LOAN DEFAULT CLASSIFICATION

Manveer Sadhal
Oct 29, 2021

NEED

Identify loans in Interested Lending, Inc.'s current loan portfolio that are likely to default.

Use cases:

- Contact borrower
- Incorporate expected defaults into operating plan
- Identify loans to sell to another institution

DATA AND SCOPE

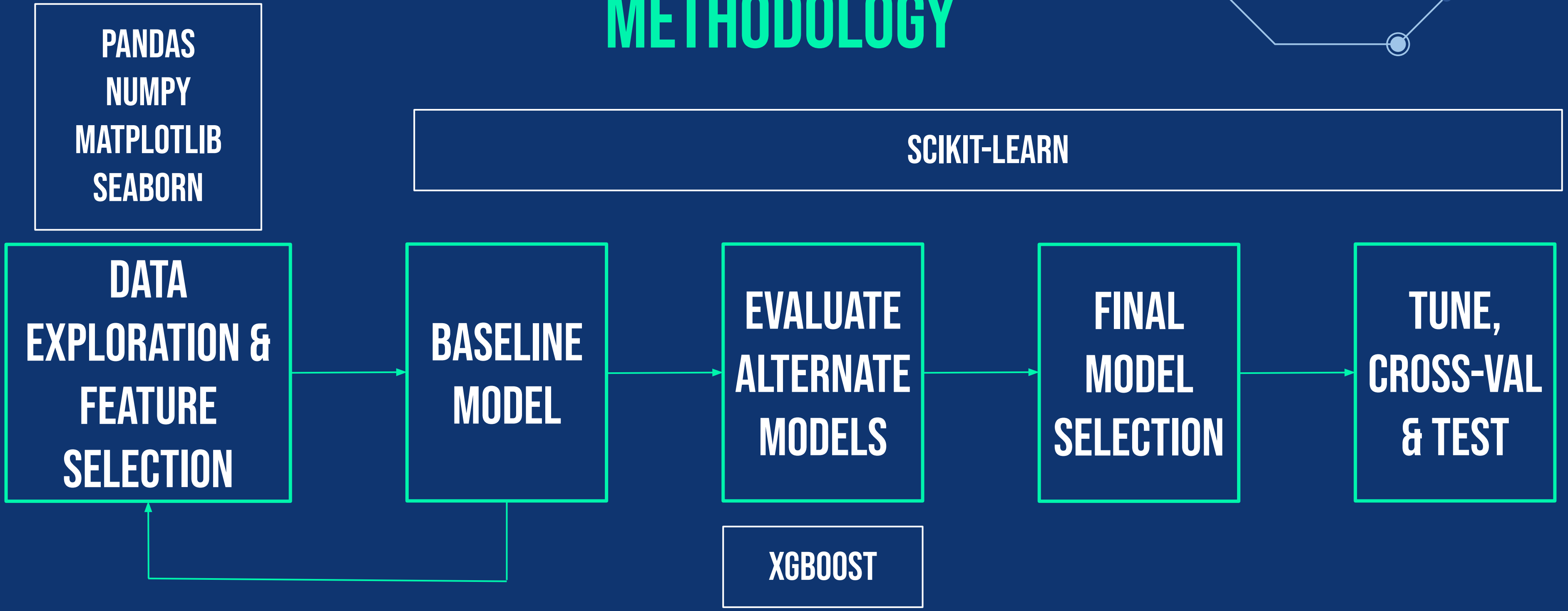
DATA

- Snapshot of 887,000+ individual loans.
- Filtered to approximately 250,000 loans.
- 73 columns reduced to 19 features for final model (e.g. annual income, interest rate)

SCOPE

- Current and closed loans
- Model developed with closed loans (fully paid, charged off, or defaulted).

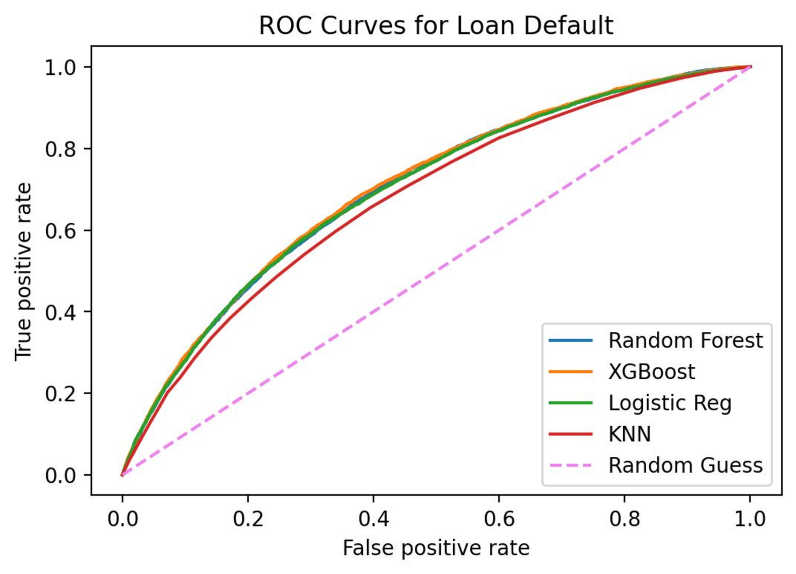
METHODOLOGY



RESULTS

Better

Better



- Logistic regression model, random forest, and XGB had nearly identical performance.
- Logistic regression selected - simpler, more interpretable.

RESULTS (LOGISTIC REGRESSION)

	Precision	Recall	F1
Paid	0.89	0.63	0.74
Default or Charge Off	0.29	0.66	0.40
Accuracy	0.63		
Macro Average	0.59	0.64	0.57
Weighted Average	0.78	0.63	0.68

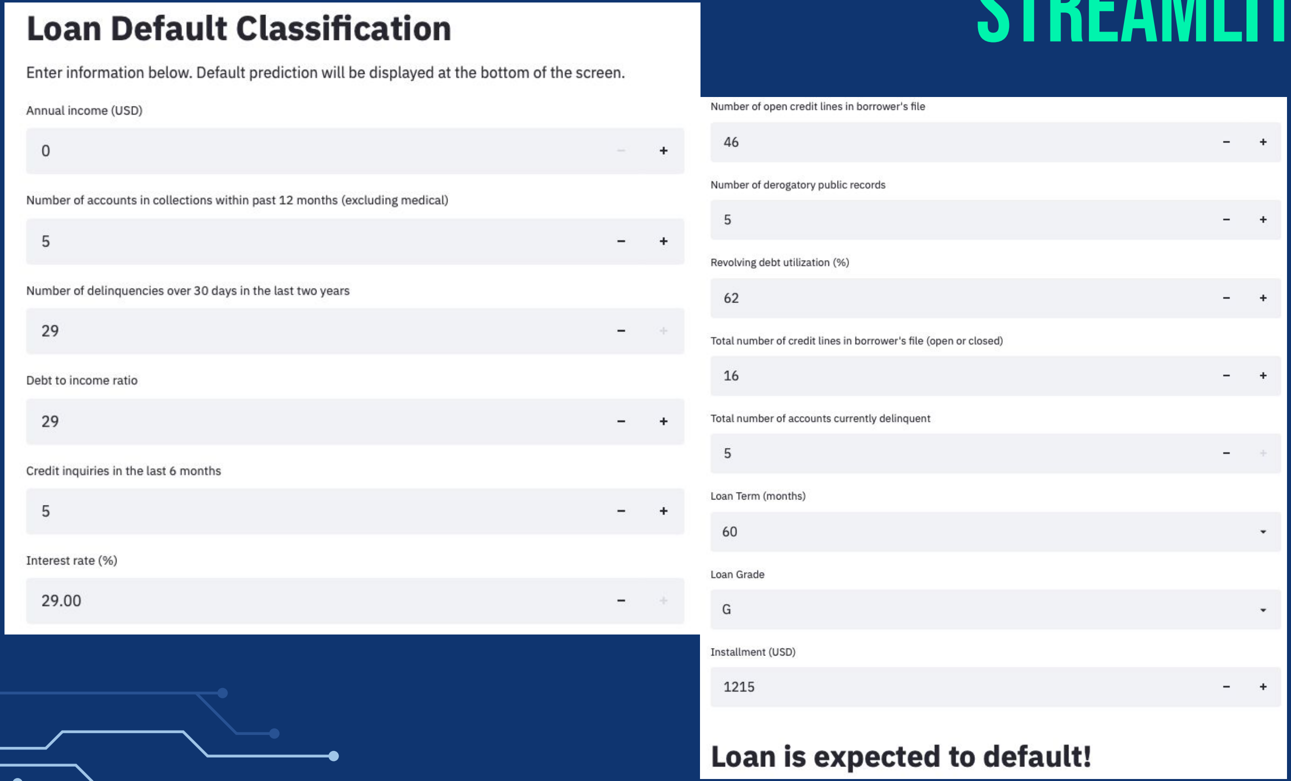
Tuned to maximize F2
Final score: 0.52

RESULTS

Greater impact

Feature	Coefficient
Interest Rate	0.469
Annual Income	-0.284
Total Number of Accounts	-0.187
Term of 60 months (vs. 36)	0.176
Debt to Income Ratio	0.174

INTERACTIVE STREAMLIT APP



CONCLUSIONS

- Model offers predictive capability for approved loans that may go into default
- Most impactful features can inform loan screening process
- Possible actions for loans likely to default:
  - Borrower outreach
  - Identify loans to sell to other institutions
  - Account for expected loan defaults in operating plan

FUTURE WORK

- Additional feature engineering
- Model stacking
- Develop separate model to evaluate whether high risk loans should be sold
- Add functionality to app to allow user to upload CSV with data on multiple loans rather than inputting one at a time.

THANKS!

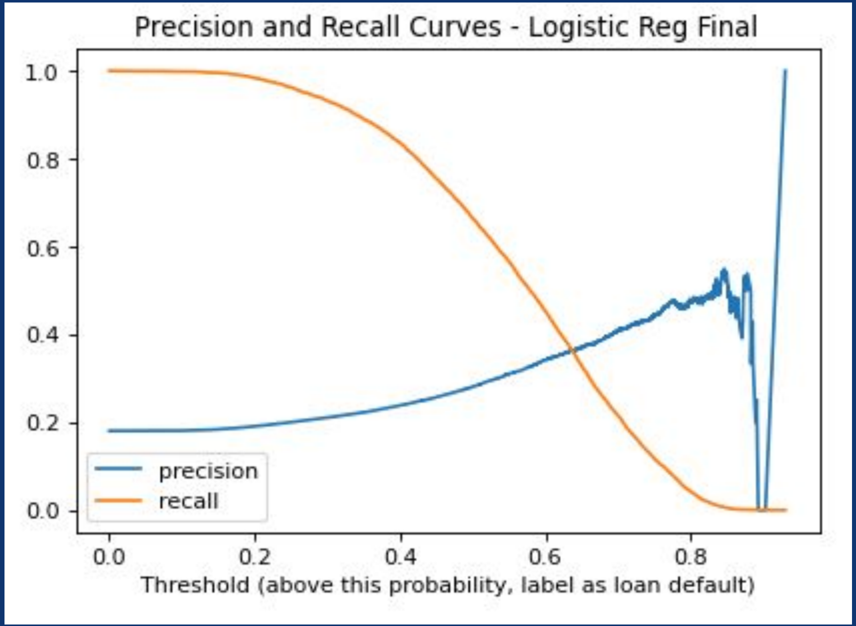
Do you have any questions?

addyouremail@freepik.com +91 620 421 838
yourcompany.com

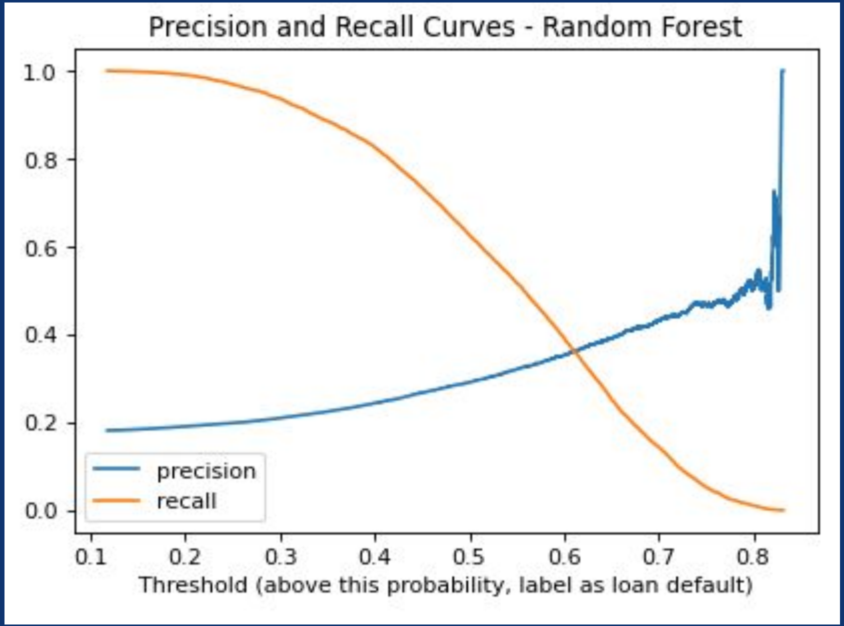
CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

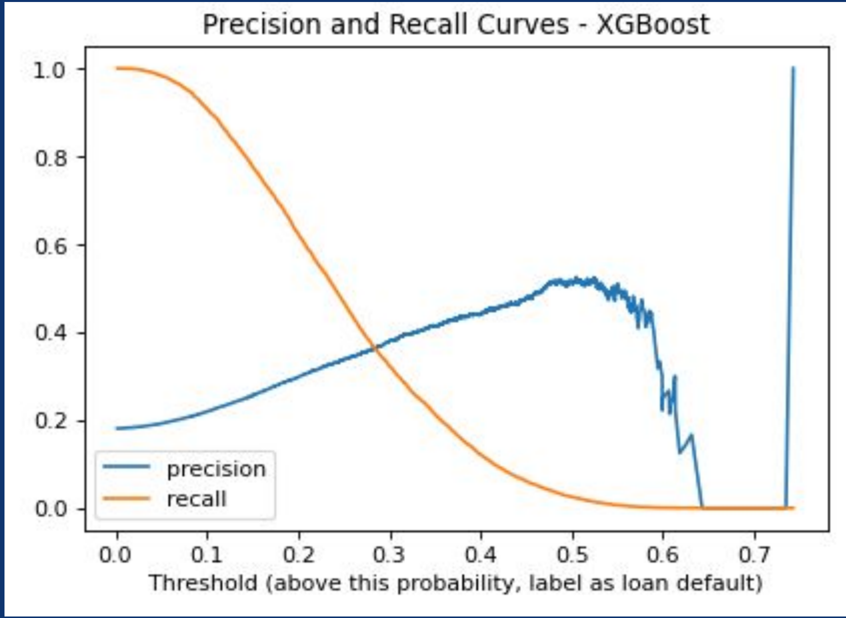
Please keep this slide for attribution

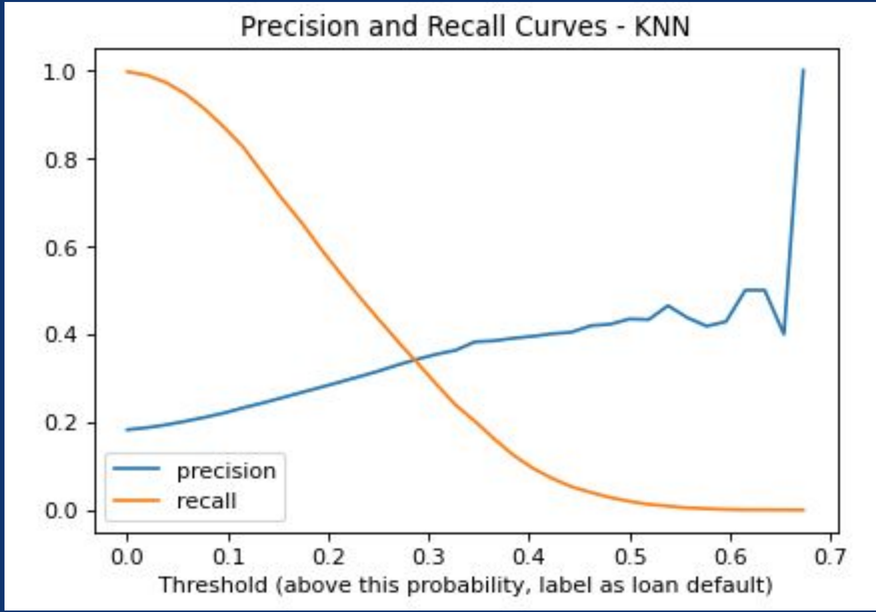
APPENDIX


Precision and Recall Curves - Logistic Reg Final
1.0
0.8
0.6
0.4
0.2
0.0
precision
recall
0.0
0.2
0.4
0.6
0.8
Threshold (above this probability, label as loan default)

```
'mean_test_score': array([0.52294971, 0.52292928, 0.52295788, 0.52292928, 0.52295248,
       0.52292928, 0.52293755, 0.52295   , 0.52294028, 0.52295   ,
       0.52294584, 0.52295   , 0.52301339, 0.52297616, 0.52300787,
       0.52297616, 0.52300787, 0.52297616, 0.52296372, 0.52298436,
       0.52296366, 0.52298436, 0.52298441, 0.52298436, 0.52296096,
       0.52298436, 0.52296366, 0.52298436, 0.5229817 , 0.52298436,
       0.52295276, 0.5229843 , 0.5229843 , 0.5229843 , 0.52297898,
       0.5229843 , 0.52298431, 0.5229843 , 0.52298154, 0.5229843 ,
       0.52297622, 0.5229843 , 0.52298154, 0.5229843 , 0.52298165,
       0.5229843 , 0.52297892, 0.5229843 , 0.52297898, 0.5229843 ,
       0.52297889, 0.5229843 , 0.52298707, 0.5229843 , 0.5229843 ,
       0.5229843 , 0.5229789 , 0.5229843 , 0.52298157, 0.5229843 ,
       0.52297616, 0.5229843 , 0.52297889, 0.5229843 , 0.5229555 ,
       0.5229843 , 0.52298433, 0.5229843 , 0.52298157, 0.5229843 ,
       0.52298983, 0.5229843 , 0.52297616, 0.5229843 , 0.52298166,
       0.5229843 , 0.52298431, 0.5229843 , 0.52298157, 0.5229843 ,
       0.52298431, 0.5229843 , 0.52297613, 0.5229843 ]),
'std_test_score': array([0.00465179, 0.00469768, 0.00464358, 0.00469768, 0.00465376,
       0.00469768, 0.00468125, 0.00471252, 0.00468019, 0.00471252,
       0.00468825, 0.00471252, 0.00472156, 0.00472239, 0.00471705,
       0.00472239, 0.00471705, 0.00472239, 0.00471271, 0.0047186 ,
       0.00470566, 0.0047186 , 0.00472521, 0.0047186 , 0.00471049,
       0.0047186 , 0.00470566, 0.0047186 , 0.00470202, 0.0047186 ,
       0.00471473, 0.00471161, 0.00471161, 0.00471161, 0.00473116,
       0.00471161, 0.00471159, 0.00471161, 0.00470943, 0.00471161,
       0.00472898, 0.00471161, 0.00470943, 0.00471161, 0.004723  ,
       0.00471161, 0.00472415, 0.00471161, 0.00473116, 0.00471161,
       0.00472127, 0.00471161, 0.00471377, 0.00471161, 0.00471161,
       0.00471161, 0.00472124, 0.00471161, 0.00471274, 0.00471161,
       0.00472242, 0.00471161, 0.00472127, 0.00471161, 0.00471359,
       0.00471161, 0.0047145 , 0.00471161, 0.00471274, 0.00471161,
       0.00471552, 0.00471161, 0.00472242, 0.00471161, 0.00472342,
       0.00471161, 0.00471159, 0.00471161, 0.00471274, 0.00471161,
       0.00471159, 0.00471161, 0.00471909, 0.00471161]),
```


Precision and Recall Curves - Random Forest
1.0
0.8
0.6
0.4
0.2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Threshold (above this probability, label as loan default)
precision
recall


Precision and Recall Curves - XGBoost
1.0
0.8
0.6
0.4
0.2
0.0
precision
recall
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Threshold (above this probability, label as loan default)


Precision and Recall Curves - KNN
1.0
0.8
0.6
0.4
0.2
0.0
precision
recall
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Threshold (above this probability, label as loan default)