Small business loan Approval Forecast

Abigail Deng

AGENDA

- Overview of Problem
- Overview of Datasets and approach
- Findings from EDA and Model
- Model comparison
- Product Demo

Overview of Problem

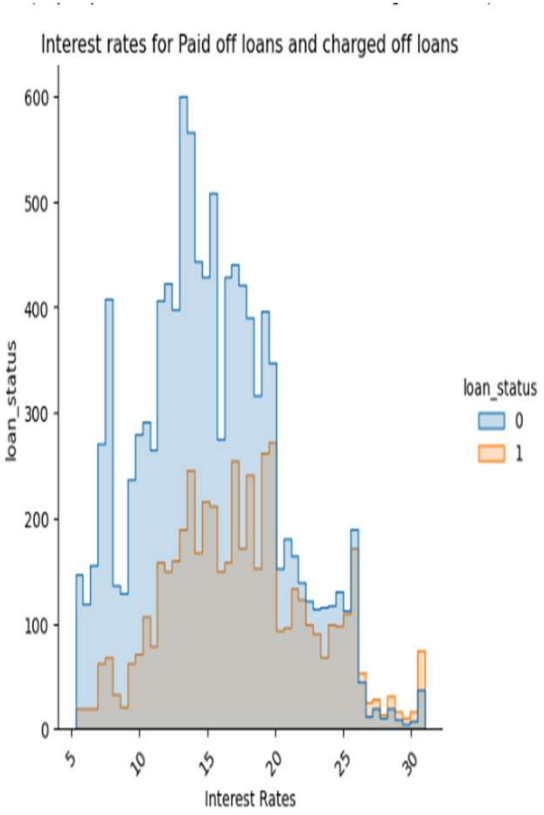
- Elevated interest rate climate
- Insufficient research on business loans
- Limited access to conventional commercial banks for small business owners

Overview of Datasets and approach

2007-2018 lending café data(more than 2 million rows and 151 columns)

Data preprocessing: reduce dataset to 16K rows and 21 columns

Use Machine Learning models to forecast the default



#	Column	Non-Null	Count	Dtype
0	loan_statu	16098	non-null	int64
1	term	16098	non-null	int64
2	int_rate	16098	non-null	float64
3	installmen	16098	non-null	float64
4	grade	16098	non-null	int64
5	home_own	16098	non-null	int64
6	annual_inc	16098	non-null	float64
7	loan_amnt	16098	non-null	float64
8	delinq_2yr	16098	non-null	float64
9	fico_range	16098	non-null	float64
10	fico_range	16098	non-null	float64
11	inq_last_6	16098	non-null	float64
12	open_acc	16098	non-null	float64
13	pub_rec	16098	non-null	float64
14	revol_bal	16098	non-null	float64
15	total_acc	16098	non-null	float64
16	out_prncp	16098	non-null	float64
17	out_prncp_	16098	non-null	float64
18	last_pymn	16098	non-null	float64
19	delinq_am	16098	non-null	float64

Findings from EDA and Model

Model comparison

Imbalanced Sample

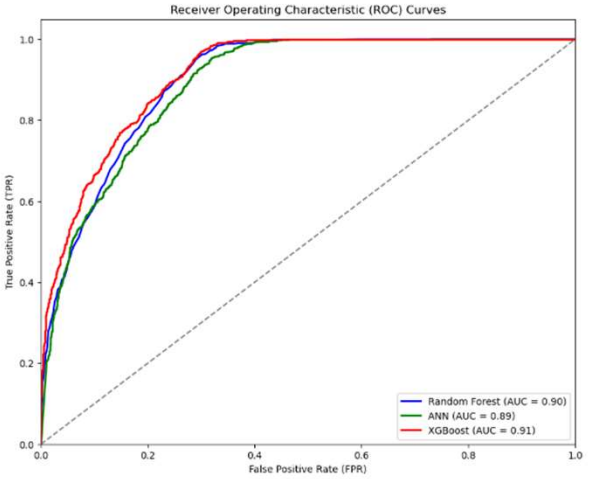
Algorithm/Metric	Accuracy	F1(Default)	AUC
Random Forest	0.81863354	0.71	0.91
ANN	0.6798	0.71	0.5
XgBoost	0.82670807	0.73	0.92

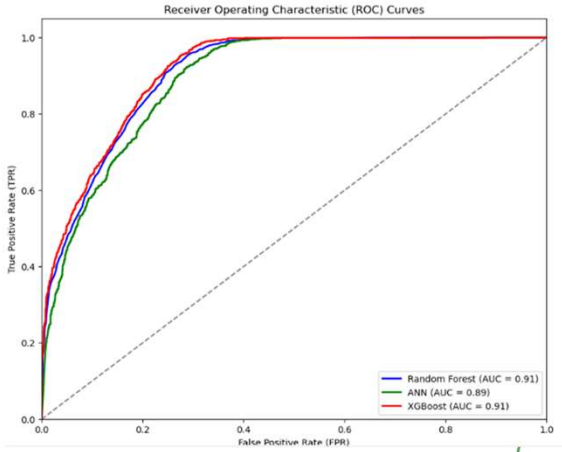
With SMOTE

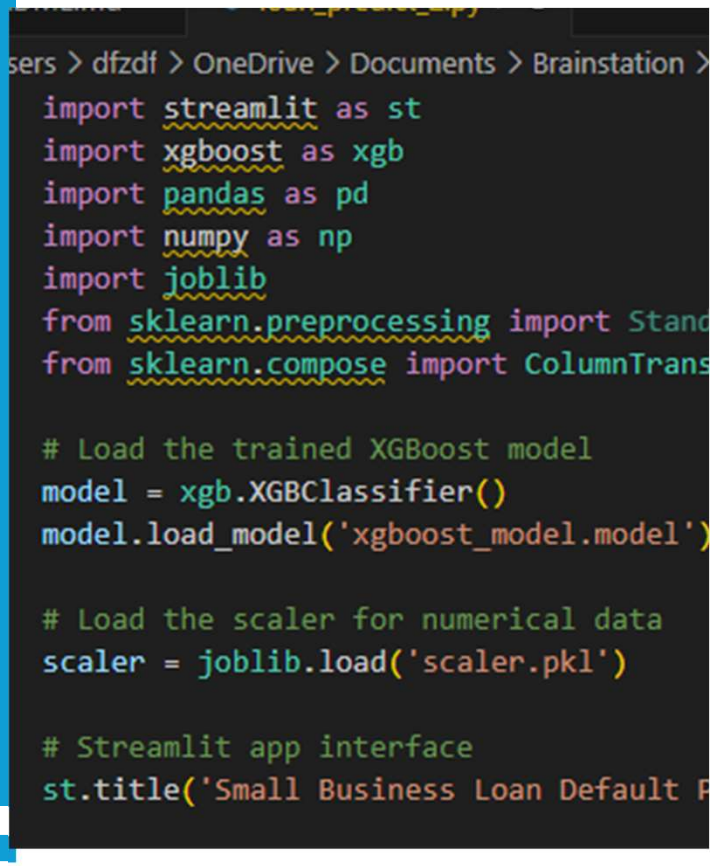
Algorithm/Metric	Accuracy	F1(Default)	AUC
Random Forest	0.80559006	0.72	0.9
ANN	0.77763975	0.73	0.89
XgBoost	0.81242237	0.74	0.91

With Random Sampling

Algorithm/Metric	Accuracy	F1(Default)	AUC
Random Forest	0.81273292	0.73	0.91
ANN	0.77950311	0.73	0.89
XgBoost	0.81055901	0.75	0.91







Product Demo

References

Lending Club dataset. Kaggle. https://www.kaggle.com/datasets/wordsforthewise/lending-club/code

Imane RHZIOUAL BERRADA, Fatimazahra BARRAMOU and Omar BACHIR ALAMI, “Towards a Machine Learning-based Model for Corporate Loan Default Prediction” International Journal of Advanced Computer Science and Applications(IJACSA), 15(3), 2024. http://dx.doi.org/10.14569/IJACSA.2024.0150357

Thank You