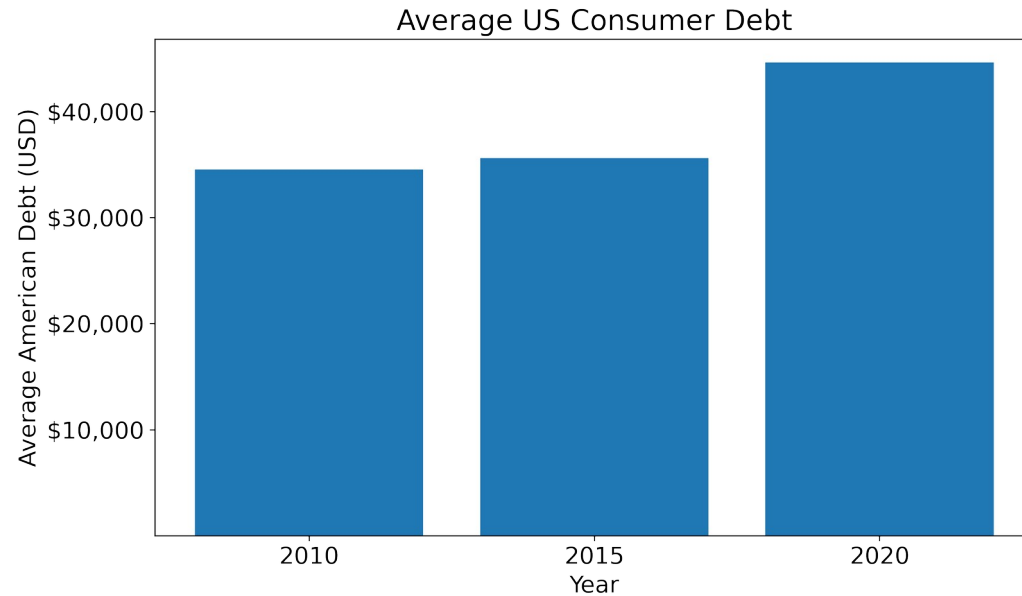


Responsible Lending

By John Lassetter
Oct. 29, 2021

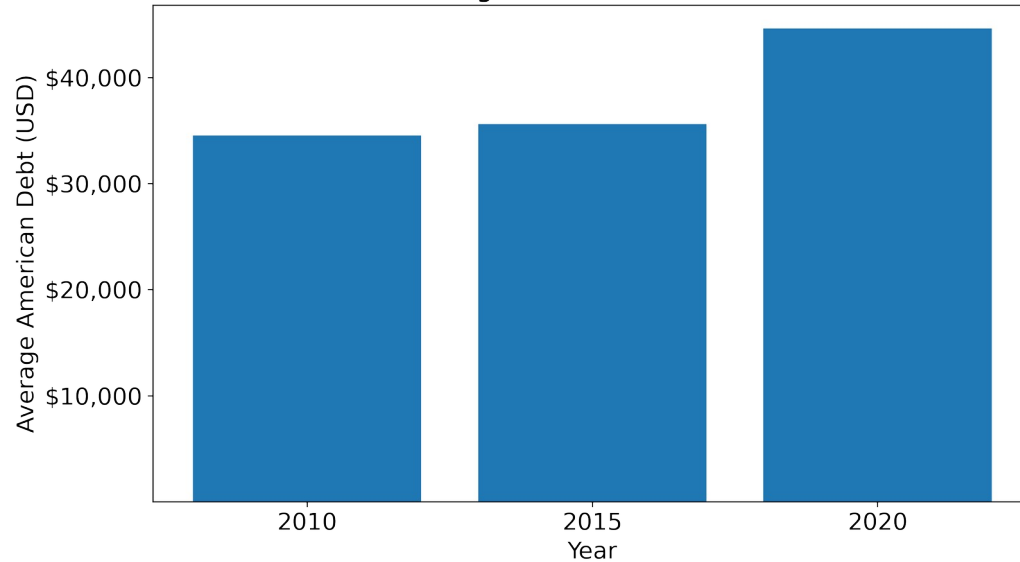
Background



Source: experian.com

Background

Average US Consumer Debt



Source: experian.com

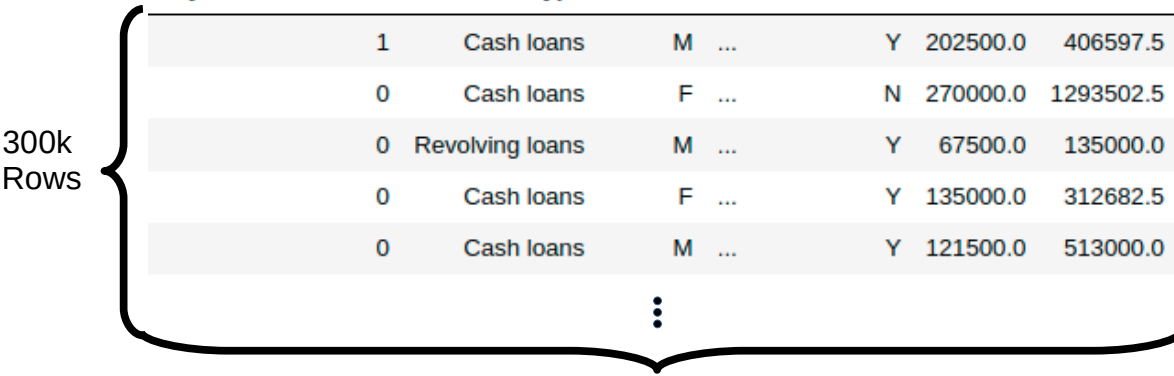
US leveraged loan default rate



Data through Dec. 31, 2020.

Sources: LCD, an offering of S&P Global Market Intelligence; S&P/LSTA Leveraged Loan Index

Data



Payment Difficulties	Contract Type	Gender	...	Owns Home	Income	Credit
1	Cash loans	M	...	Y	202500.0	406597.5
0	Cash loans	F	...	N	270000.0	1293502.5
0	Revolving loans	M	...	Y	67500.0	135000.0
0	Cash loans	F	...	Y	135000.0	312682.5
0	Cash loans	M	...	Y	121500.0	513000.0
⋮						

300k Rows

121 Features

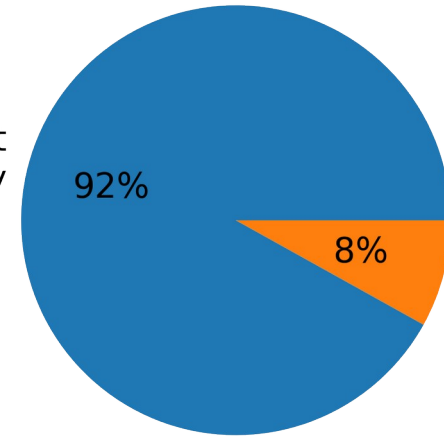
Data

300k Rows

Payment Difficulties	Contract Type	Gender	...	Owns Home	Income	Credit
1	Cash loans	M	...	Y	202500.0	406597.5
0	Cash loans	F	...	N	270000.0	1293502.5
0	Revolving loans	M	...	Y	67500.0	135000.0
0	Cash loans	F	...	Y	135000.0	312682.5
0	Cash loans	M	...	Y	121500.0	513000.0
⋮						

121 Features

No Payment
Difficulty



Payment
Difficulty

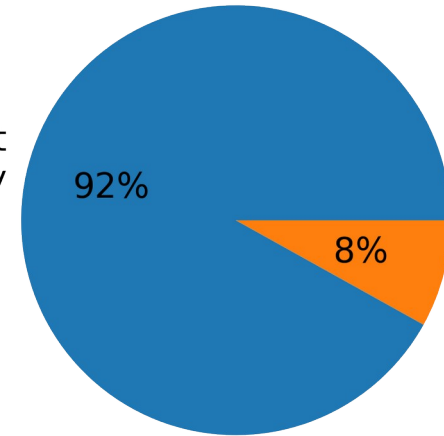
Data

300k Rows

Payment Difficulties	Contract Type	Gender	...	Owns Home	Income	Credit
1	Cash loans	M	...	Y	202500.0	406597.5
0	Cash loans	F	...	N	270000.0	1293502.5
0	Revolving loans	M	...	Y	67500.0	135000.0
0	Cash loans	F	...	Y	135000.0	312682.5
0	Cash loans	M	...	Y	121500.0	513000.0
⋮						

121 Features

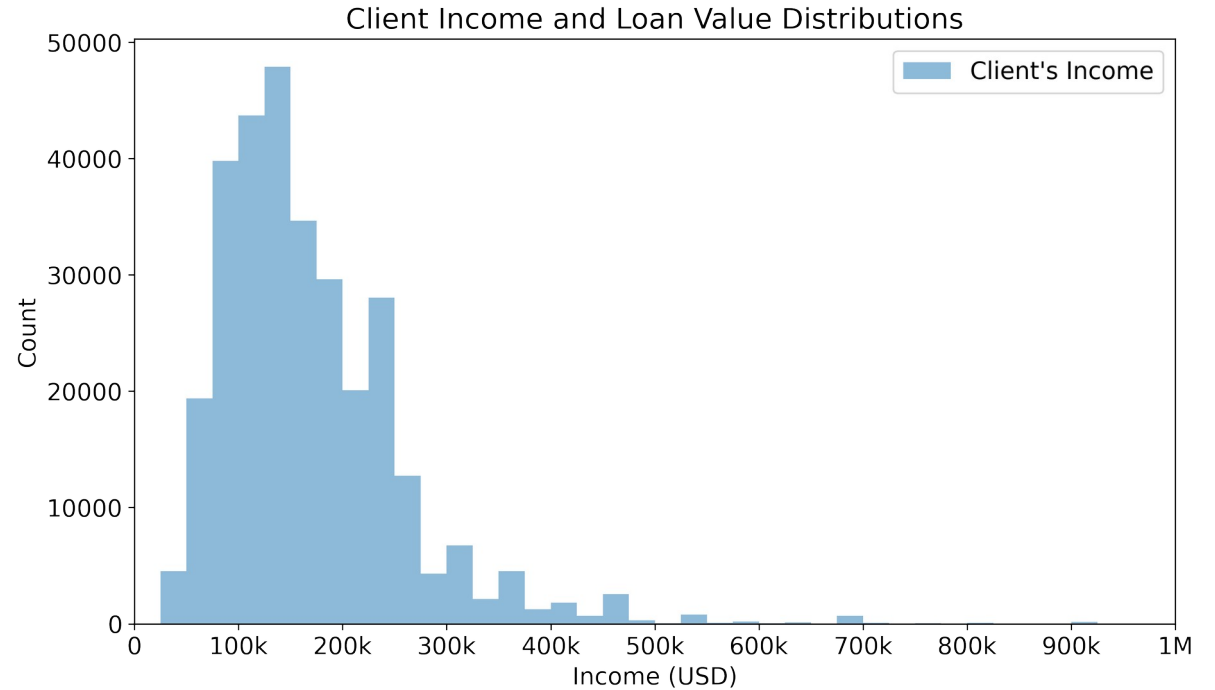
No Payment
Difficulty



Payment
Difficulty

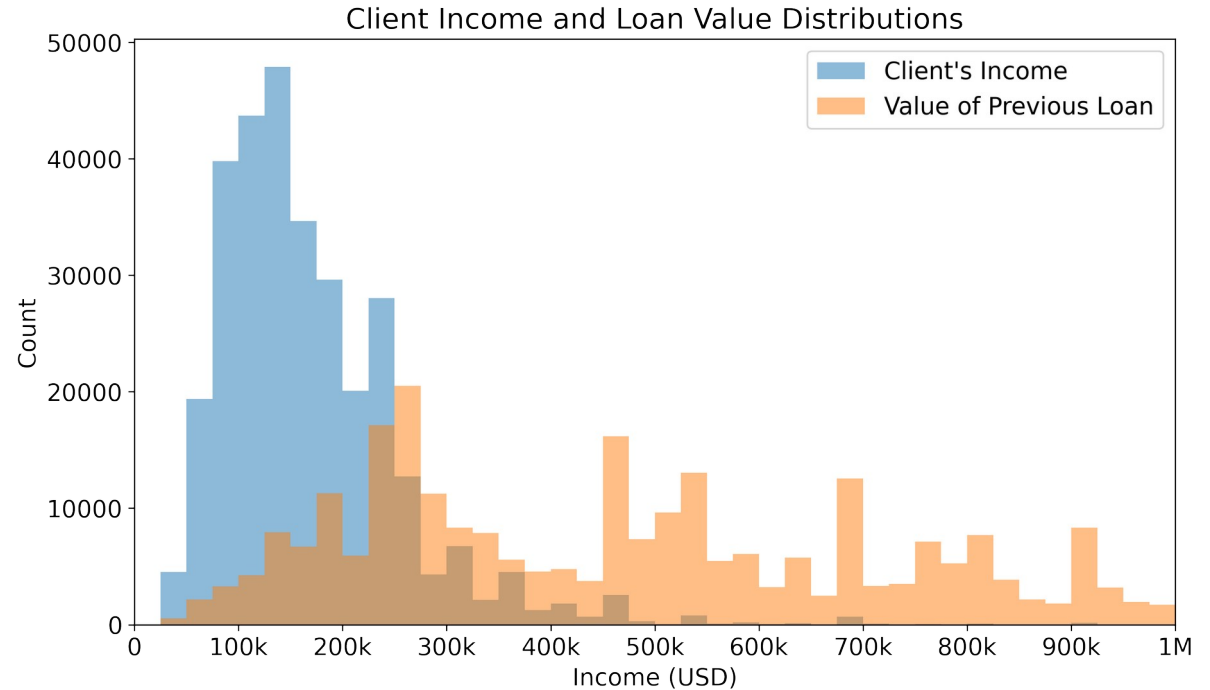
Data

- Wealthy Clients



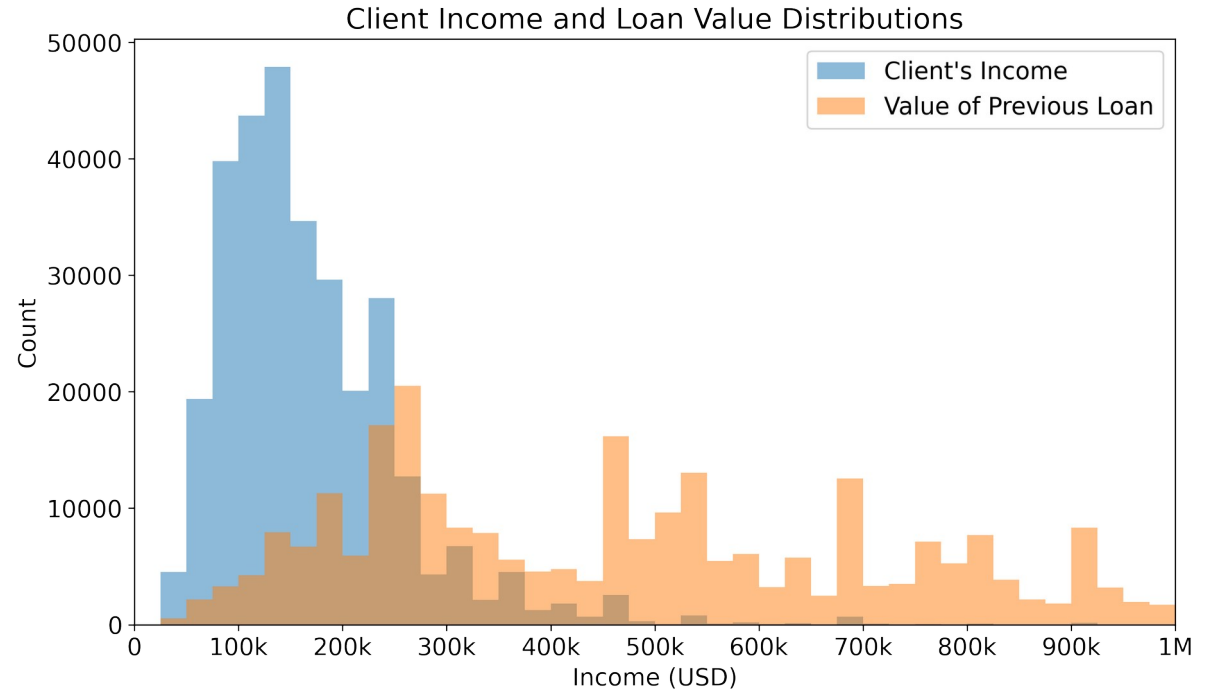
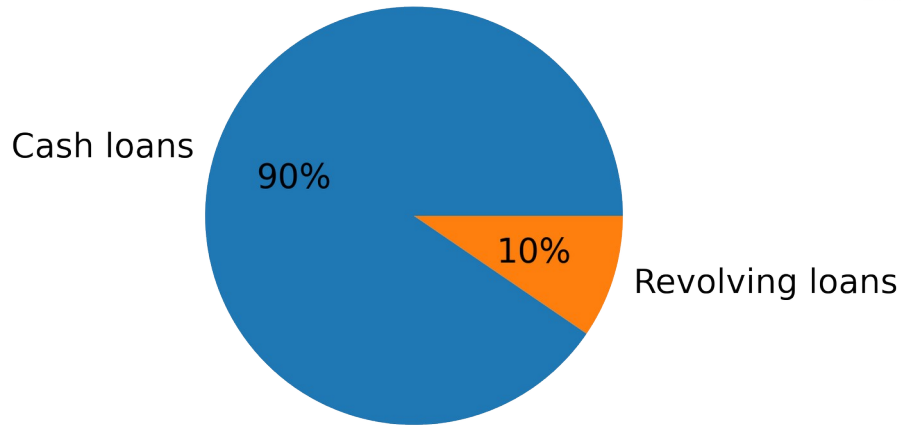
Data

- Wealthy Clients
- Loans range 100k - >1M

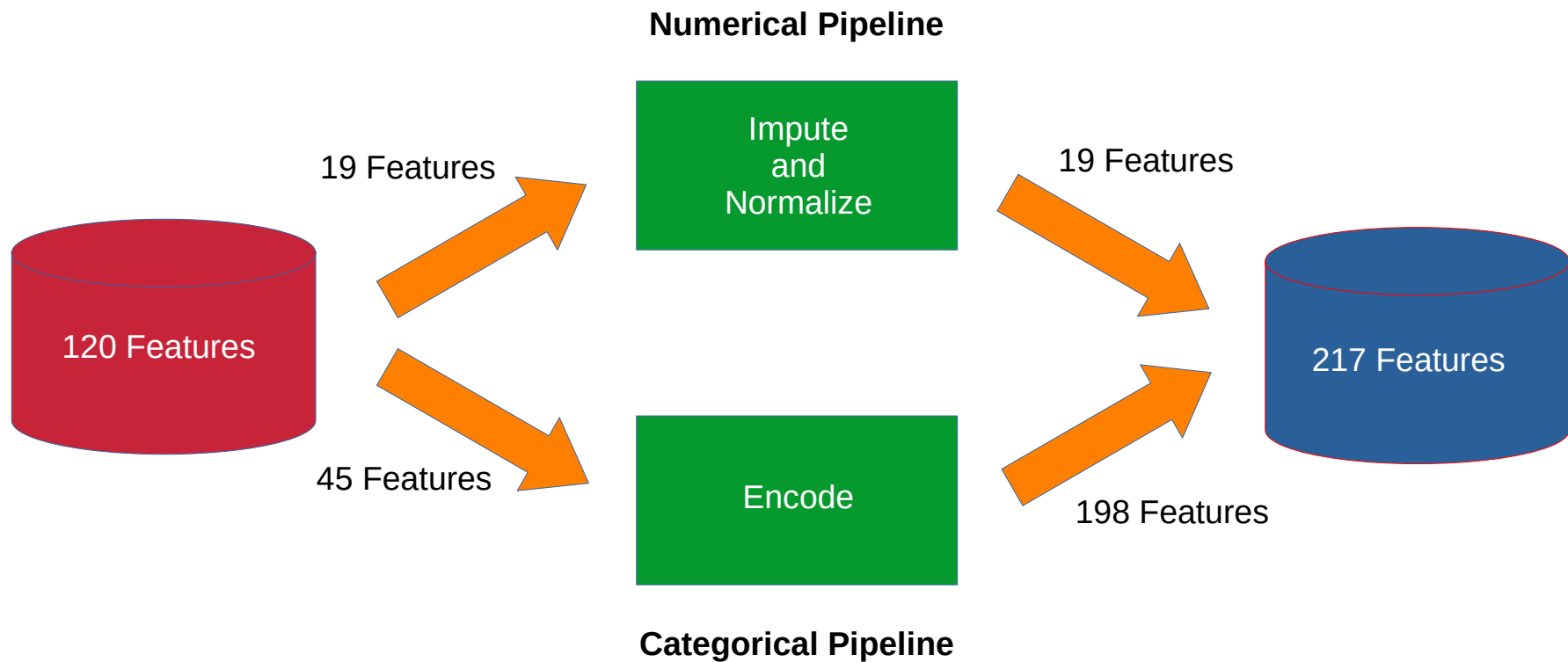


Data

- Wealthy Clients
- Loans range 100k - >1M
- Mostly Cash Loans



Pipeline



Model Comparison

Model
Logistic Regression
Random Forest
Gradient Boosted Forest

Model Comparison

Model	F1 Score
Logistic Regression	0.23
Random Forest	0.22
Gradient Boosted Forest	0.25

Model Comparison

Model	F1 Score	% of PD flagged
Logistic Regression	0.23	36%
Random Forest	0.22	43%
Gradient Boosted Forest	0.25	37%

Model Comparison

Model	F1 Score	% of PD flagged	$\frac{\text{True} +}{\text{False} +}$
Logistic Regression	0.23	36%	1:5
Random Forest	0.22	43%	1:5.7
Gradient Boosted Forest	0.25	37%	1:4.4

Important Features

Numerical	Categorical
Days in current job	Employment Industry
Age	Doc. 14
Price of Loan	Doc. 13
# Defaults in social circle	Region rating
Days since changing phone	Brought someone to apply for loan

Important Features

Takeaways:

(P = Prob. of P.D.)

- Steady job = $\downarrow P$
- Older = $\downarrow P$
- Bigger loan = $\downarrow P$
- More defaults in social circle = $\uparrow P$

Numerical	Categorical
Days in current job	Employment Industry
Age	Doc. 14
Price of Loan	Doc. 13
# Defaults in social circle	Region rating
Days since changing phone	Brought someone to apply for loan

Dashboard Prototype

Example Output

Higher Risk Loan (probability of PD = 0.223)

Top 5 reasons for denial and their normalized values:

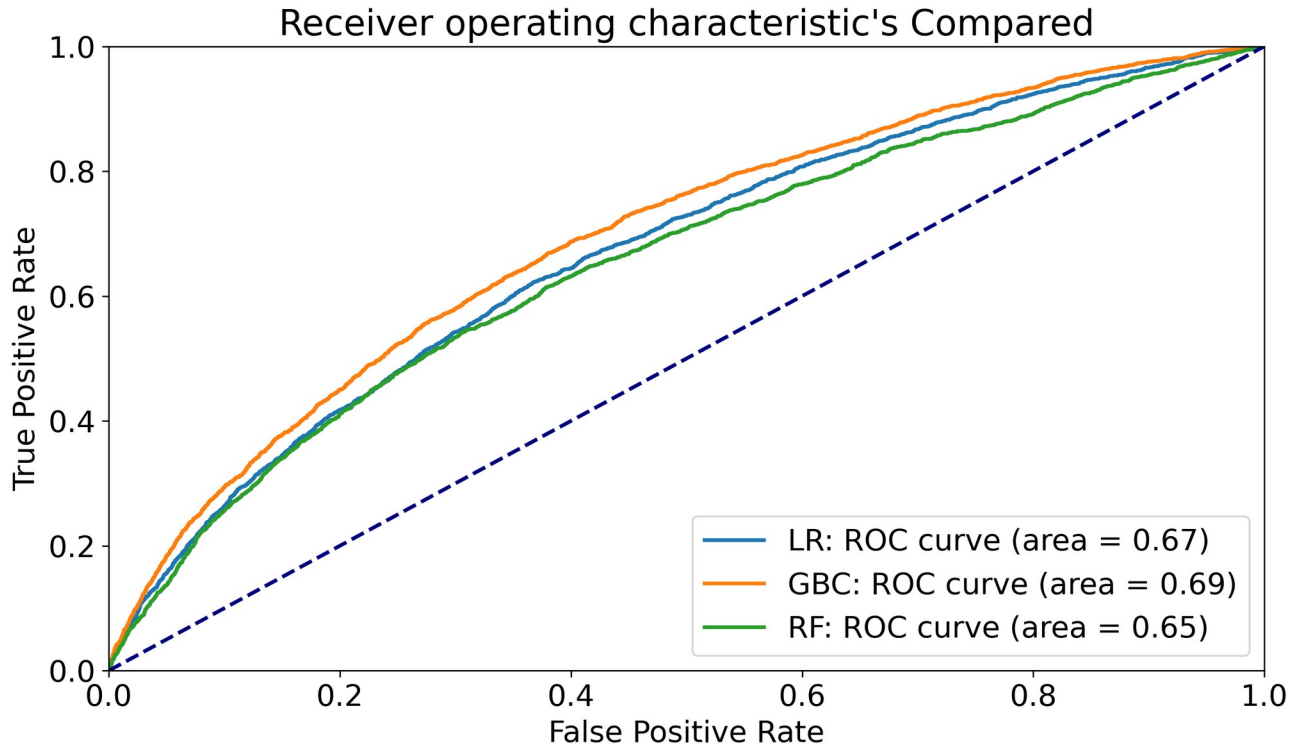
	normalized_feat_value
:	:
FLAG_DOCUMENT_2_0	1
FLAG_DOCUMENT_5_0	1
FLAG_DOCUMENT_6_0	1
FLAG_DOCUMENT_9_0	1
FLAG_DOCUMENT_7_0	1

Potential Improvements

- **More features**
 - Had payment difficulties or default previously
 - Reason for loan (e.g. debt consolidation)
- **More feature engineering**
 - Explore interaction terms more thoroughly
- **Implement Dashboard for Prediction**
 - Probability of payment difficulties
 - Top 5 reasons for decision

Questions

Appendix



Appendix

