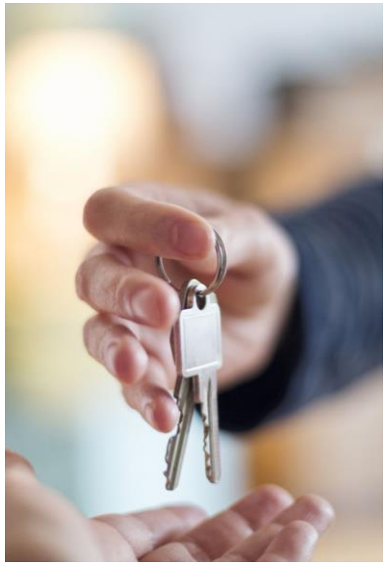


KEY DRIVERS FOR LOAN PAYBACK FOUND USING SEGMENTED LOGISTIC CLASSIFICATION MODELS

Towards More Customized Approval Criteria

Max Sydow

September, 2021

GOAL: USE COEFFICIENTS FROM MULTIPLE LOGIT MODELS TO PROVIDE INSIGHT TO WHAT INFLUENCES LOAN PAYBACK.

- Subset data according to unique values in categorical columns.
- Fit logit models for each subset.
- Accuracy and coefficients can be used to identify key influencers.

- ***Will there be enough accurate models with enough predictor variables to make these kinds of conclusion?***

  - use of AUC score in recursive feature elimination to determine final models.

DATA SUMMARY

110,000 rows of Loan data, 18 columns. (Kaggle)

Target variable: Loan Status – 1 = Fully Paid, 0 = Charged Off (not paid in full).

12 Predictor variables: Current_Loan_Amount, Credit_Score, Annual_Income, Monthly_Debt, Years_of_Credit_History, Number_of_Open_Accounts, Number_of_Credit_Problems, Current_Credit_Balance, Maximum_Open_Credit, Bankruptcies, Tax_Liens.

Categorical columns to subset by: Term, Years in current job, Home Ownership, Purpose.

DATA PREPARATION

30 subsets with number of rows ranging from 113 to 77090.

Subsets with less than 100 rows dropped to ensure enough data for accurate models.

Outliers removed, missing values replaced with median per distribution shape.

Purpose: Wedding and moving had 2-3 columns dropped due to limited variation of values. The rest of the subsets used all 12 predictor columns in their models.

SUBSETS.
ONE MULIT-LOGIT MODEL FOR EACH.

- Term: Short Term, Long Term
- Years in current job : < 1 year, 1 year, 2 years, 3 years, 4 years, 5 years, 6 years, 7 years, 8 years, 9 years, 10+ years
- Home Ownership: Home Mortgage, Own Home, Rent
- Purpose: Home Improvements, Debt Consolidation, Buy House, other, Business Loan, Buy a Car, major_purchase, Take a Trip, small_business, Medical Bills, wedding, vacation, moving

BUILDING MODELS

- RFECV – Recursive Feature Elimination with Cross Validation.

  Logit models with random collections of column subsets compared.

  AUC scores used as performance metric.

    - ensures models perform better than random guessing.
- 70/30 train/test split. All features normalized.

* If accuracy used instead of AUC models may still reflect random guessing performance, but may have given better predictions on test data. Something to try in future iterations of this effort. (Zach, 2021)

FINDINGS

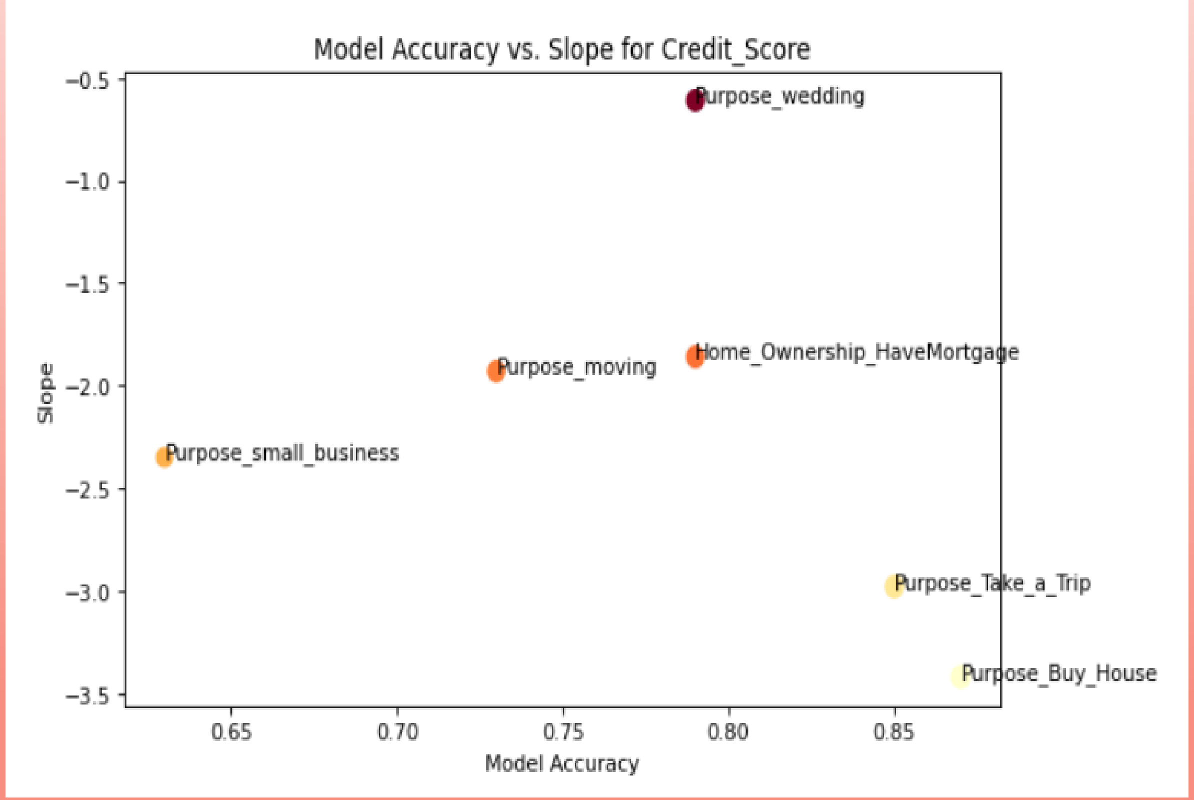
- Loans for buying a house predict payoff with highest accuracy.
- High predictive accuracy on repayment for taking a trip, short term, mortgage and home improvement loans for homeowners, and other.

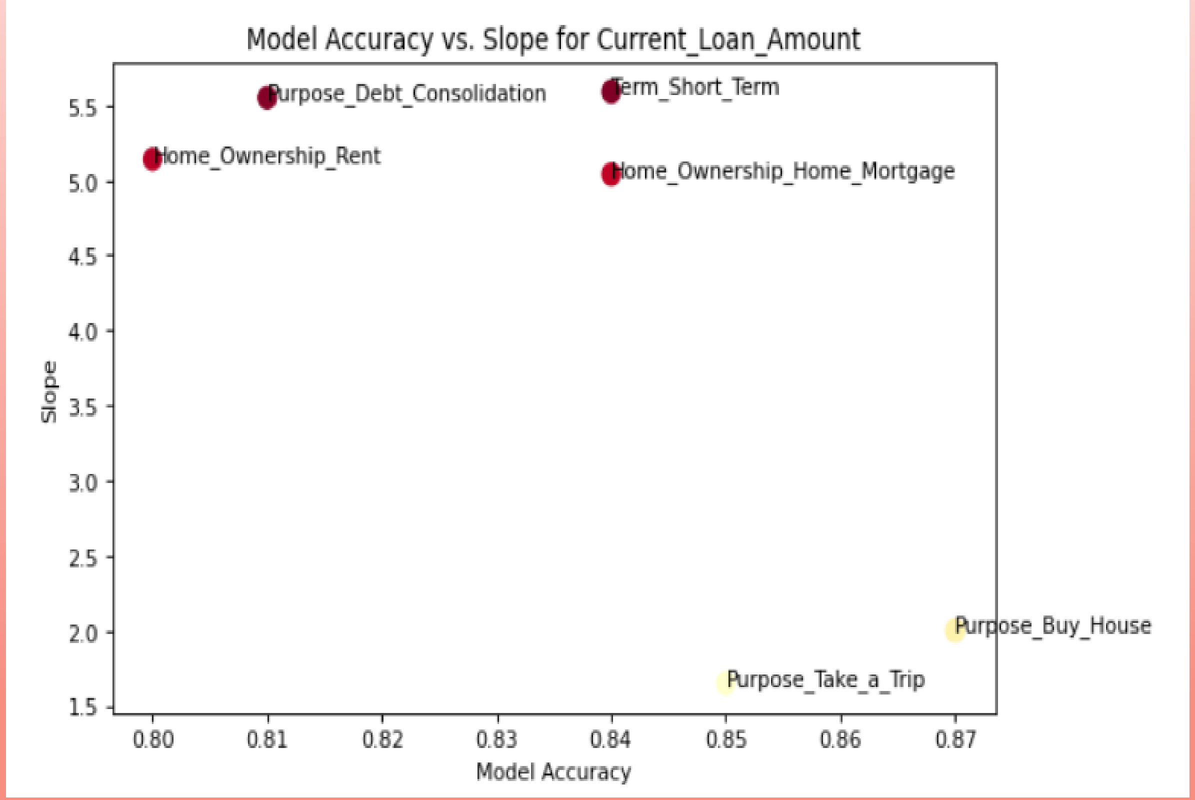
Segment	Accuracy	Current_Loan_Amount	Credit_Score	Annual_Income	Monthly_Debt	Years_of_Credit_History
Term_Short_Term	0.84	5.59	-7.27	4.50	0.56	0.59
Term_Long_Term	0.76	4.85	-6.58	5.57	-1.72	0.17
Years_in_current_job_less1_year	0.81	3.84	-5.39	2.84	-0.24	-0.29
Years_in_current_job_1_year	0.81	3.54	-5.38	3.15	-0.01	-0.07
Years_in_current_job_2_years	0.83	3.93	-5.48	2.65	-0.12	-0.22
Years_in_current_job_3_years	0.83	3.68	-5.48	2.43	-1.37	0.35
Years_in_current_job_4_years	0.81	3.59	-5.04	1.59	-0.53	0.22
Years_in_current_job_5_years	0.82	3.72	-5.12	1.91	-1.05	0.37
Years_in_current_job_6_years	0.81	3.55	-5.11	1.76	0.03	1.10
Years_in_current_job_7_years	0.82	3.47	-5.13	2.73	-0.92	0.32
Years_in_current_job_8_years	0.81	3.26	-4.85	1.84	-0.41	0.85
Years_in_current_job_9_years	0.82	3.36	-4.62	1.85	-0.89	0.65
Years_in_current_job_10plus_years	0.83	4.77	-6.49	2.21	-0.53	1.07
Home_Ownership_Home_Mortgage	0.84	5.04	-6.89	2.93	-0.86	0.50
Home_Ownership_Own_Home	0.81	3.90	-5.50	1.47	-0.13	-0.03
Home_Ownership_Rent	0.80	5.14	-6.64	6.72	-2.24	0.08
Home_Ownership_HaveMortgage	0.79	1.10	-1.86	0.90	0.09	-0.26
Purpose_Home_Improvements	0.84	3.45	-5.04	3.25	-0.69	0.59
Purpose_Debt_Consolidation	0.81	5.55	-7.15	4.96	-1.30	0.38
Purpose_Buy_House	0.87	2.00	-3.42	1.06	0.89	0.11
Purpose_other	0.80	3.52	-5.06	2.64	-0.54	0.14
Purpose_Business_Loan	0.75	2.55	-3.96	0.64	-0.11	0.55
Purpose_Buy_a_Car	0.83	2.12	-3.98	1.11	0.54	0.81
Purpose_major_purchase	0.81	1.66	-2.60	-0.18	-1.02	0.11
Purpose_Take_a_Trip	0.85	1.65	-2.98	0.59	0.21	0.85
Purpose_Other	0.84	2.97	-4.49	1.18	-0.36	0.43
Purpose_small_business	0.63	1.79	-2.35	0.64	-0.90	-0.34
Purpose_Medical_Bills	0.83	2.45	-3.56	0.35	-0.31	1.59
Purpose_wedding	0.79	0.99	-0.61	0.43	0.25	-0.01
Purpose_moving	0.73	1.35	-1.93	0.29	-0.11	-0.34

Segment	Number_of_Open_Accounts	Number_of_Credit_Problems	Current_Credit_Balance	Maximum_Open_Credit	Bankruptcies	Tax_Liens
Term_Short_Term	-0.54	-0.46	1.10	0.70	0.51	0.00
Term_Long_Term	0.01	0.39	1.41	0.25	-0.16	0.00
Years_in_current_job_less1_year	0.02	0.04	-0.28	0.06	-0.47	0.00
Years_in_current_job_1_year	-0.02	-0.24	-0.24	0.38	1.04	-0.21
Years_in_current_job_2_years	-0.20	0.36	0.26	0.36	0.26	0.00
Years_in_current_job_3_years	-0.52	-0.04	0.81	0.11	-0.23	0.00
Years_in_current_job_4_years	-0.06	0.25	-0.47	0.50	-0.13	0.40
Years_in_current_job_5_years	-0.33	0.08	0.20	-0.09	-0.23	0.00
Years_in_current_job_6_years	-0.57	0.22	0.52	0.08	0.13	0.00
Years_in_current_job_7_years	-1.04	0.12	0.39	0.89	0.67	0.00
Years_in_current_job_8_years	0.02	-0.17	0.55	0.76	0.32	0.00
Years_in_current_job_9_years	0.08	-0.12	0.37	1.31	-0.31	0.00
ars_in_current_job_10plus_years	-0.66	0.22	0.35	0.50	0.22	0.00
ome_Ownership_Home_Mortgage	-0.62	-0.39	0.77	0.39	0.14	0.00
Home_Ownership_Own_Home	-0.34	0.35	1.01	0.35	-0.21	0.00
Home_Ownership_Rent	-0.21	-0.42	-0.89	1.10	0.27	0.00
Home_Ownership_HaveMortgage	-0.49	0.43	-0.29	0.12	0.43	0.00
Purpose_Home_Improvements	-0.52	0.79	0.83	-0.75	0.49	0.00
Purpose_Debt_Consolidation	-0.55	0.03	1.47	0.51	-0.05	0.00
Purpose_Buy_House	-0.28	-0.28	0.40	0.34	0.32	0.00
Purpose_other	-0.50	-0.00	-0.02	0.33	1.48	0.00
Purpose_Business_Loan	-0.81	-0.10	-1.02	0.04	0.39	0.00
Purpose_Buy_a_Car	-0.29	0.11	0.22	0.29	0.43	0.00
Purpose_major_purchase	-0.04	0.20	0.05	0.27	0.55	0.00
Purpose_Take_a_Trip	-0.19	0.92	0.62	0.11	0.42	0.00
Purpose_Other	-0.22	-0.29	0.26	0.23	0.21	0.00
Purpose_small_business	-0.31	0.36	0.08	-0.02	0.15	0.00
Purpose_Medical_Bills	-0.51	0.61	-0.64	0.41	-0.55	0.00
Purpose_wedding	-0.63	0.22	-0.03	0.18	0.00	0.00
Purpose_moving	1.06	-0.26	0.27	0.43	-0.26	0.00

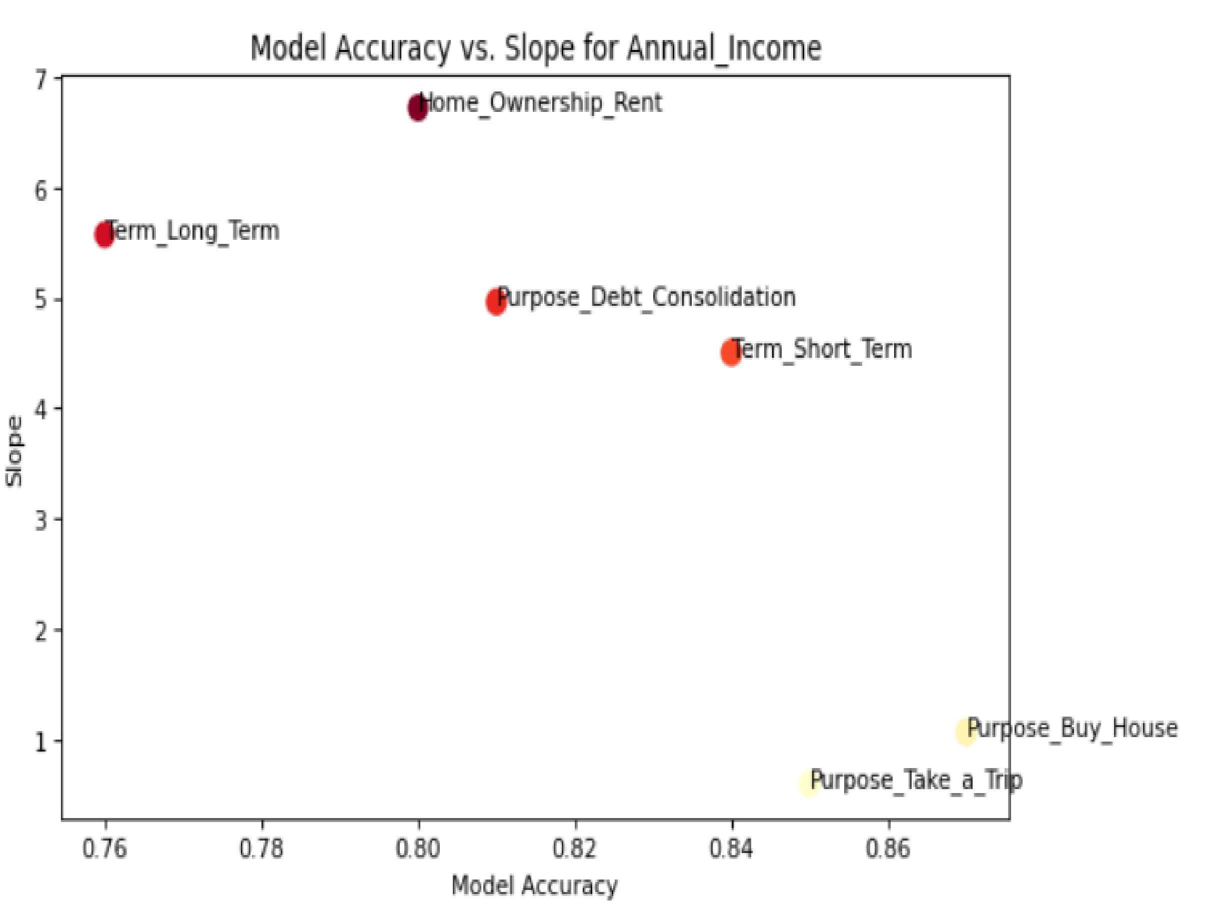
Model coefficients vs. accuracy.

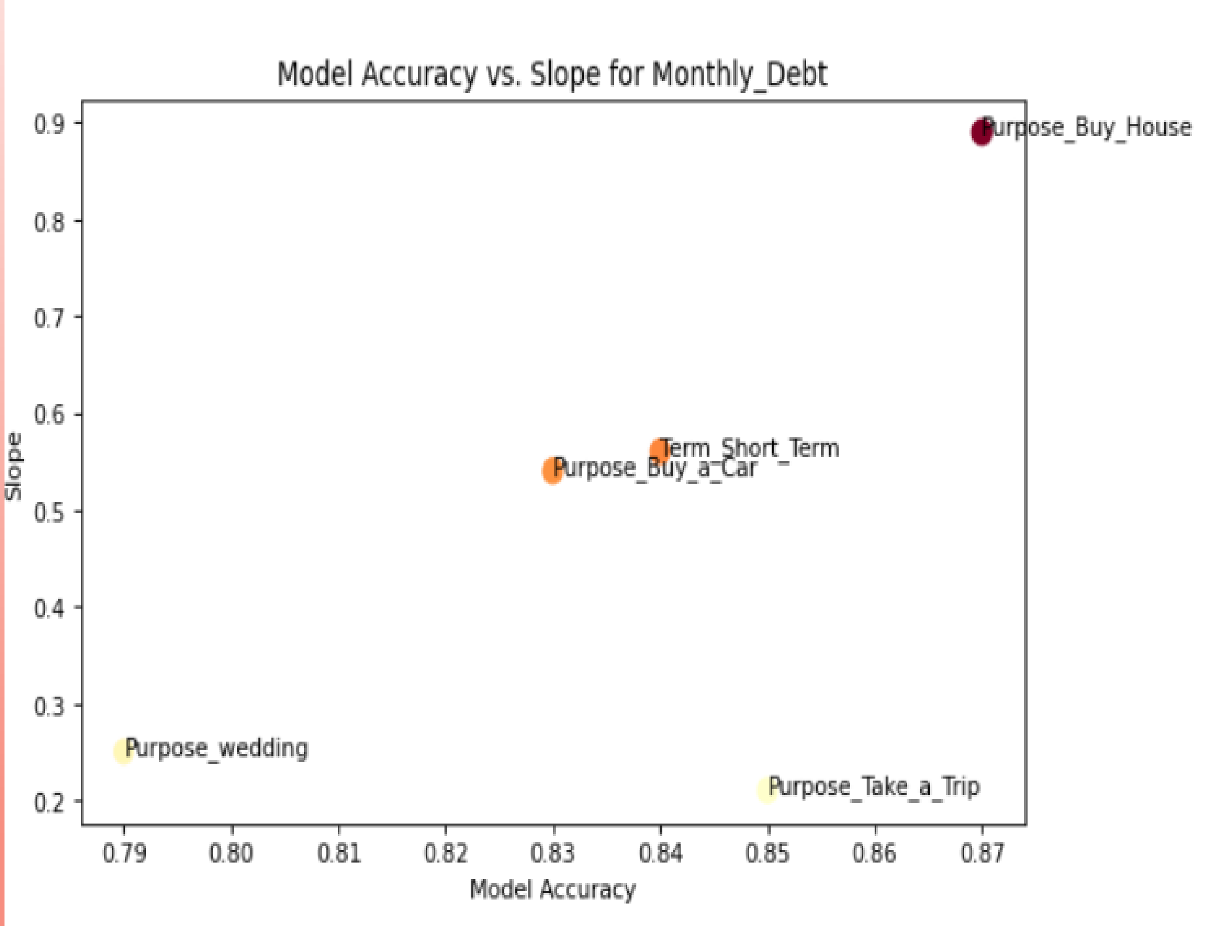
High accuracy and large positive coefficient = high influence.



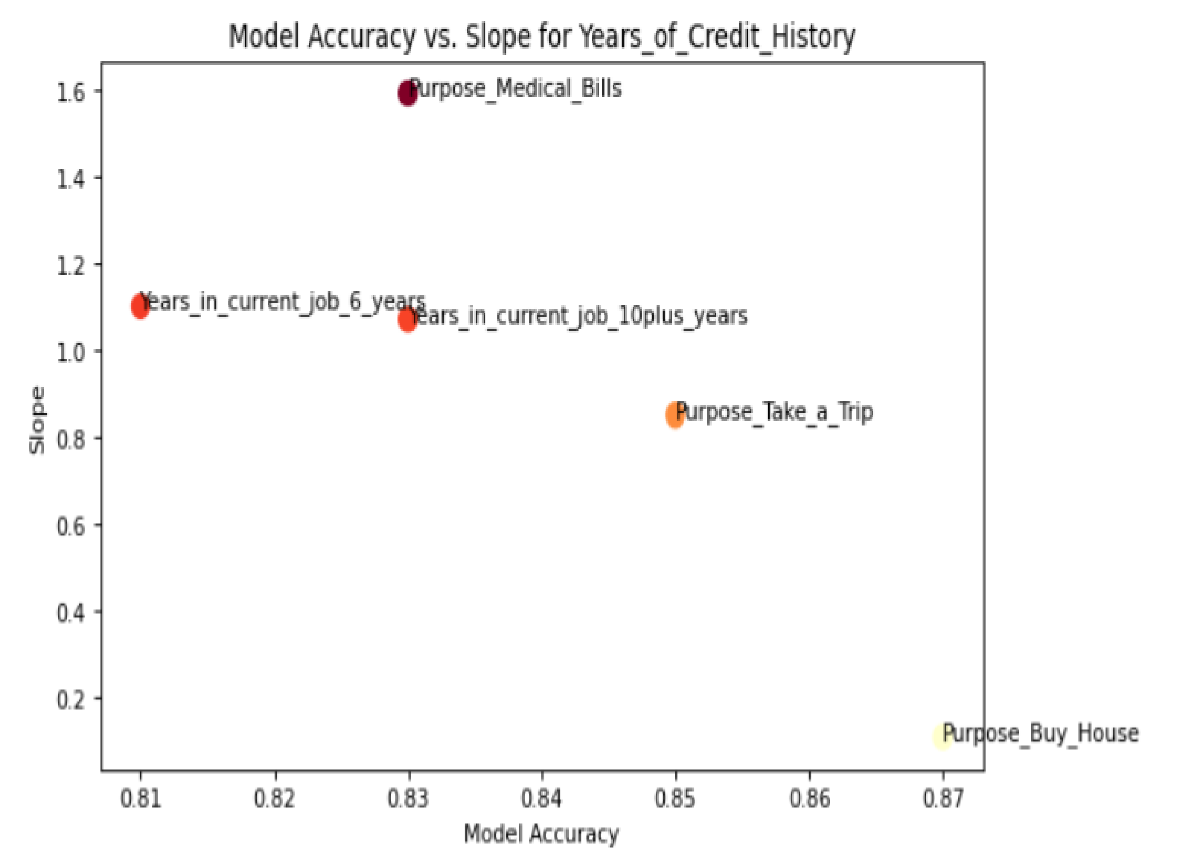


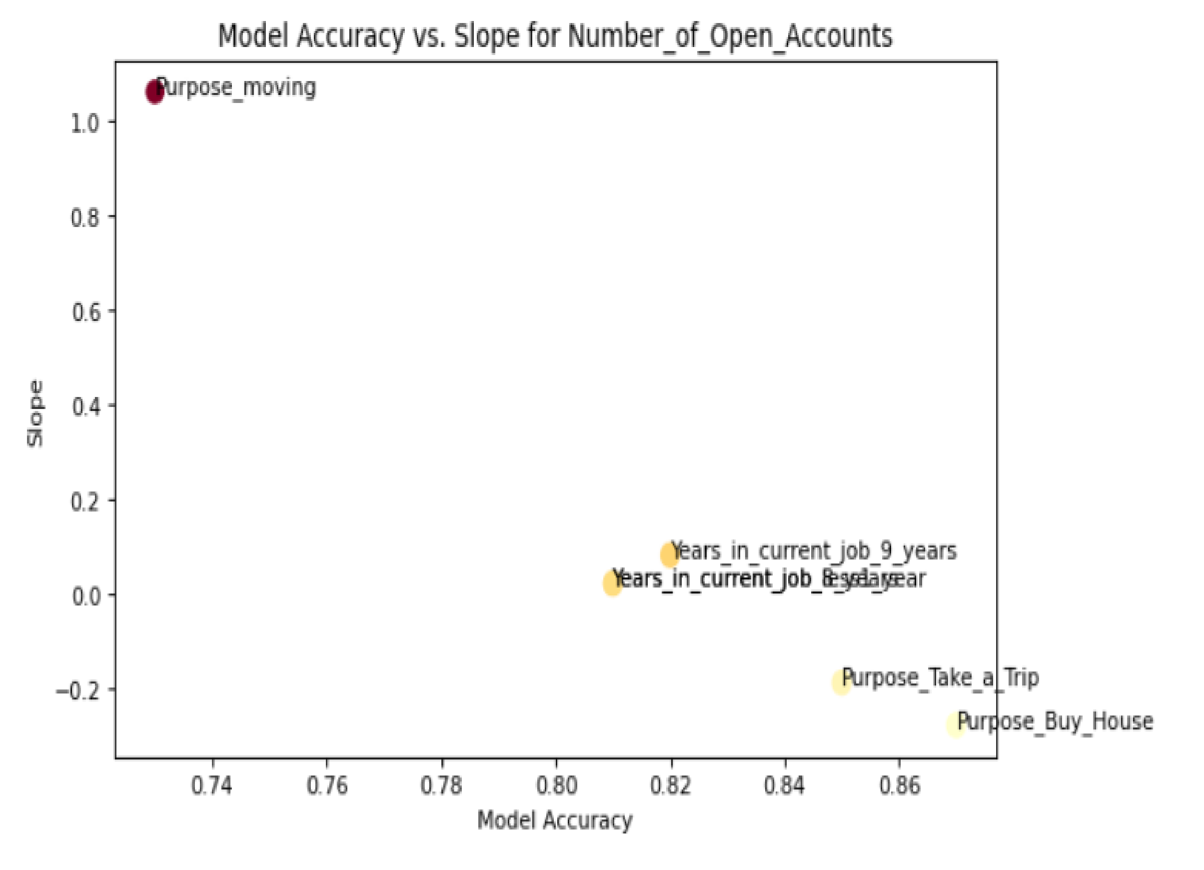
Short term loans and home mortgage loan payoffs are more effected by greater current loan amounts. Credit score has a negative influence on repayment for even the most accurate model segments.



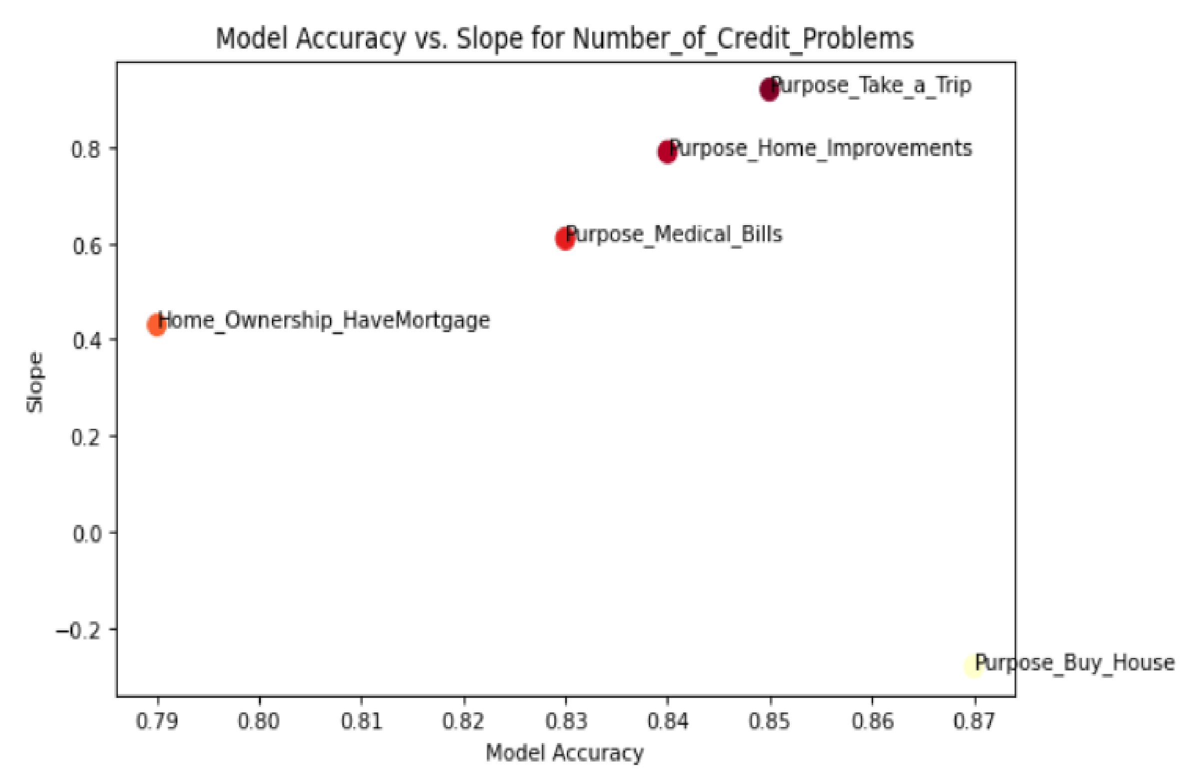


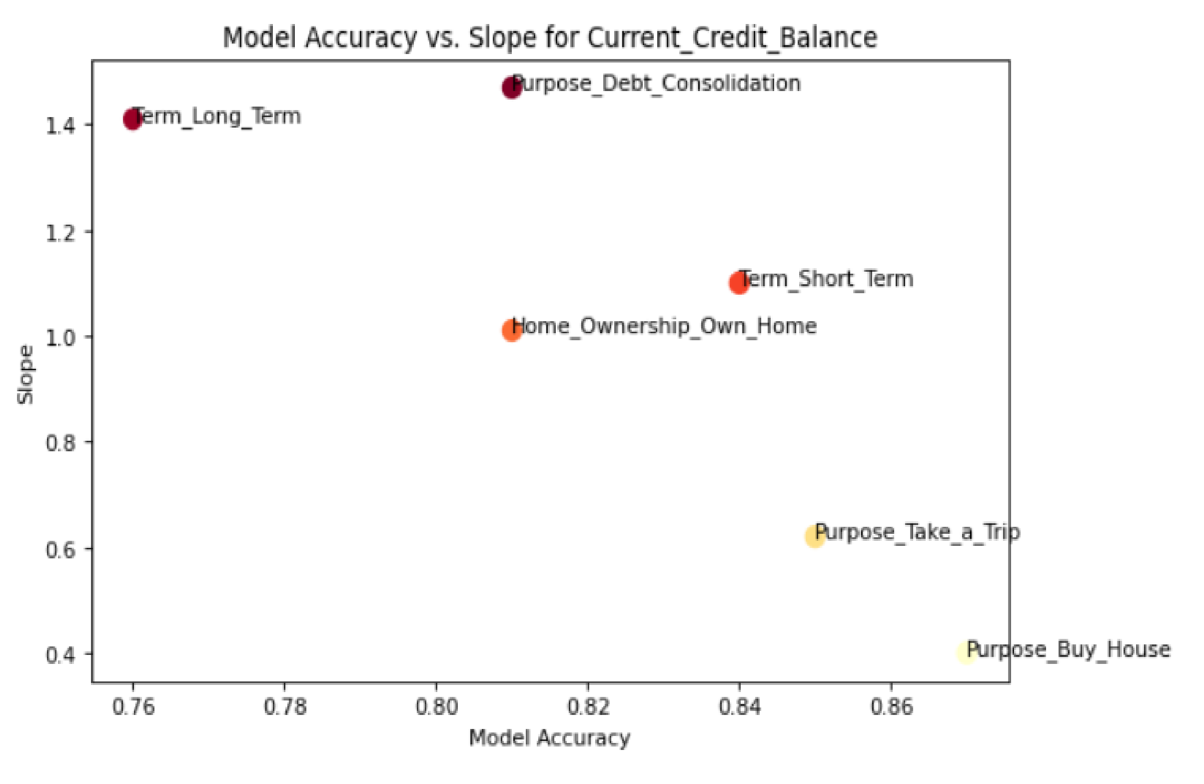
Annual income influences repayment the most for rent, debt consolidation, and sort term loans. Customers who are buying a house are more likely to pay off the loan if they have a high monthly debt.



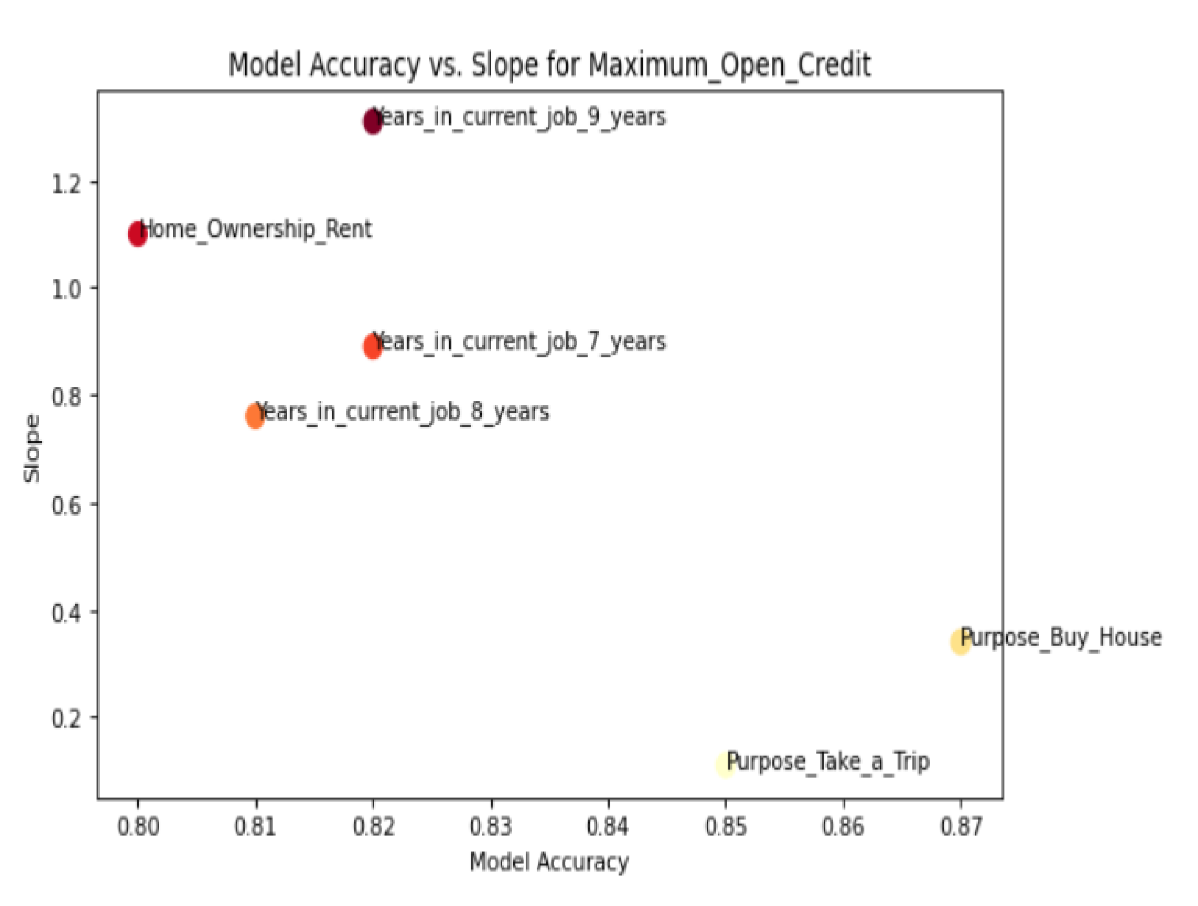


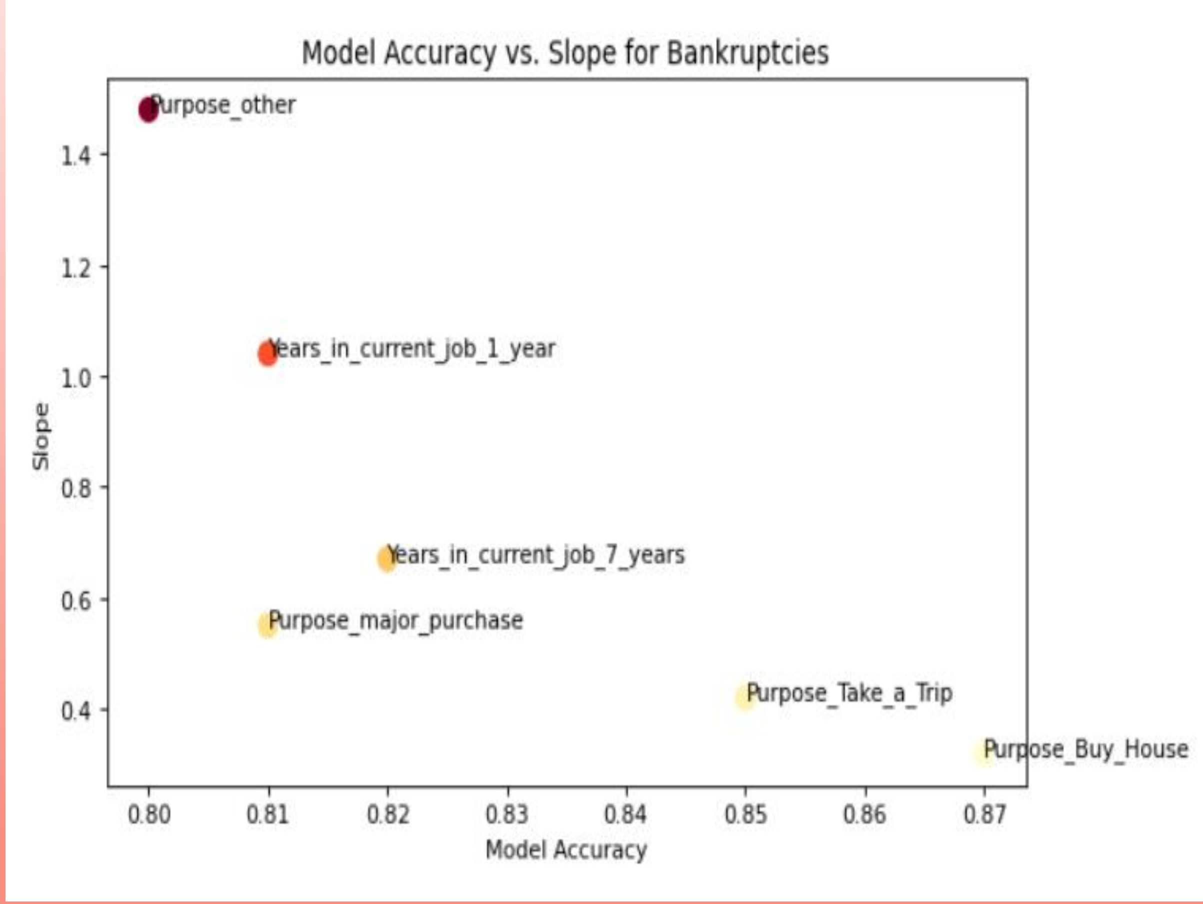
Payoff of loans for medical bills seem to be most influenced by credit history. The number of open accounts is most influential on payback for moving loans, but is not a very accurate predictor as other metrics.



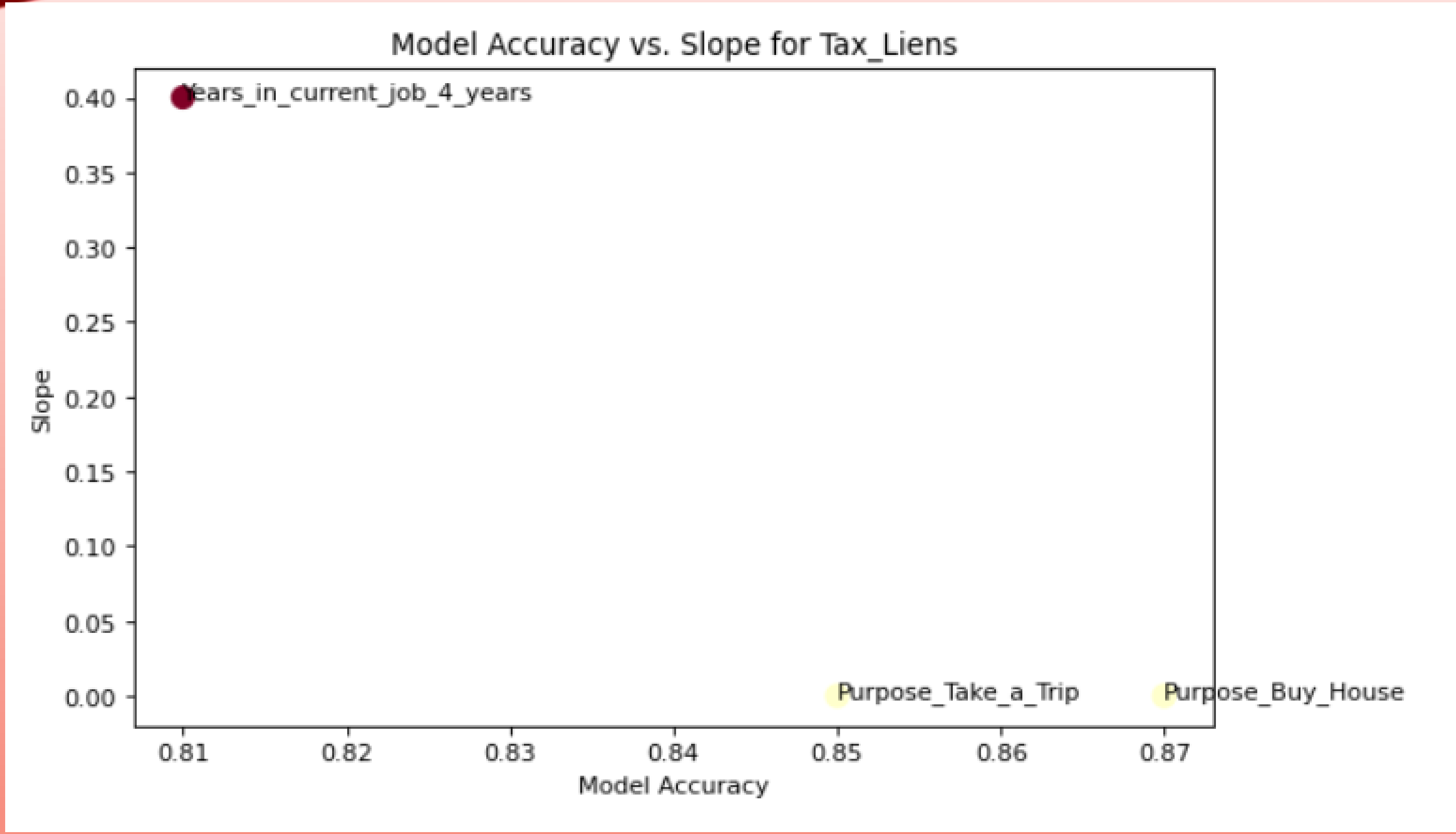


Number of credit problems is weakly influential on payback for loans to take a trip, make home improvements and pay medical bills. Credit balance is most influential on loans for debt consolidation.





Maximum open credit seems to have the most impact on repayment for renter's loans and for those who have been at their jobs for 7-9 years. The number of bankruptcies have heavy influence on non-specified loan types and for those who have less than 1 year on the job.



Number of tax liens is most impacted by those who have been on the job for 4 years, but the strength of influence is rather weak. It may be concluded that this metric does little to predict loan payback.

Benefits

- Knowing key influencers for certain loan types can expedite approval process.
- Allows lenders to focus on individual customer attributes.

Further Actions

- Compare these results with tree-based model feature importance.
- Similar procedures with linear regression using loan amount as target variable.
- Identify possible clusters and interactions amongst predictor variables.

REFERENCE SOURCES

https://www.kaggle.com/datasets/zaurbegiev/my-dataset?select=credit_test.csv

(Zach. (2021, September 9). What is Considered a Good AUC Score? Statology. https://www.statology.org/what-is-a-good-auc-score/)