

Credit Scoring Report

(HMEQ)

Rows: 5,960, Columns: 13, Target: BAD (1 = bad, 0 = good)

Exploratory Data Analysis / Phân tích khám phá dữ liệu

Dataset head

BAD	LOAN	MORTDUE	VALUE	REASON	JOB	YOJ	DEROG	DELINQ	CLAGE	NINQ	CLNO	
1	1100	25860.0	39025.0	HomeImp	Other	10.5	0.0	0.0	94.366667	1.0	9.0	
1	1300	70053.0	68400.0	HomeImp	Other	7.0	0.0	2.0	121.833333	0.0	14.0	
1	1500	13500.0	16700.0	HomeImp	Other	4.0	0.0	0.0	149.466667	1.0	10.0	
1	1500	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
0	1700	97800.0	112000.0	HomeImp	Office	3.0	0.0	0.0	93.333333	0.0	14.0	
1	1700	30548.0	40320.0	HomeImp	Other	9.0	0.0	0.0	101.466002	1.0	8.0	37
1	1800	48649.0	57037.0	HomeImp	Other	5.0	3.0	2.0	77.100000	1.0	17.0	
1	1800	28502.0	43034.0	HomeImp	Other	11.0	0.0	0.0	88.766030	0.0	8.0	36
1	2000	32700.0	46740.0	HomeImp	Other	3.0	0.0	2.0	216.933333	1.0	12.0	
1	2000	NaN	62250.0	HomeImp	Sales	16.0	0.0	0.0	115.800000	0.0	13.0	

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5960 entries, 0 to 5959
Data columns (total 13 columns):
#   Column      Non-Null Count  Dtype
---  -
0   BAD          5960 non-null   int64
1   LOAN         5960 non-null   int64
2   MORTDUE      5442 non-null   float64
3   VALUE        5848 non-null   float64
4   REASON       5708 non-null   object
5   JOB          5681 non-null   object
6   YOJ          5445 non-null   float64
7   DEROG        5252 non-null   float64
8   DELINQ       5380 non-null   float64
9   CLAGE        5652 non-null   float64
10  NINQ         5450 non-null   float64
11  CLNO         5738 non-null   float64
12  DEBTINC      4693 non-null   float64
dtypes: float64(9), int64(2), object(2)
memory usage: 605.4+ KB
```

Descriptive Statistics / Thống kê mô t

	count	unique	top	freq	mean	std	min	25%
BAD	5960.0	NaN	NaN	NaN	0.199497	0.399656	0.0	0.0
LOAN	5960.0	NaN	NaN	NaN	18607.969799	11207.480417	1100.0	11100.0
MORTDUE	5442.0	NaN	NaN	NaN	73760.8172	44457.609458	2063.0	46276.0
VALUE	5848.0	NaN	NaN	NaN	101776.048741	57385.775334	8000.0	66075.5
REASON	5708	2	DebtCon	3928	NaN	NaN	NaN	NaN
JOB	5681	6	Other	2388	NaN	NaN	NaN	NaN
YOJ	5445.0	NaN	NaN	NaN	8.922268	7.573982	0.0	3.0
DEROG	5252.0	NaN	NaN	NaN	0.25457	0.846047	0.0	0.0
DELINQ	5380.0	NaN	NaN	NaN	0.449442	1.127266	0.0	0.0
CLAGE	5652.0	NaN	NaN	NaN	179.766275	85.810092	0.0	115.116702
NINQ	5450.0	NaN	NaN	NaN	1.186055	1.728675	0.0	0.0
CLNO	5738.0	NaN	NaN	NaN	21.296096	10.138933	0.0	15.0
DEBTINC	4693.0	NaN	NaN	NaN	33.779915	8.601746	0.524499	29.140031

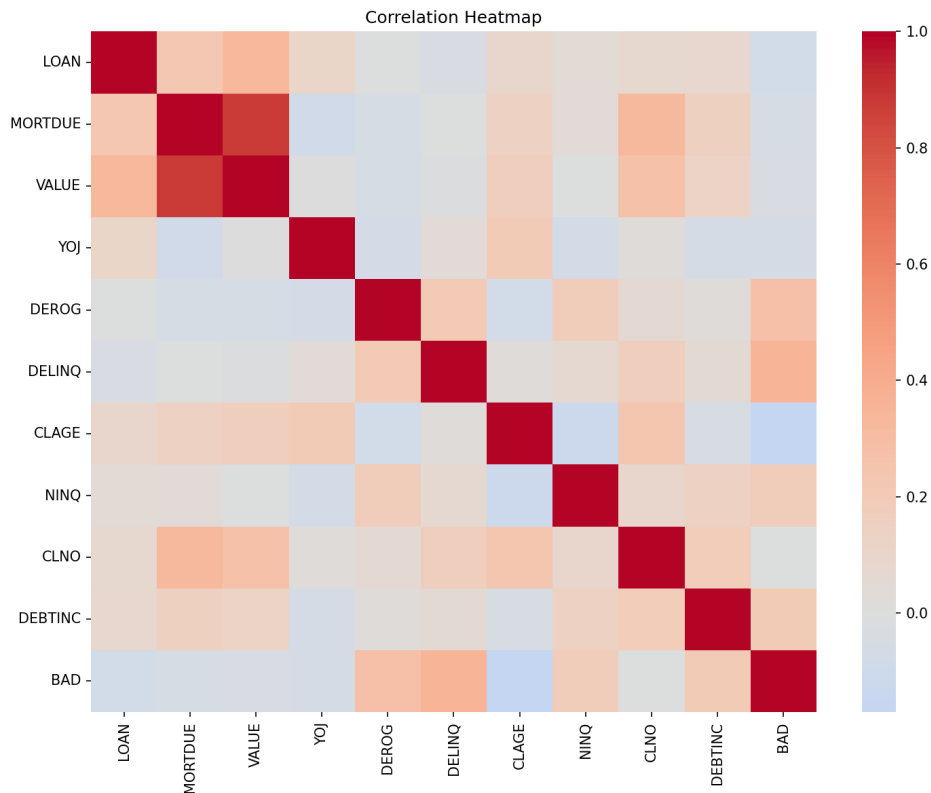
Target distribution

BAD	
0	4771
1	1189

Missing Values / Giá trị khuyết

DEBTINC	0.212584
DEROG	0.118792
DELINQ	0.097315
MORTDUE	0.086913
YOJ	0.086409
NINQ	0.085570
CLAGE	0.051678
JOB	0.046812
REASON	0.042282
CLNO	0.037248
VALUE	0.018792
LOAN	0.000000
BAD	0.000000

Correlation & Multicollinearity / T ng quan & Đa c ng tuy n



Variable	VIF
VALUE	12.152187
MORTDUE	11.128686
DEBTINC	9.068683
CLNO	6.741402
CLAGE	5.722485
LOAN	4.214961
YOJ	2.552094
NINQ	1.587066
DELINQ	1.214113
DEROG	1.143633

WoE & IV + Feature Selection / WoE & IV + Ch n bi n

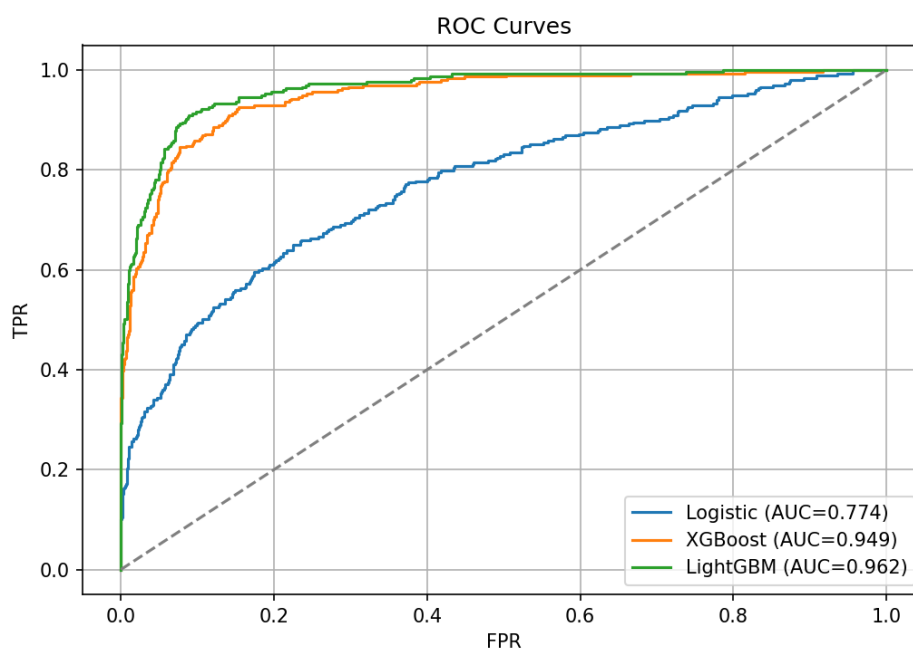
Variable	IV
DEBTINC	1.877128
VALUE	0.470688
DELINQ	0.452361
DEROG	0.276585
CLAGE	0.230638
NINQ	0.165557
LOAN	0.160155
JOB	0.123728

YOJ 0.083533
CLNO 0.079103
MORTDUE 0.048920
REASON 0.008618

Selected features: DEBTINC, VALUE, DELINQ, DEROG,
CLAGE, NINQ, LOAN, JOB

	Coef.	Std.Err.	z	P> z	[0.025	0.975]	CI95_low
const	-1.374353	0.046131	-29.792394	4.901588e-195	-1.464768	-1.283937	-1.464768
DEBTINC_WOE	-0.929462	0.031224	-29.767908	1.017120e-194	-0.990659	-0.868265	-0.990659
VALUE_WOE	-0.980790	0.083000	-11.816726	3.199095e-32	-1.143467	-0.818112	-1.143467
DELINQ_WOE	-0.979279	0.067809	-14.441822	2.822796e-47	-1.112181	-0.846376	-1.112181
DEROG_WOE	-0.791717	0.084455	-9.374393	6.957488e-21	-0.957246	-0.626187	-0.957246
CLAGE_WOE	-1.129374	0.094778	-11.916033	9.764832e-33	-1.315135	-0.943613	-1.315135
NINQ_WOE	-0.588538	0.106634	-5.519250	3.404504e-08	-0.797536	-0.379540	-0.797536
LOAN_WOE	-0.346030	0.111754	-3.096361	1.959116e-03	-0.565063	-0.126997	-0.565063
JOB_WOE	-0.940376	0.132242	-7.111000	1.152050e-12	-1.199567	-0.681186	-1.199567

Model Training + Cross Validation / Hu n luy n mô hình + Cross Validation



Model	AUC_Train	AUC_Test	CV_Mean	KS
Logistic	0.820363	0.773767	0.804198	0.425188
XGBoost	0.986464	0.948572	0.954572	0.772531
LightGBM	0.999650	0.962339	0.966603	0.820239