Bankruptcy Detection

Name: 陳怡仁　Group: 11

Motivation

When debt ratio = 0.95, but ROE = 0.2.

Data Description

Name: Taiwanese Bankruptcy Prediction

Time: 1999 to 2009

Instances: 6819

Features: 96

Class definition: By the business regulations of Taiwan Stock Exchange

Dataset web: UC Irvine Mechine Learning Repository

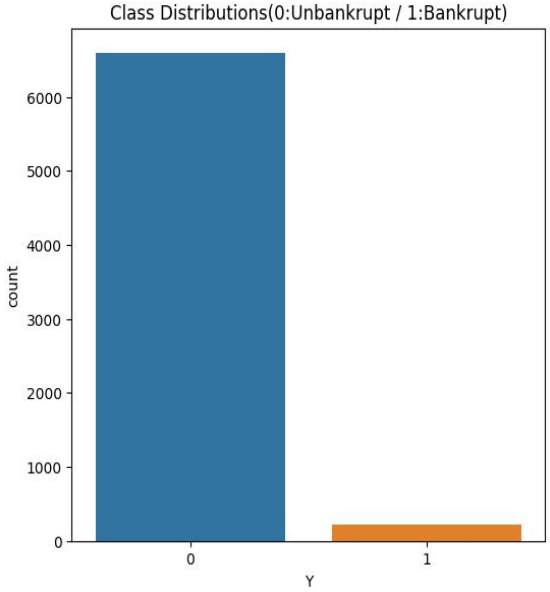
Class and Variables

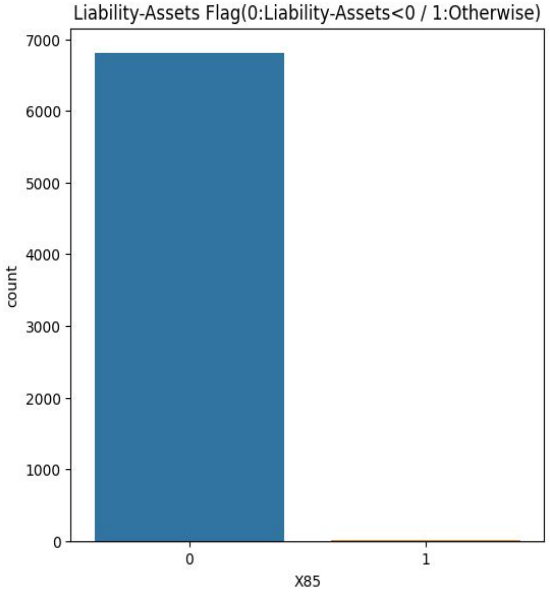
Y	Bankrupt?
X1	Cost of Interest-bearing Debt
X2	Cash Reinvestment Ratio
X3	Current Ratio
X4	Acid Test
X5	Interest Expenses/Total Revenue
X6	Total Liability/Equity Ratio
X7	Liability/Total Assets
X8	Interest-bearing Debt/Equity
X9	Contingent Liability/Equity
X10	Operating Income/Capital
X11	Pretax Income/Capital
X12	Working Capital to Total Assets
X13	Quick Assets/Total assets
X14	Current Assets/Total Assets
X15	Cash/Total Assets
X16	Quick Assets/Current Liability
X17	Cash/Current Liability
X18	Current Liability to Assets
X19	Operating Funds to Liability
X20	Inventory/Working Capital
X21	Inventory/Current Liability
X22	Current Liabilities/Liability
X23	Working Capital/Equity
X24	Current Liabilities/Equity
X25	Long-term Liability to Current Assets
X26	Current Liability to Current Assets
X27	One if Total Liability exceeds Total Assets;
X28	Equity to Liability
X29	Equity/Total Assets
X30	(Long-term Liability+Equity)/Fixed Assets
X31	Fixed Assets to Assets
X32	Current Liability to Liability
X33	Current Liability to Equity
X34	Equity to Long-term Liability
X35	Liability to Equity
X36	Degree of Financial Leverage
X37	Interest Coverage Ratio
X38	Operating Expenses/Net Sales
X39	(Research and Development Expenses)/Net Sales
X40	Effective Tax Rate
X41	Book Value Per Share(B)
X42	Book Value Per Share(A)
X43	Book Value Per Share(C)
X44	Cash Flow Per Share
X45	Sales Per Share
X46	Operating Income Per Share
X47	Sales Per Employee
X48	Operation Income Per Employee
X49	Fixed Assets Per Employee
X50	total assets to GNP price
X51	Return On Total Assets(C)
X52	Return On Total Assets(A)
X53	Return On Total Assets(B)
X54	Gross Profit /Net Sales
X55	Realized Gross Profit/Net Sales
X56	Operating Income /Net Sales
X57	Pre-Tax Income/Net Sales
X58	Net Income/Net Sales
X59	Net Non-operating Income Ratio
X60	Net Income-Exclude Disposal Gain or Loss/Net Sales
X61	EPS-Net Income
X62	Pretax Income Per Share
X63	Retained Earnings to Total Assets
X64	Total Income to Total Expenses
X65	Total Expenses to Assets
X66	Net Income to Total Assets
X67	Gross Profit to Sales
X68	Net Income to Stockholder's Equity
X69	One if Net Income is Negative for the Last Two Years; Zero Otherwise
X70	(Inventory +Accounts Receivables) /Equity
X71	Total Asset Turnover
X72	Accounts Receivable Turnover
X73	Days Receivable Outstanding
X74	Inventory Turnover
X75	Fixed Asset Turnover
X76	Equity Turnover
X77	Current Assets to Sales
X78	Quick Assets to Sales
X79	Working Capital to Sales
X80	Cash to Sales
X81	Cash Flow to Sales
X82	No-credit Interval
X83	Cash Flow from Operating/Current Liabilities
X84	Cash Flow to Total Assets
X85	Cash Flow to Liability
X86	CFO to Assets
X87	Cash Flow to Equity
X88	Realized Gross Profit Growth Rate
X89	Operating Income Growth
X90	Net Income Growth
X91	Continuing Operating Income after Tax Growth
X92	Net Income-Excluding Disposal Gain or Loss Growth
X93	Total Asset Growth
X94	Total Equity Growth
X95	Return on Total Asset Growth

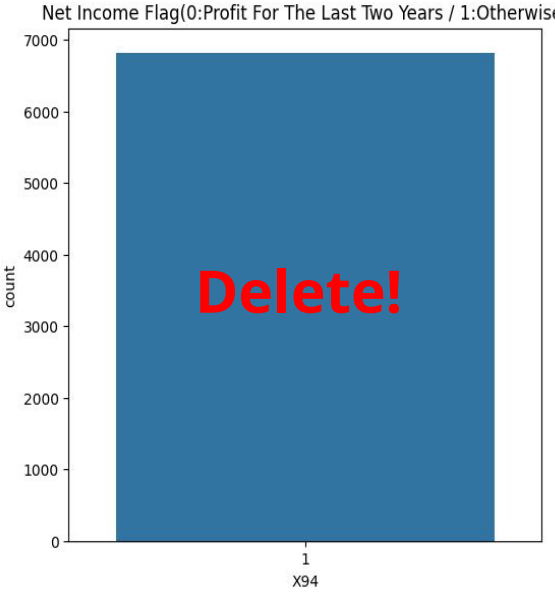
Data Visualization

	Bankrupt?	ROA(C) before interest and depreciation before interest	ROA(A) before interest and % after tax	ROA(B) before interest and depreciation after tax	Operating Gross Margin	Realized Sales Gross Margin	Operating Profit Rate	Pre-tax net Interest Rate	After-tax net Interest Rate	Non-industry income and expenditure/revenue	...	Net Income to Total Assets	Total assets to GNP price
0	1	0.370594	0.424389	0.405750	0.601457	0.601457	0.998969	0.796887	0.808809	0.302646	...	0.716845	0.009219
1	1	0.464291	0.538214	0.516730	0.610235	0.610235	0.998946	0.797380	0.809301	0.303556	...	0.795297	0.008323
2	1	0.426071	0.499019	0.472295	0.601450	0.601364	0.998857	0.796403	0.808388	0.302035	...	0.774670	0.040003
3	1	0.399844	0.451265	0.457733	0.583541	0.583541	0.998700	0.796967	0.808966	0.303350	...	0.739555	0.003252
4	1	0.465022	0.538432	0.522298	0.598783	0.598783	0.998973	0.797366	0.809304	0.303475	...	0.795016	0.003878
...	...	...	...	...	...	...	...	...	...	...	...	...	...
6814	0	0.493687	0.539468	0.543230	0.604455	0.604462	0.998992	0.797409	0.809331	0.303510	...	0.799927	0.000466
6815	0	0.475162	0.538269	0.524172	0.598308	0.598308	0.998992	0.797414	0.809327	0.303520	...	0.799748	0.001959
6816	0	0.472725	0.533744	0.520638	0.610444	0.610213	0.998984	0.797401	0.809317	0.303512	...	0.797778	0.002840
6817	0	0.506264	0.559911	0.554045	0.607850	0.607850	0.999074	0.797500	0.809399	0.303498	...	0.811808	0.002837
6818	0	0.493053	0.570105	0.549548	0.627409	0.627409	0.998080	0.801987	0.813800	0.313415	...	0.815956	0.000707

01Features Description





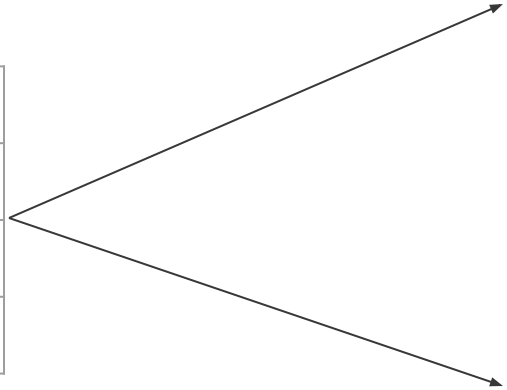


Experimental Scheme

Sample	X	y	ML Model	Measure	Result
Training Data: 80% Testing Data: 20%	X1~X95(delete X94)	0 / 1	Logistic Regression	Recall Rate	0.023
			XGBoost		0.045
			Random Forest		0.227
			LightGBM		0.341
			LightGBM With Bayesian Optimization		0.91

Train/Test Stratified Split

Dataset	
Class	Amount
0	6599
1	220



Train(80%) for X and y	
Class	Amount
0	5279
1	176

Test(20%) for X and y	
Class	Amount
0	1320
1	44

ML Model Result(1)

Logistic Regression				
	precision	recall	f1-score	support
0	0.968	0.999	0.984	1320
1	0.500	**0.023**	0.043	44
accuracy			0.968	1364
macro avg	0.651	0.511	0.513	1364
weighted avg	0.948	0.967	0.953	1364
parameter	random_state=0			

XGBoost				
	precision	recall	f1-score	support
0	0.969	1.000	0.984	1320
1	1.000	**0.045**	0.087	44
accuracy			0.969	1364
macro avg	0.985	0.523	0.536	1364
weighted avg	0.970	0.969	0.955	1364
parameters	eval_metric = "logloss", max_depth=5, learning_rate=0.01, n_estimators=100, gamma=0, min_child_weight=1, subsample=0.8, colsample_bytree=0.8, reg_alpha=0.005,seed = 0			

ML Model Result(2)

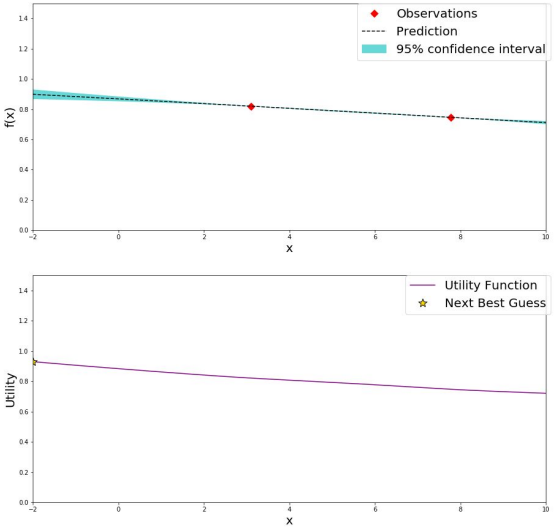
Random Forest				
	precision	recall	f1-score	support
0	0.975	0.994	0.984	1320
1	0.556	**0.227**	0.323	44
accuracy			0.969	1364
macro avg	0.765	0.611	0.653	1364
weighted avg	0.961	0.969	0.963	1364
parameters	n_estimators = 50, max_depth = 50, n_jobs = -1, random_state = 0			

LightGBM				
	precision	recall	f1-score	support
0	0.978	0.994	0.986	1320
1	0.652	**0.341**	0.448	44
accuracy			0.973	1364
macro avg	0.815	0.667	0.717	1364
weighted avg	0.968	0.973	0.969	1364
parameter	random_state=0			

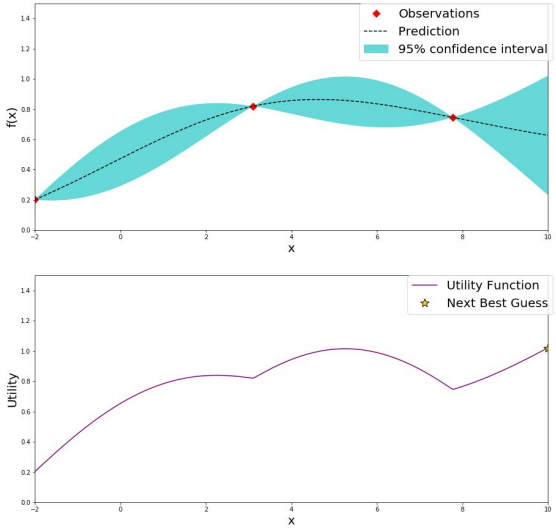
Bayesian Optimization Introduction(1)

(Initializing the model)

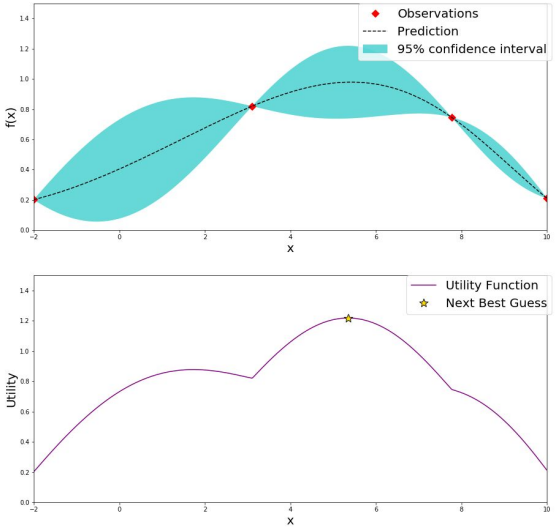
Gaussian Process and Utility Function After 2 Steps



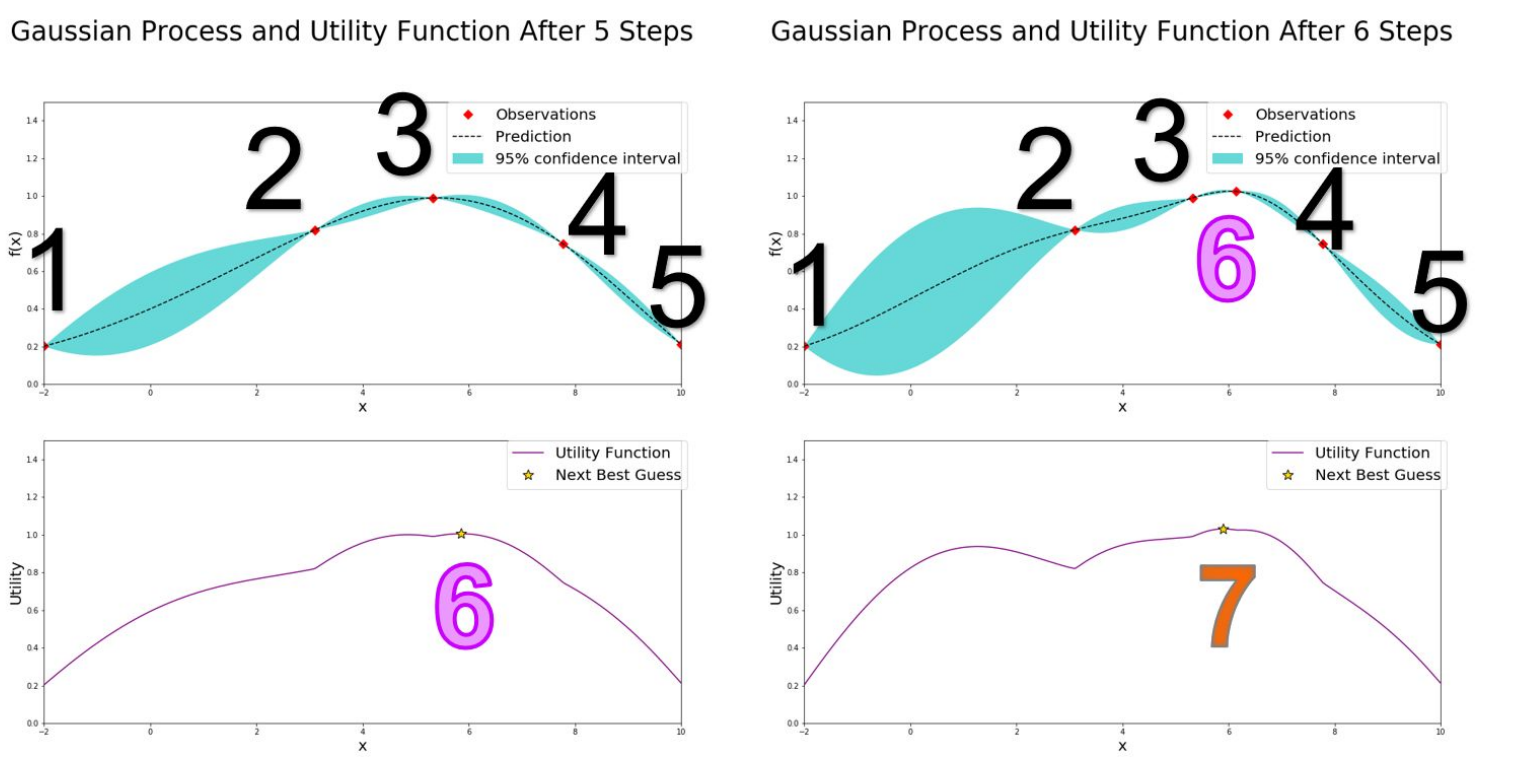
Gaussian Process and Utility Function After 3 Steps



Gaussian Process and Utility Function After 4 Steps



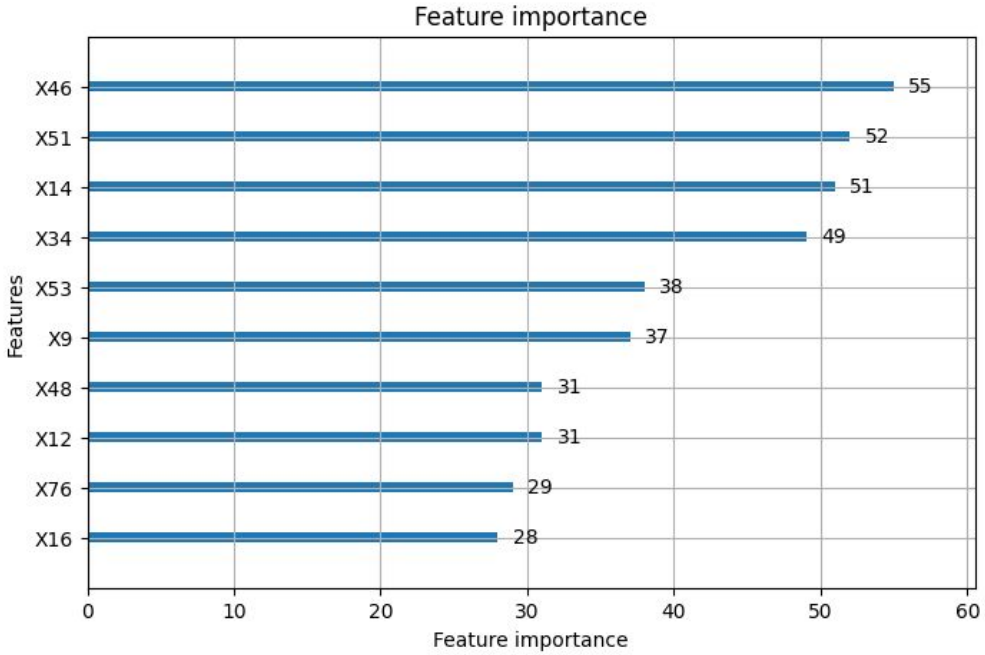
Bayesian Optimization Introduction(2)



The Best Parameters For lightGBM

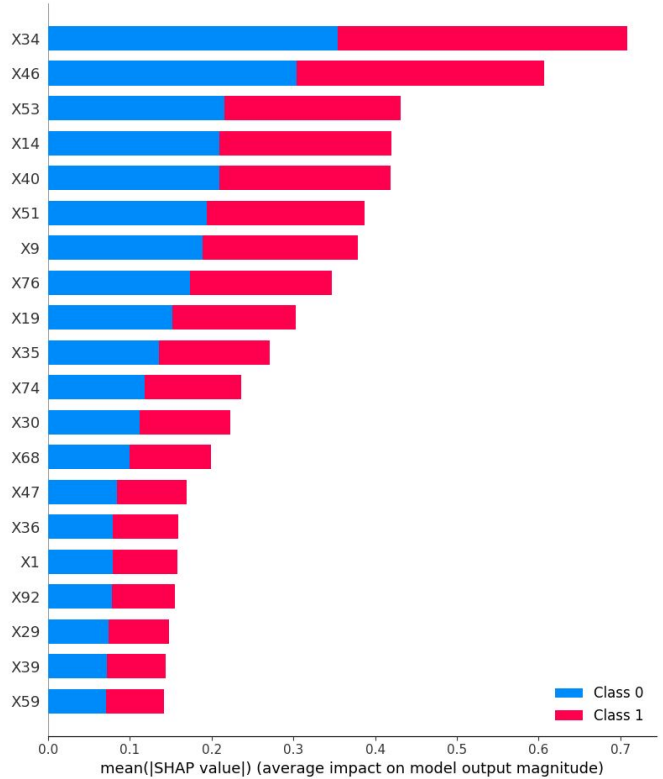
Iter	Target	colsample _bytree	lambda_l1	lambda_l2	learning _rate	max_depth	min_child _weight	n_estimators	num _leaves	subsample
1	**0.9418**	**0.7043**	**3.602**	**0.0003431**	**0.09768**	**11.6**	**6.432**	**133.8**	**47.64**	**0.7397**
2	0.9401	0.764	2.096	2.056	0.06929	44.52	3.315	351.7	53.38	0.7559
3	0.9366	0.5688	0.9905	2.402	0.2908	19.1	35.23	444.4	91.57	0.7085
. . .										. . .
10	0.9395	0.8912	4.809	2.004	0.03557	16.91	8.162	136.8	55.41	0.7847

Feature Importance with Adjusted LightGBM



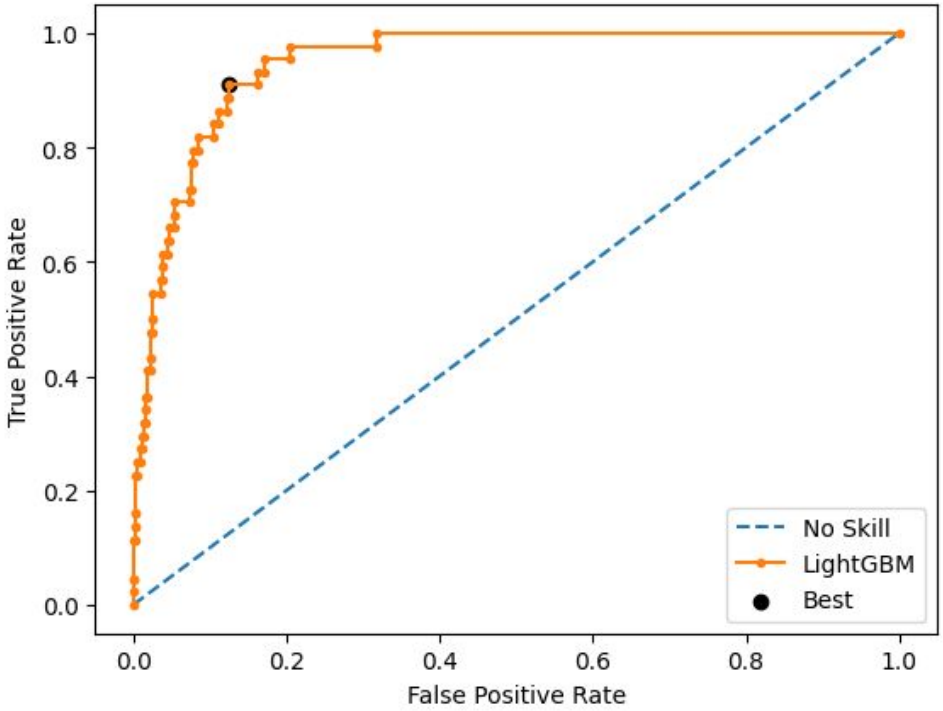
- Feature Importance is just a relative indicator rather than an absolute percentage.
- These features contribute 28%~55% to the model's predictions.
- From the macro perspective, X46 (Accounts Receivable Turnover) has the most influence.

SHAP Value With Adjusted LightGBM



- The SHAP value indicates the relative contribution of a particular feature to a single prediction.
- If these features are substitued into the model, it will let the MSE rise/reduce 0.07~0.35.
- From the micro perspective, X34 (Quick Ratio) has the most influence.

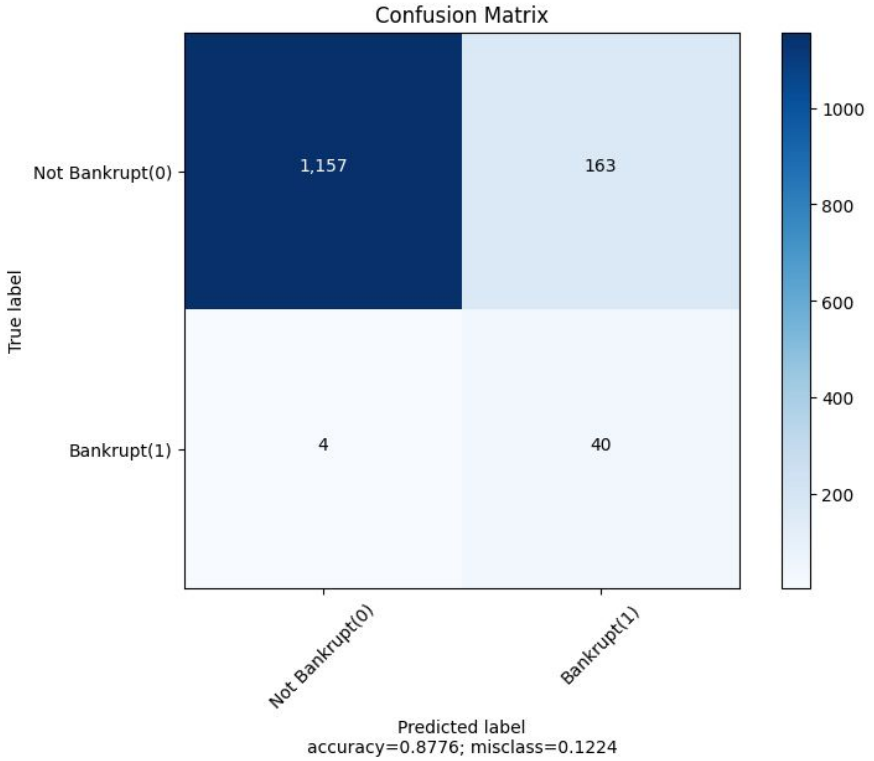
ROC Curve With Adjusted LightGBM



Best Threshold=0.021575

- On best threshold point, true positive rate and false positive rate can achive optimal balance.
- The probability value of the model output >= 0.021575, positive(1)
  < 0.021575, negetive(0)

Confusion Matrix With Adjusted LightGBM



$$G-Mean=\sqrt{\frac{40}{40+4}\ \text{x}\ \frac{163}{1157+163}}=0.893$$

Adjusted LightGBM				
	precision	recall	f1-score	support
0	1.00	0.88	0.93	1320
1	0.20	**0.91**	0.32	44
accuracy			0.88	1364
macro avg	0.60	0.89	0.63	1364
weighted avg	0.97	0.88	0.91	1364
parameters	colsample_bytree : 0.7043, lambda_l1 : 3.602, lambda_l2 : 0.0003431, learning_rate : 0.09768, max_depth : 12, min_child_weight : 6.432, n_estimators : 134, num_leaves : 48, subsample : 0.74			

Reference

https://blog.csdn.net/Leon_winter/article/details/86604553

https://www.kaggle.com/code/pauljkk/bankruptcyclassifier-using-lightgbm-91-recall/edit

https://archive.ics.uci.edu/dataset/572/taiwanese+bankruptcy+prediction

Thanks everyone

It’s Q&A time!