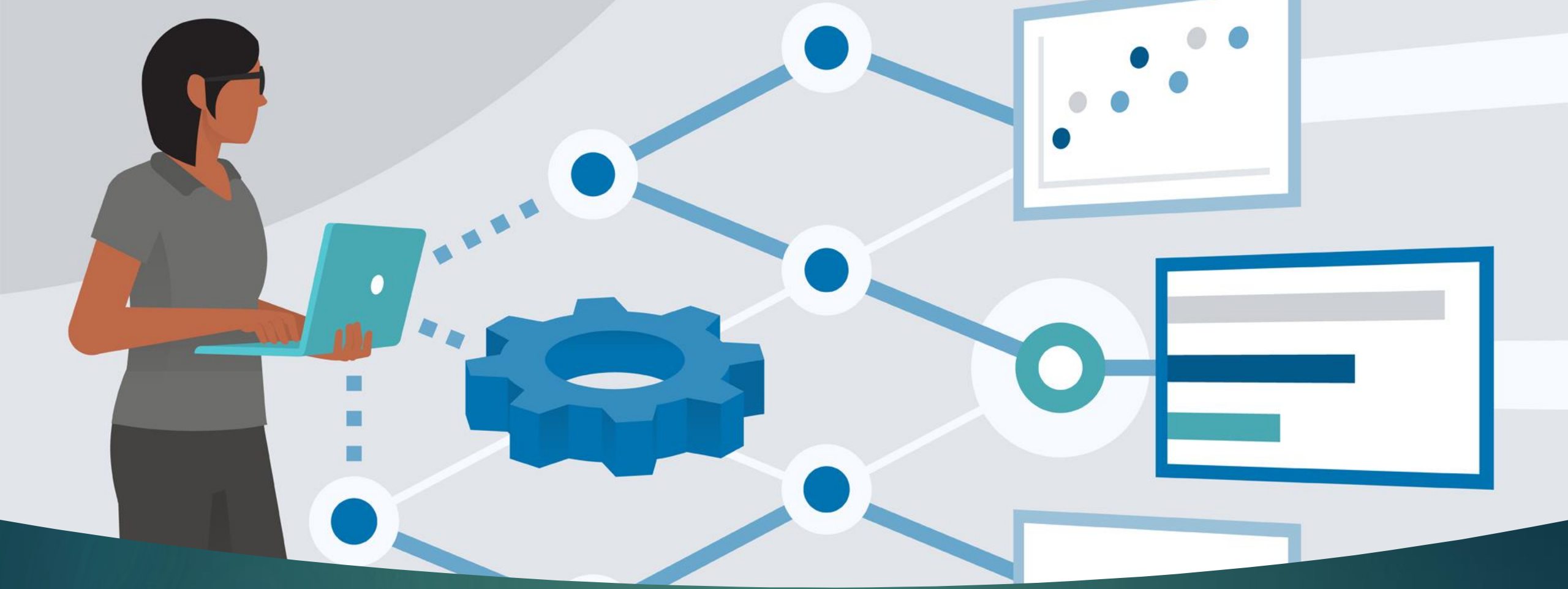


An Explainable Model for Credit Risk Performance

MSBA COHORT 2 TEAM 26

Introduction

- Random forest Model
- Logistic Model
- Other models



Data Cleaning (scaler)

Set Risk Performance to 0 and 1
Replace values that contains -7, -8, -9 to median or mean (base on standard deviation)

Model Selection (RMSE)

- Linear Regression: 0.42847
- Linear SVM: 0.43374
- SVM (Polynomial): 0.43389
- SVM (Radial basis): 0.52985
- Linear Discriminant Analysis: 0.51470
- KNN: 0.61209
- Logistic Regression: 0.52366

Best Model Explanation – Random Forest

- Prevalence:
- When we classified randomly, the positive rate = 47%(4735/(4735+5136))
- After our analytics, our accuracy improved 53%

Root Mean Squared Error: 43%

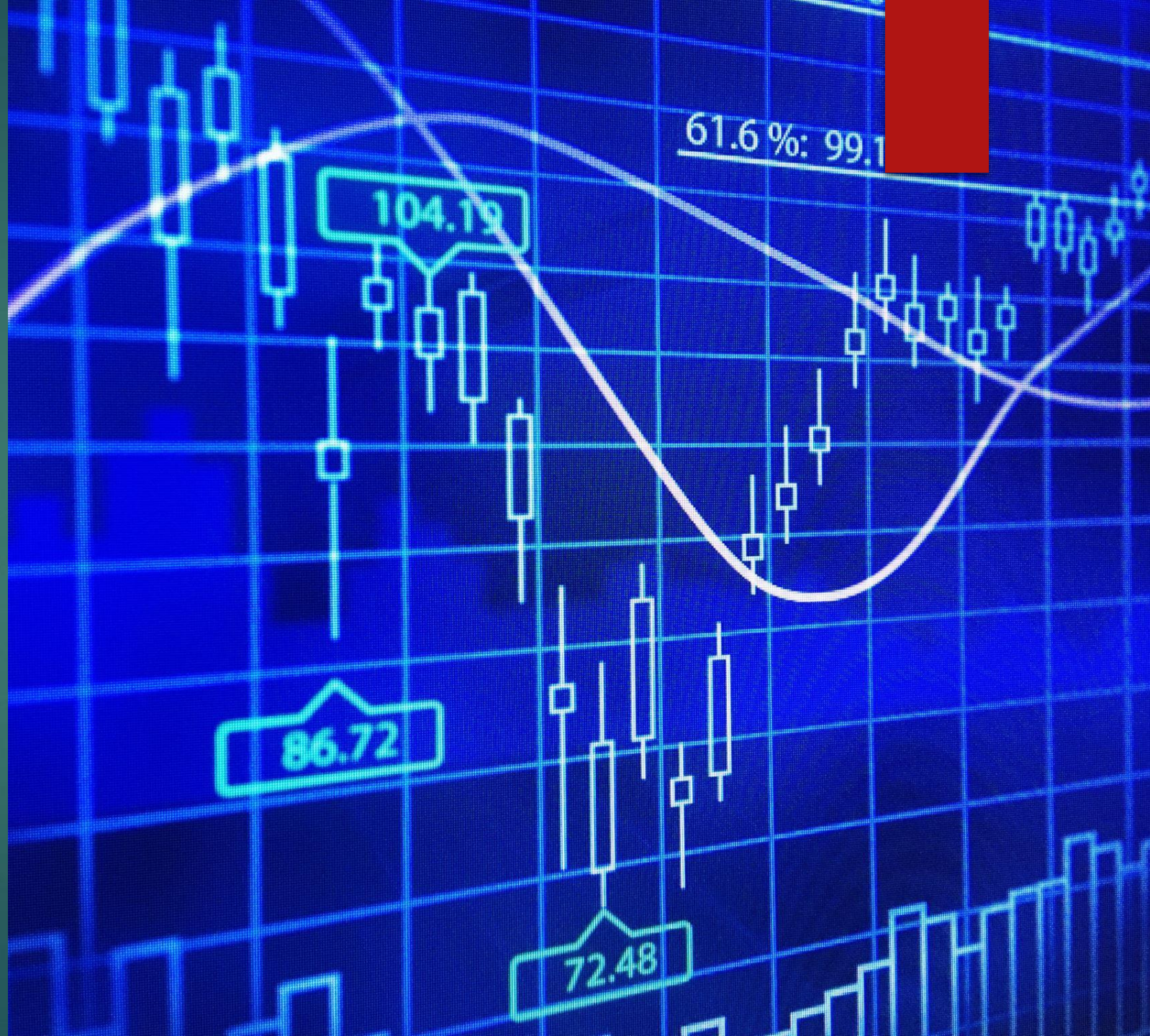
Accuracy: 71%

Confusion Matrix

```
[[1295  399]
 [ 540 1024]]
```

	precision	recall	f1-score	support
0	0.71	0.76	0.73	1694
1	0.72	0.65	0.69	1564
accuracy			0.71	3258
macro avg	0.71	0.71	0.71	3258
weighted avg	0.71	0.71	0.71	3258

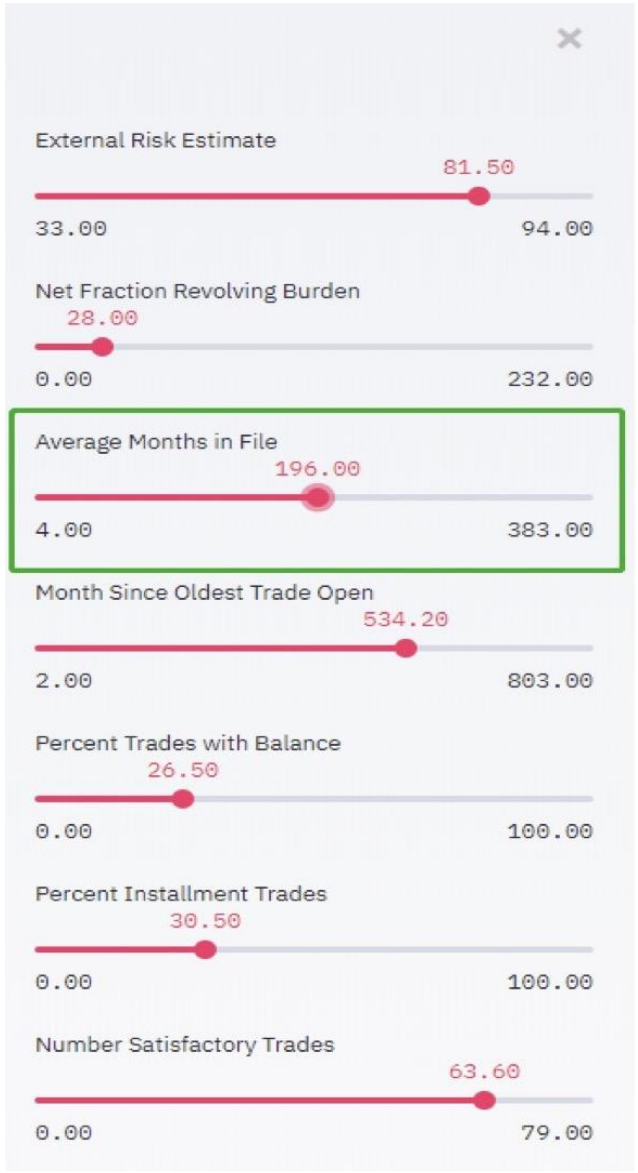
Interface



Importance Ranking

- importance ranking of all the feature.
- select the top seven features to demonstrate the model by using Random Forest.

	Imporatance
ExternalRiskEstimate	0.112755
NetFractionRevolvingBurden	0.075978
AverageMInFile	0.068570
MSinceOldestTradeOpen	0.068552
PercentTradesWBalance	0.060203
NumSatisfactoryTrades	0.055841
NumTotalTrades	0.050649
PercentInstallTrades	0.049474
PercentTradesNeverDelq	0.046529
NetFractionInstallBurden	0.044785
MSinceMostRecentTradeOpen	0.042940
NumRevolvingTradesWBalance	0.038512
MSinceMostRecentDelq	0.036841
NumBank2NatlTradesWHighUtilization	0.035212
MSinceMostRecentInqexcl7days	0.034130
NumTradesOpeninLast12M	0.028264
NumInqLast6M	0.026673
NumInstallTradesWBalance	0.026461
NumInqLast6Mexcl7days	0.023931
MaxDelq2PublicRecLast12M	0.023166
MaxDelqEver	0.022388
NumTrades60Ever2DerogPubRec	0.016121
NumTrades90Ever2DerogPubRec	0.012026



HELOC Data

Show dataframe

Choose a row of information in the dataset (0~1974):

10

test current sample

Which algorithm?

Random Forest

Prediction: Bad

Accuracy: 0.7179746835443038

Confusion Matrix:

	0	1
0	791	234
1	323	627

Result: 0

Random Forest Chosen

Bad Prediction

Good Prediction

HELOC Data

Show dataframe

Choose a row of information in the dataset (0~1974):

10

test current sample

Which algorithm?

Random Forest

Prediction: Good

Accuracy: 0.7179746835443038

Confusion Matrix:

	0	1
0	791	234
1	323	627

Result: 1

Random Forest Chosen

Summary

Thank you.