Credit Card Payments

Presented by Deautaun Ross, Marci Morrell, and Jasmine Castañeda

March 2022

Outline

Data Overview
Data Cleaning
Card Status
Target Distributions
Model Building
Limitations
Conclusion

Data Overview

Data from Kaggle about credit card payments

Original Dataset	777,715 Rows	36,457 Rows	Model Dataset
	20 Columns	8 Columns	

Median percent of payments missed: 45%

Overall Mode Card Status: 0

Data Cleaning

- Renamed columns
- Dropped occupation
- Converted data to the proper types
- Label encoded the objects
- Used clients who have been a client for 6 months or longer

&

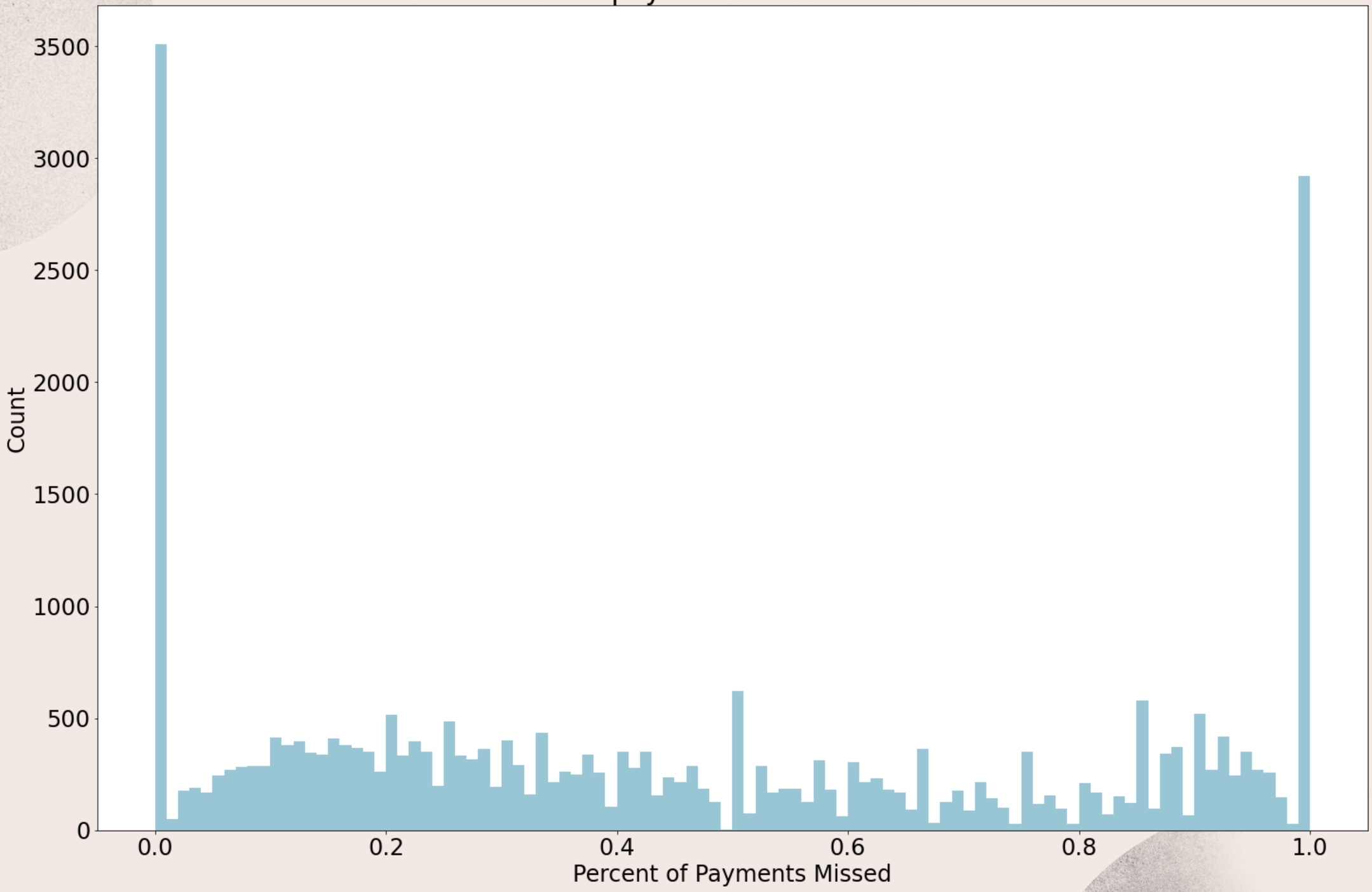
Feature Engineering

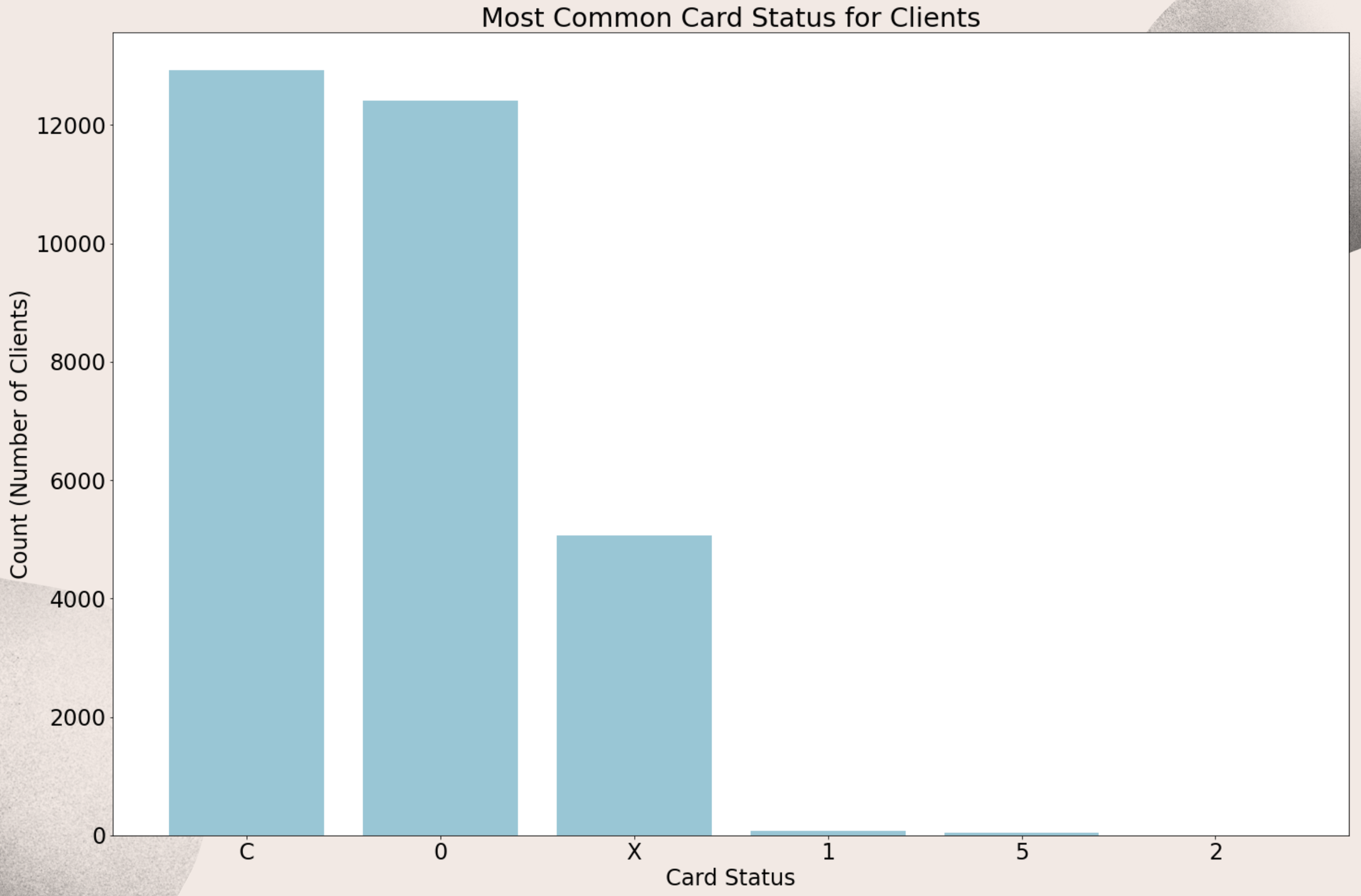
- Total amount of card status entries for a unique card holder
- Percentage of payments they have missed
- Mode card status of unique card holder

Card Status

0	<1 month overdue
1	1-2 months overdue
2	2-3 months overdue
3	3-4 months overdue
4	4-5 months overdue
5	>5 months overdue
C	Paid off that month
X	No loan for the month

Distribution of Target Columns


Distribution of payments missed for each client
Count
3500
3000
2500
2000
1500
1000
500
0
0.0
0.2
0.4
0.6
0.8
1.0
Percent of Payments Missed


Most Common Card Status for Clients
Count (Number of Clients)
0
2000
4000
6000
8000
10000
12000
C
0
X
1
5
2
Card Status

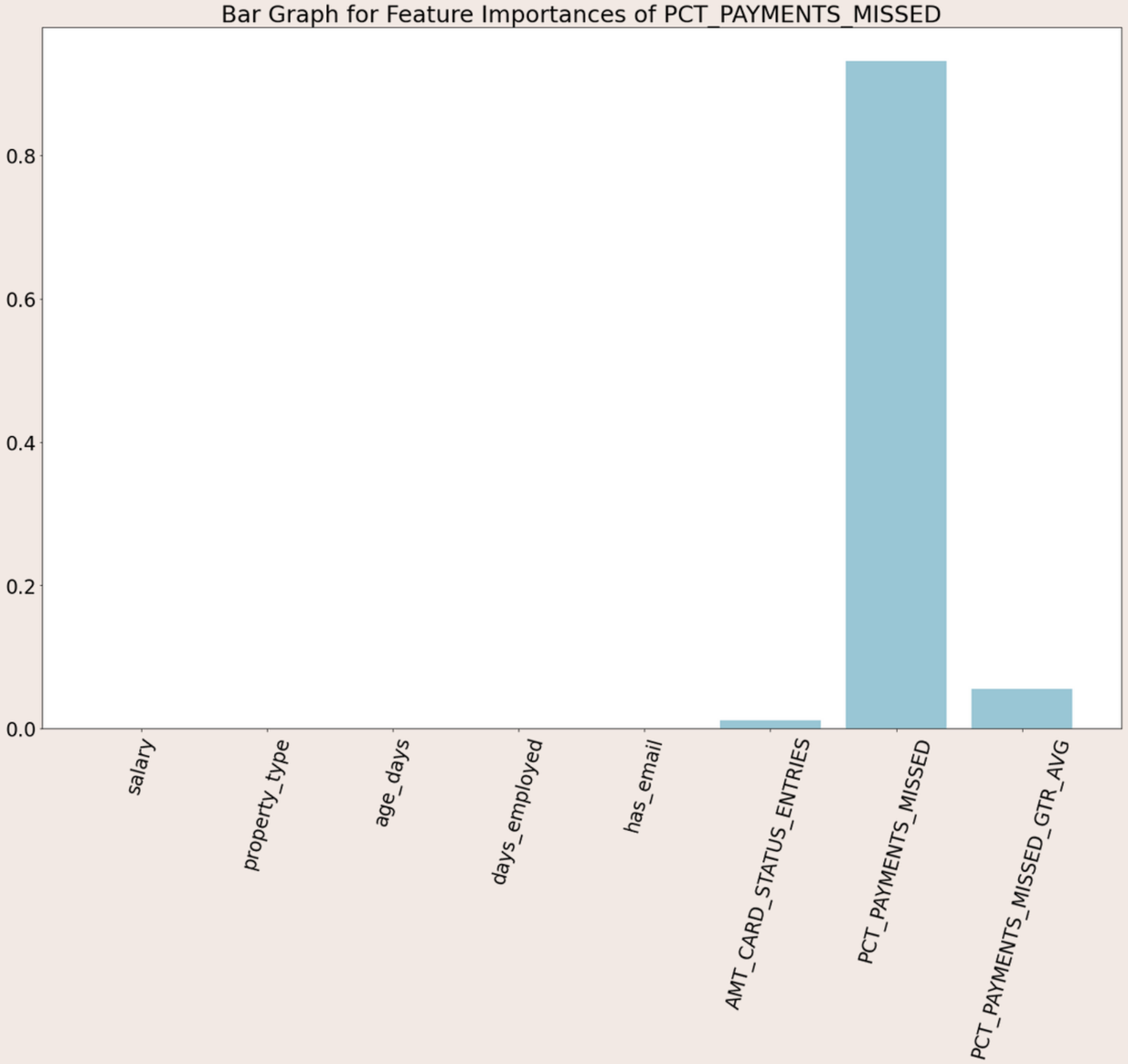
Model Building

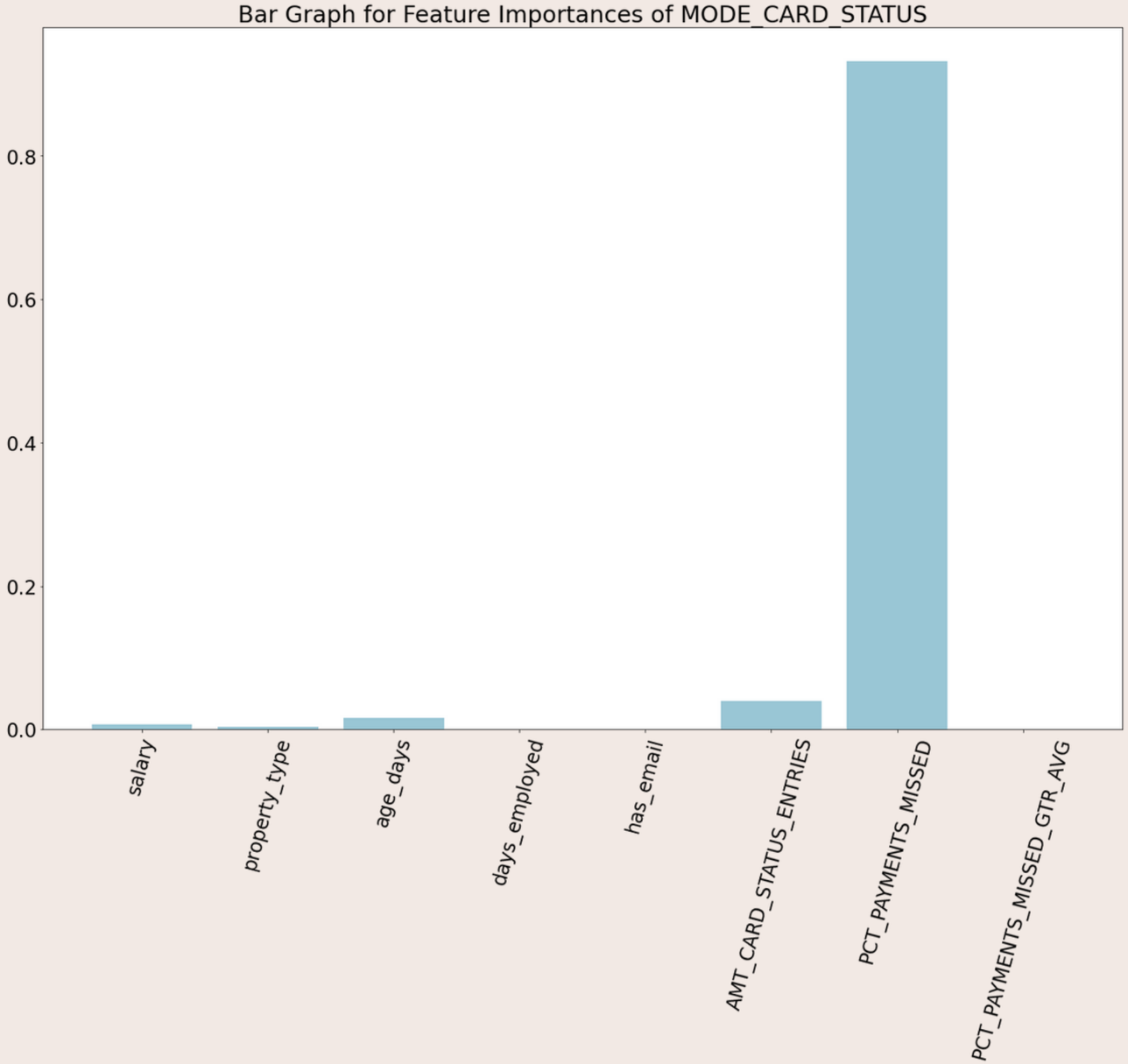
Percent of Missed Payments

- Regression Model
- Gradient Boost Regressor

Mode of Card Status

- Classification Model
- Gradient Boost Classifier


Bar Graph for Feature Importances of PCT_PAYMENTS_MISSED
0.0
0.2
0.4
0.6
0.8
salary
property_type
age_days
days_employed
has_email
AMT_CARD_STATUS_ENTRIES
PCT_PAYMENTS_MISSED
PCT_PAYMENTS_MISSED_GTR_AVG


Bar Graph for Feature Importances of MODE_CARD_STATUS
0.0
0.2
0.4
0.6
0.8
salary
property_type
age_days
days_employed
has_email
AMT_CARD_STATUS_ENTRIES
PCT_PAYMENTS_MISSED
PCT_PAYMENTS_MISSED_GTR_AVG

Features Used in Models

Salary

Property Type

Age

Days Employed

Has E-mail

Length of Card Holding

Mode Card Status

Is the Percent of Missed Payments Greater than the Median (Boolean)

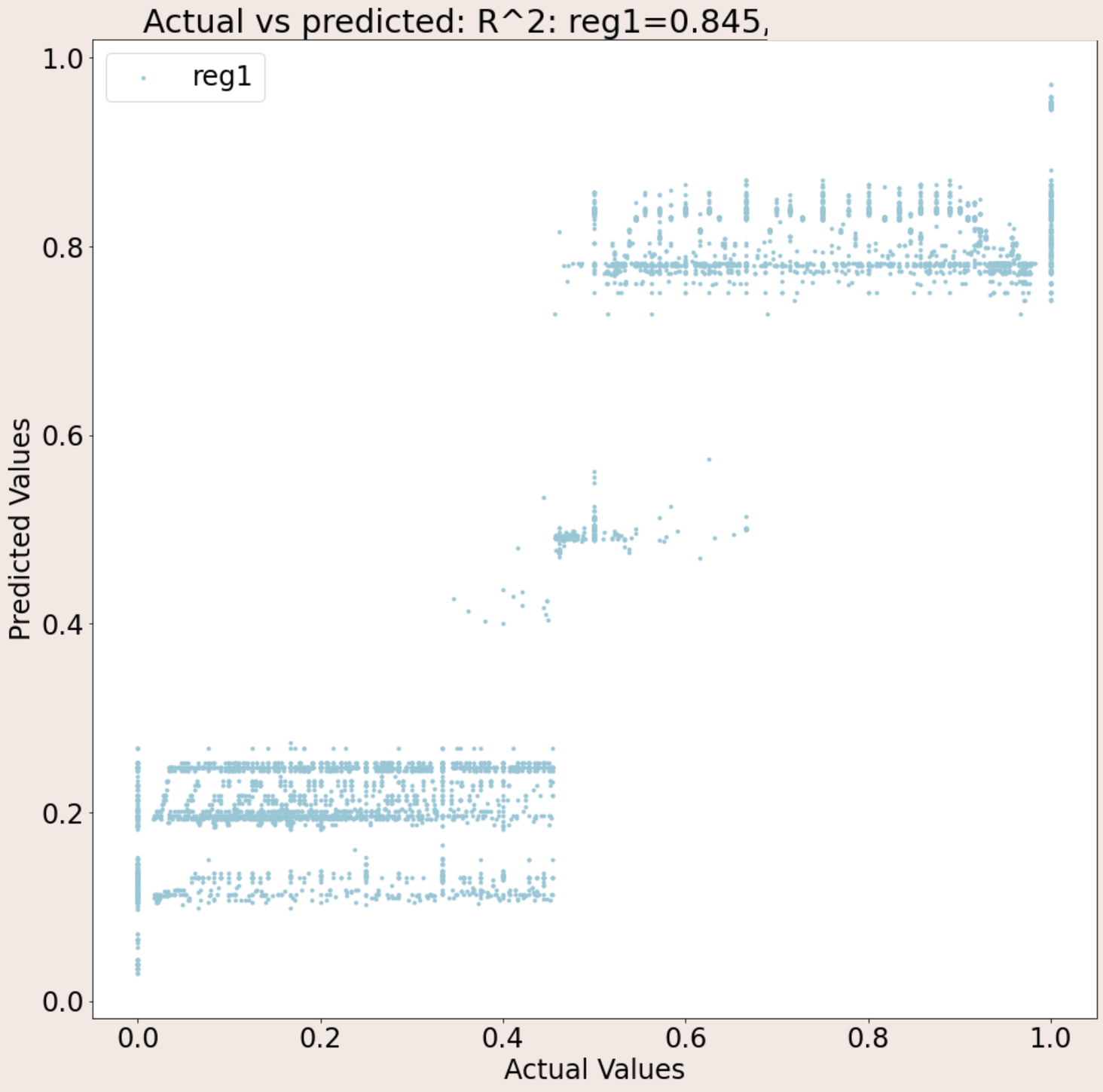
Percent of Payments Missed

Percent of Payments Missed

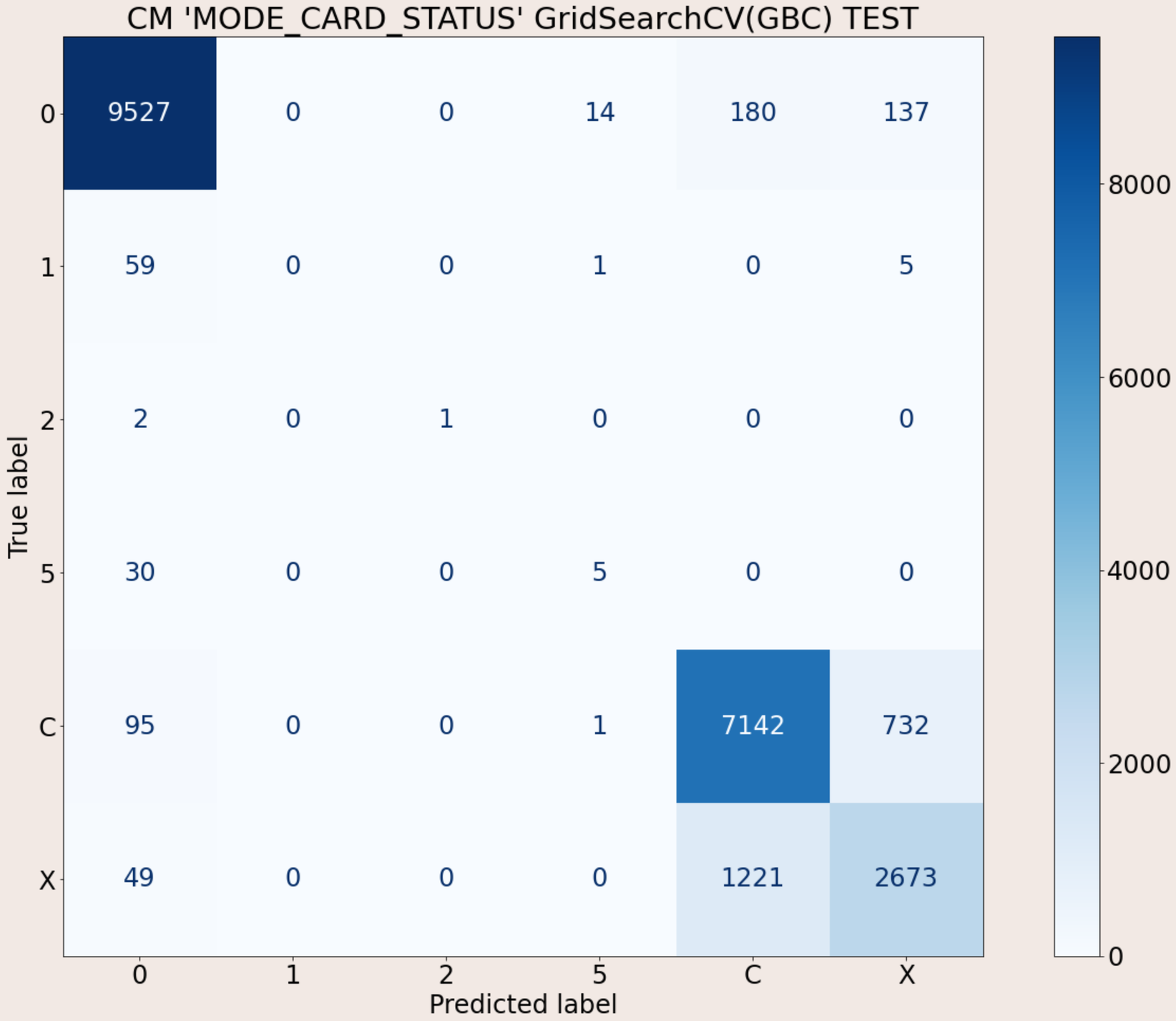
-Gradient Boosting Regressor

-Used GridSearchCV for hyperparameter tuning

- Hyperparameters:
  - min_samples_split (5)
  - min_samples_leaf (6)
  - max_depth (4)
  - loss (Squared Error)
  - learning_rate (.3)
  - n_estimators (150)
  - criterion (Squared Error)
  - tol (1e-3)
  - validation_fraction (.1)
  - n_iter_no_change (2)



Mode Card Status



Mode Card Status

-Gradient Boosting Classifier

- Used GridSearchCV for hyperparameter tuning

- 88% accuracy

- Hyperparameters:

  - min_samples_split (7)
  - min_samples_leaf (7)
  - max_depth (3)
  - loss (Squared Error)
  - learning_rate (.3)
  - n_estimators (150)
  - criterion (Squared Error)
  - tol (1e-3)
  - validation_fraction (.1)
  - n_iter_no_change (2)

Problems and Limitations

- Our model had a hard time predicting values between 30-70 for percent missing

- Data may not be able to tell us the values we were looking for, the featured engineered columns

- Data is unbalanced for classification but we are able to get clear predictions. There is an absence of data from the card status 1-5

Conclusions

-Majority of card holders miss all or none of their payments

-Salary, Property Type, Age, Days Employed, Has E-mail, Length of Card Holding, and Mode Card Status are important when analyzing cardholders

Thank you

Any Questions?