PREVENTING CREDIT CARD DEFAULT

Quant Crew

OUR TEAM

Connor
Layne

Patrick
Sweeney

Ananya
Lal

Junzhi
Meng

Soumil
Zachariah

THE PROBLEM

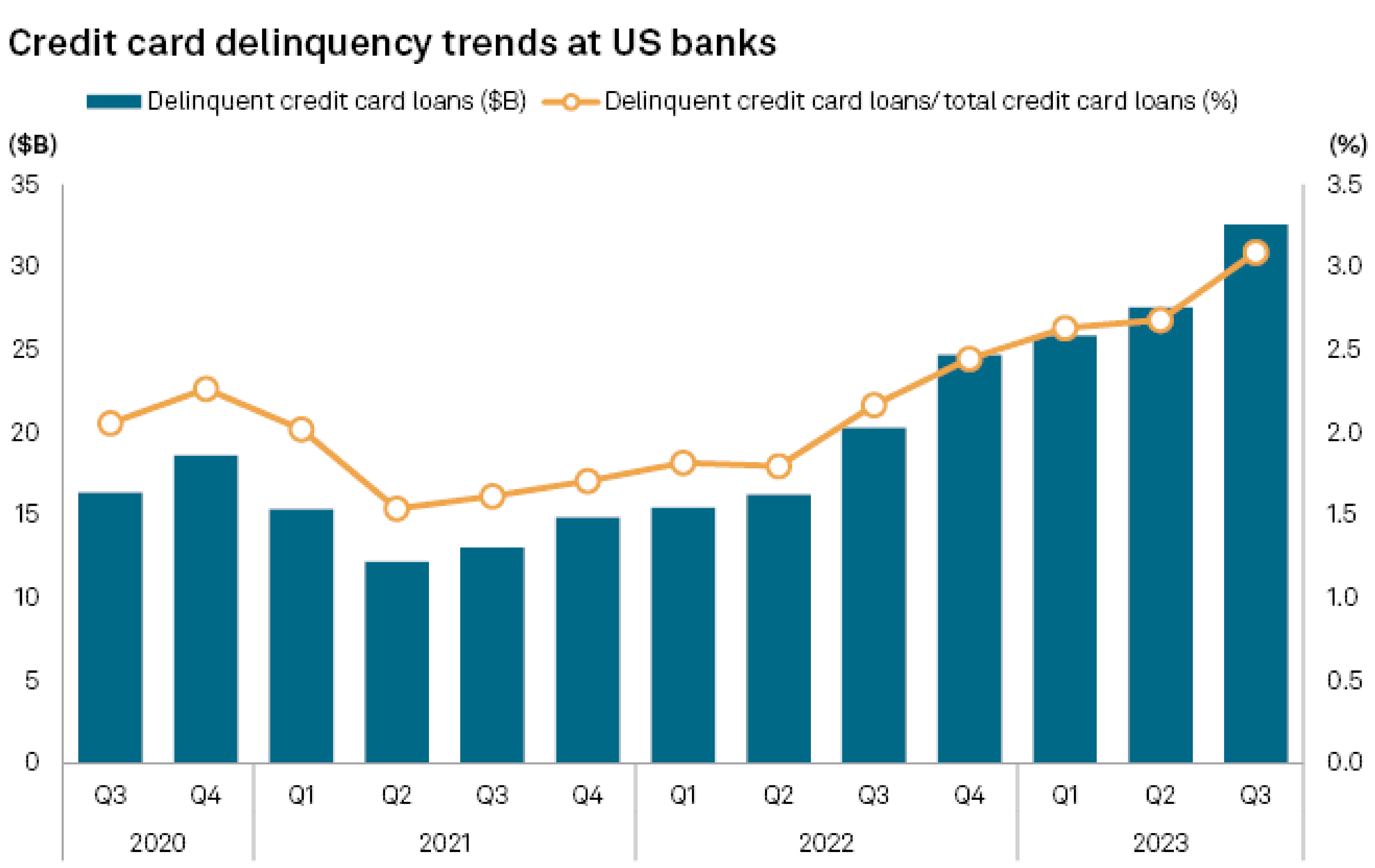
- Banks want to identify high-risk customers to prevent defaults.
- Proactive strategies are essential to minimize default risks.
- We can leverage bank data to predict and understand default likelihood and key variables.

UNDERSTANDING THE IMPACT

“Trust, But Verify”

Operational Efficiency

Relationships


Credit card delinquency trends at US banks
Delinquent credit card loans ($B)
Delinquent credit card loans/ total credit card loans (%)
($B)
(%)
35
30
25
20
15
10
5
0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Q3
Q4
Q1
Q2
Q3
Q4
Q1
Q2
Q3
Q4
Q1
Q2
Q3
2020
2021
2022
2023

DATA OVERVIEW

Demographics
Age, Marriage, Education, Gender

Monthly Payment and Billing Information
Amount Paid, Bill Amount, Repayment Status

Outcome
Whether or not the customer defaulted

Data consists of 30,000 customers from First Central Bank

MODEL SELECTION

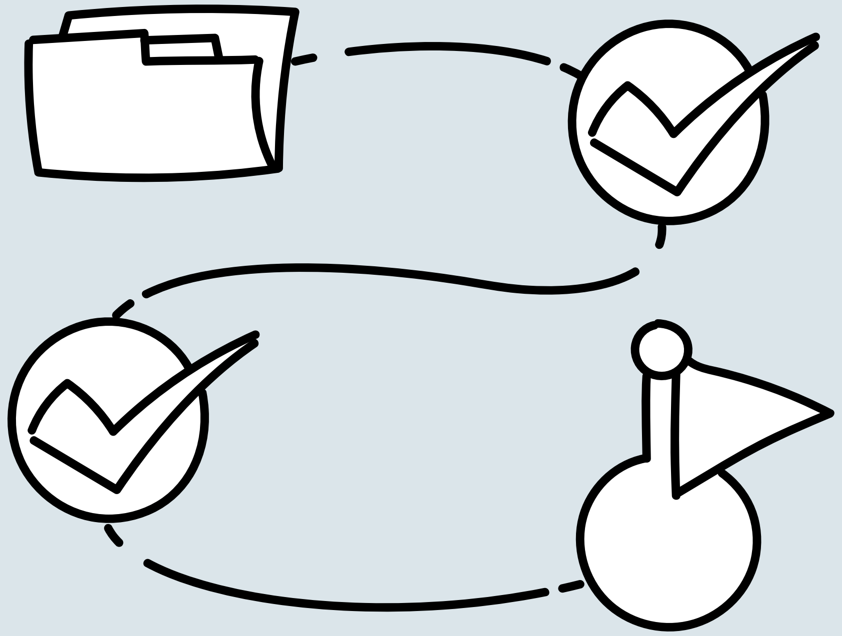
- Selection criteria:
  - 1) Highest accuracy, recall, and precision
  - 2) Practical business applicability

- Two final models:
  - 1) **Decision Tree:** identifies key factors for default
  - 2) **Logistic Regression:** evaluates future customers

THE MODELS

Decision Tree

- Helps us understand **existing** customers
- Assesses features such as late payments, credit card balance



Logistic Regression

- Helps us assess **future** customers
- External factors such as marital status or education

KEY PREDICTORS OF HIGH DEFAULT RISK

QUANT CREW

OVER
2 months
late on the **first payment**

OVER
$779
in the **bill amount**

OVER
$12,157
in credit card payments after the **third month**

OVER
$1,842
in credit card payments after the **second month**

KEY PREDICTORS OF LOW DEFAULT RISK

UNDER

2 months

late on the **first payment**

UNDER

$884

in credit card payments
after the **fourth month**

UNDER

2 months

late on the **second payment**

UNDER

$1,500

in credit card payments
after the **second month**

KEY PREDICTORS FOR FUTURE CUSTOMERS

MARITAL STATUS

Being **single** decreases future default chances by **15%**

EDUCATION

Having a **high school education** decreases future default chances by **140%**

INSIGHTS

Key Predictor

Having a **high school degree** is a significant factor for default chances

Flag Customers

Identifies those who make **first payment late**

Future Directions

Adding more **demographical information** – Credit Score, Income, Job Status, Rent, Number of Previous Cards etc.

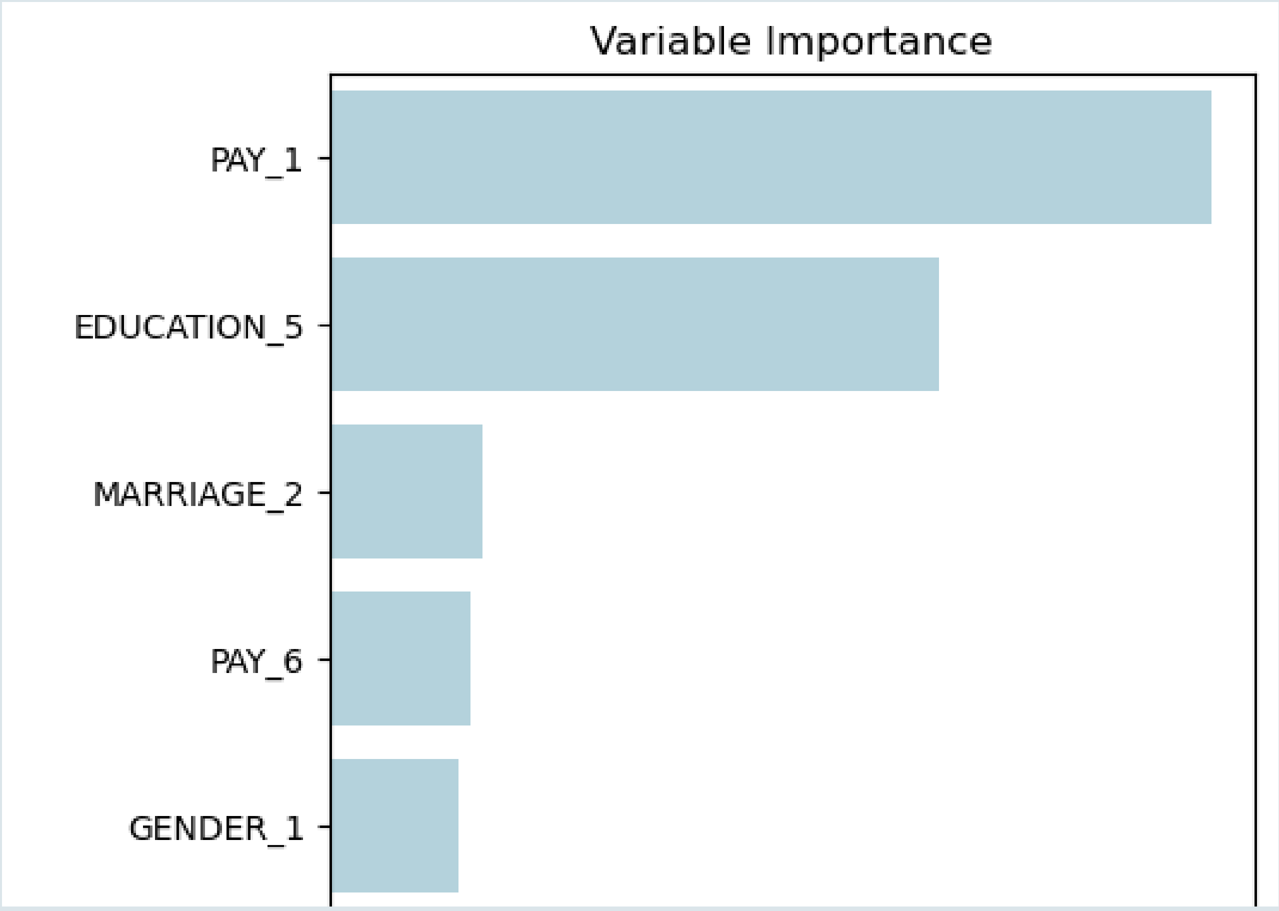
APPENDIX

Model Type	Accuracy (Train)	Precision (Train)	Recall (Train)	AUC (Train)	Accuracy (Test)	Precision (Test)	Recall (Test)	AUC (Test)
Logistic Model	0.8209	0.6933	0.3379	0.7631	0.8146	0.6756	0.3262	0.7529
Logistic Stepwise Model	0.8207	0.6943	0.3346	0.7604	0.8139	0.6732	0.3237	0.7571
L1 with c=0.1	0.8211	0.6942	0.3379	0.7626	0.8142	0.6746	0.3237	0.7587
L1 with c=0.01	0.8198	0.6958	0.3260	0.7555	0.8144	0.6821	0.3157	0.7546
L2	0.7794	0	0	0.6555	0.7766	0	0	0.6425
Elastic Net	0.7793	0	0	0.6577	0.7768	0	0	0.6471
Base Decision Tree	0.8247	0.6897	0.3731	0.7674	0.8177	0.6710	0.3601	0.7517
Tuned Decision Tree	0.8240	0.6975	0.3571	0.7660	0.8177	0.6800	0.3466	0.7535

LOGISTIC REGRESSION RESULTS

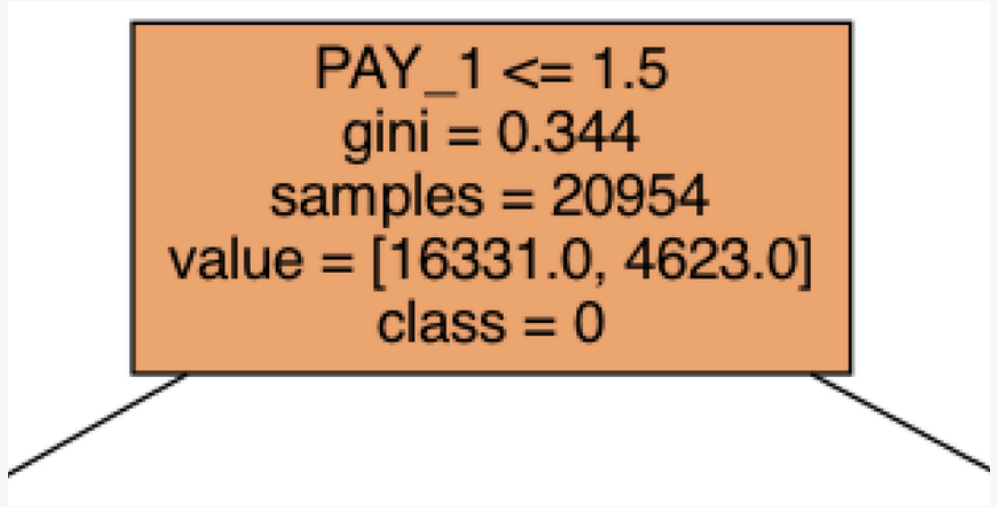
QUANT CREW

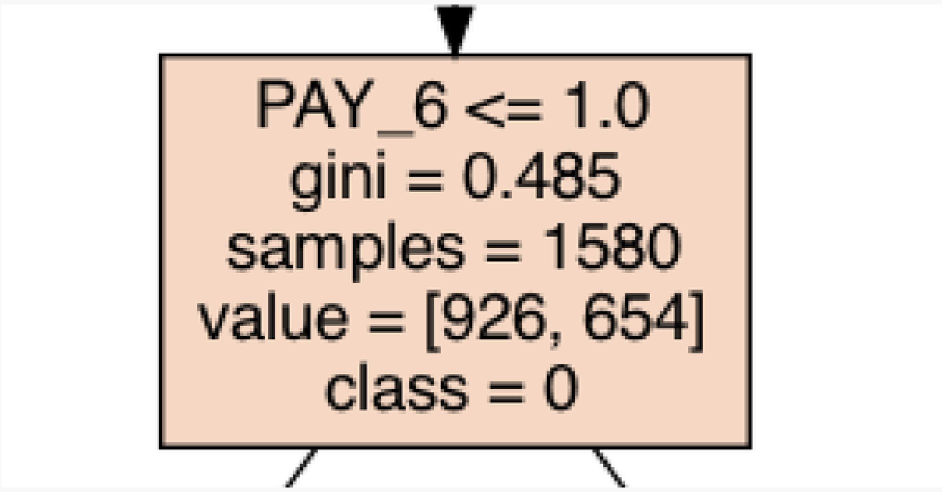
```
Optimization terminated successfully.
         Current function value: 0.443645
         Iterations 7
                           Logit Regression Results
==============================================================================
Dep. Variable:                DEFAULT   No. Observations:                20954
Model:                          Logit   Df Residuals:                    20942
Method:                           MLE   Df Model:                           11
Date:                Mon, 18 Nov 2024   Pseudo R-squ.:                  0.1593
Time:                        14:47:41   Log-Likelihood:                -9296.1
converged:                       True   LL-Null:                       -11057.
Covariance Type:            nonrobust   LLR p-value:                     0.000
===============================================================================
                  coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------
const          -1.6134      0.097    -16.677      0.000      -1.803      -1.424
LIMIT_BAL   -1.587e-06   1.68e-07     -9.448      0.000   -1.92e-06   -1.26e-06
AGE             0.0038      0.002      1.682      0.093      -0.001       0.008
PAY_1           0.9221      0.026     34.996      0.000       0.870       0.974
PAY_3           0.1795      0.027      6.723      0.000       0.127       0.232
PAY_5           0.1300      0.036      3.639      0.000       0.060       0.200
PAY_6           0.1511      0.034      4.401      0.000       0.084       0.218
PAY_AMT1    -1.179e-05   2.45e-06     -4.816      0.000   -1.66e-05   -6.99e-06
PAY_AMT2     -8.28e-06   2.15e-06     -3.857      0.000   -1.25e-05   -4.07e-06
GENDER_1        0.1513      0.038      3.990      0.000       0.077       0.226
EDUCATION_5    -1.3665      0.323     -4.231      0.000      -2.000      -0.734
MARRIAGE_2     -0.1484      0.042     -3.506      0.000      -0.231      -0.065
===============================================================================
```

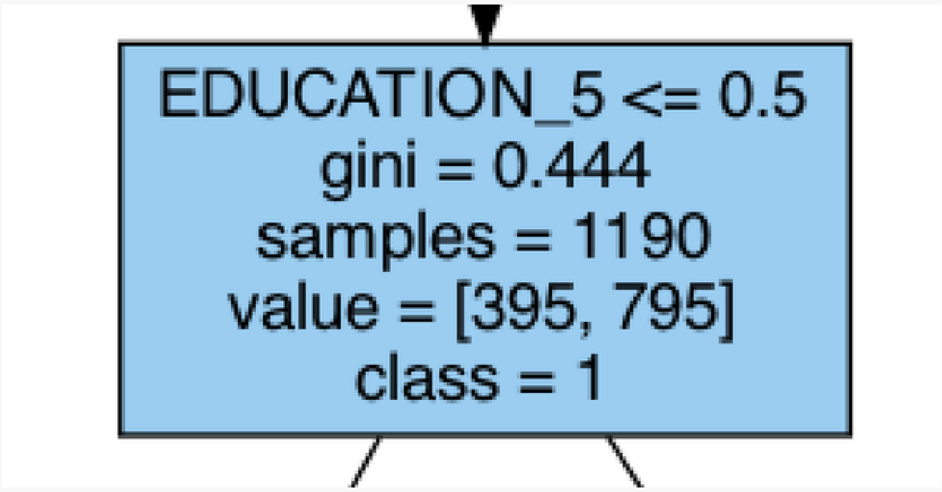


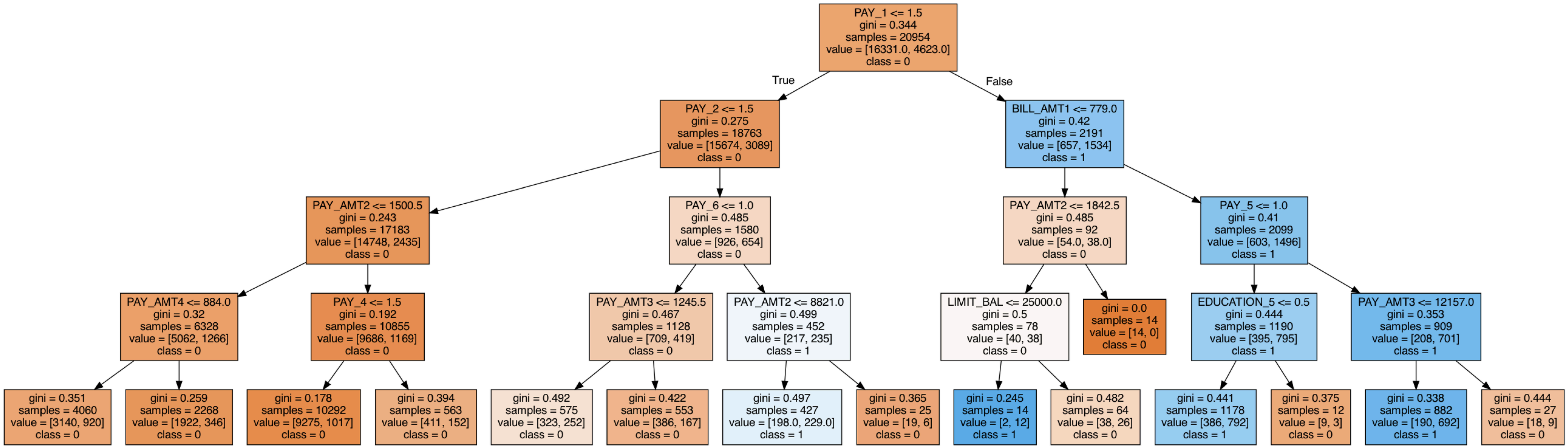
PLOT OF VARIABLE IMPORTANCE

DECISION TREE VARIABLES








PAY_1 <= 1.5
gini = 0.344
samples = 20954
value = [16331.0, 4623.0]
class = 0
True
False
PAY_2 <= 1.5
gini = 0.275
samples = 18763
value = [15674, 3089]
class = 0
BILL_AMT1 <= 779.0
gini = 0.42
samples = 2191
value = [657, 1534]
class = 1
PAY_AMT2 <= 1500.5
gini = 0.243
samples = 17183
value = [14748, 2435]
class = 0
PAY_6 <= 1.0
gini = 0.485
samples = 1580
value = [926, 654]
class = 0
PAY_AMT2 <= 1842.5
gini = 0.485
samples = 92
value = [54.0, 38.0]
class = 0
PAY_5 <= 1.0
gini = 0.41
samples = 2099
value = [603, 1496]
class = 1
PAY_AMT4 <= 884.0
gini = 0.32
samples = 6328
value = [5062, 1266]
class = 0
PAY_4 <= 1.5
gini = 0.192
samples = 10855
value = [9686, 1169]
class = 0
PAY_AMT3 <= 1245.5
gini = 0.467
samples = 1128
value = [709, 419]
class = 0
PAY_AMT2 <= 8821.0
gini = 0.499
samples = 452
value = [217, 235]
class = 1
LIMIT_BAL <= 25000.0
gini = 0.5
samples = 78
value = [40, 38]
class = 0
gini = 0.0
samples = 14
value = [14, 0]
class = 0
EDUCATION_5 <= 0.5
gini = 0.444
samples = 1190
value = [395, 795]
class = 1
PAY_AMT3 <= 12157.0
gini = 0.353
samples = 909
value = [208, 701]
class = 1
gini = 0.351
samples = 4060
value = [3140, 920]
class = 0
gini = 0.259
samples = 2268
value = [1922, 346]
class = 0
gini = 0.178
samples = 10292
value = [9275, 1017]
class = 0
gini = 0.394
samples = 563
value = [411, 152]
class = 0
gini = 0.492
samples = 575
value = [323, 252]
class = 0
gini = 0.422
samples = 553
value = [386, 167]
class = 0
gini = 0.497
samples = 427
value = [198.0, 229.0]
class = 1
gini = 0.365
samples = 25
value = [19, 6]
class = 0
gini = 0.245
samples = 14
value = [2, 12]
class = 1
gini = 0.482
samples = 64
value = [38, 26]
class = 0
gini = 0.441
samples = 1178
value = [386, 792]
class = 1
gini = 0.375
samples = 12
value = [9, 3]
class = 0
gini = 0.338
samples = 882
value = [190, 692]
class = 1
gini = 0.444
samples = 27
value = [18, 9]
class = 0