Mitigating Credit Risk: A Predictive Model for Charge Off

by Random Seed
Xin Guo, Xinya Dan, Xinkai Zhao, Yumeng Li

Introduction

Discoveries

Gradient Boosted Model

Perks

Credit charge offs is an essential factor to institution's financial stability. It occurs when a customer fails to makepayment and eventually is written off as a loss. Therefore, it is becoming more and more crucial for banks to make precise predictions about their charge offs to make informed decisions for the future.

Problem Setting

Dataset

EDA

Problem Setting

Survival analysis is a statistical method used to analyze time-to-event data. It can be used to model the probability of an event occurring at a particular time. To predict the number of charge-offs in the future, it is sensible to first estimate the likelihood of a customer being charged off at a given duration from their snapshot date, then the difference between total obs and summation of the estimated likelihood at each time is the number of charge-offs.

Credit charge offs is an essential factor to institution's financial stability. It occurs when a customer fails to makepayment and eventually is written off as a loss. Therefore, it is becoming more and more crucial for banks to make precise predictions about their charge offs to make informed decisions for the future.

Problem Setting

Dataset

EDA

Dataset

Customer dataset : 5 million records of customers' monthly financial status from 2018-19

Macro dataset: macro ecomonic conditions in a monthly basis

Forecast dataset: starting point to predict the charge off in the following 12 month.

Credit charge offs is an essential factor to institution's financial stability. It occurs when a customer fails to makepayment and eventually is written off as a loss. Therefore, it is becoming more and more crucial for banks to make precise predictions about their charge offs to make informed decisions for the future.

Problem Setting

Dataset

EDA

Proportion of Charge Off

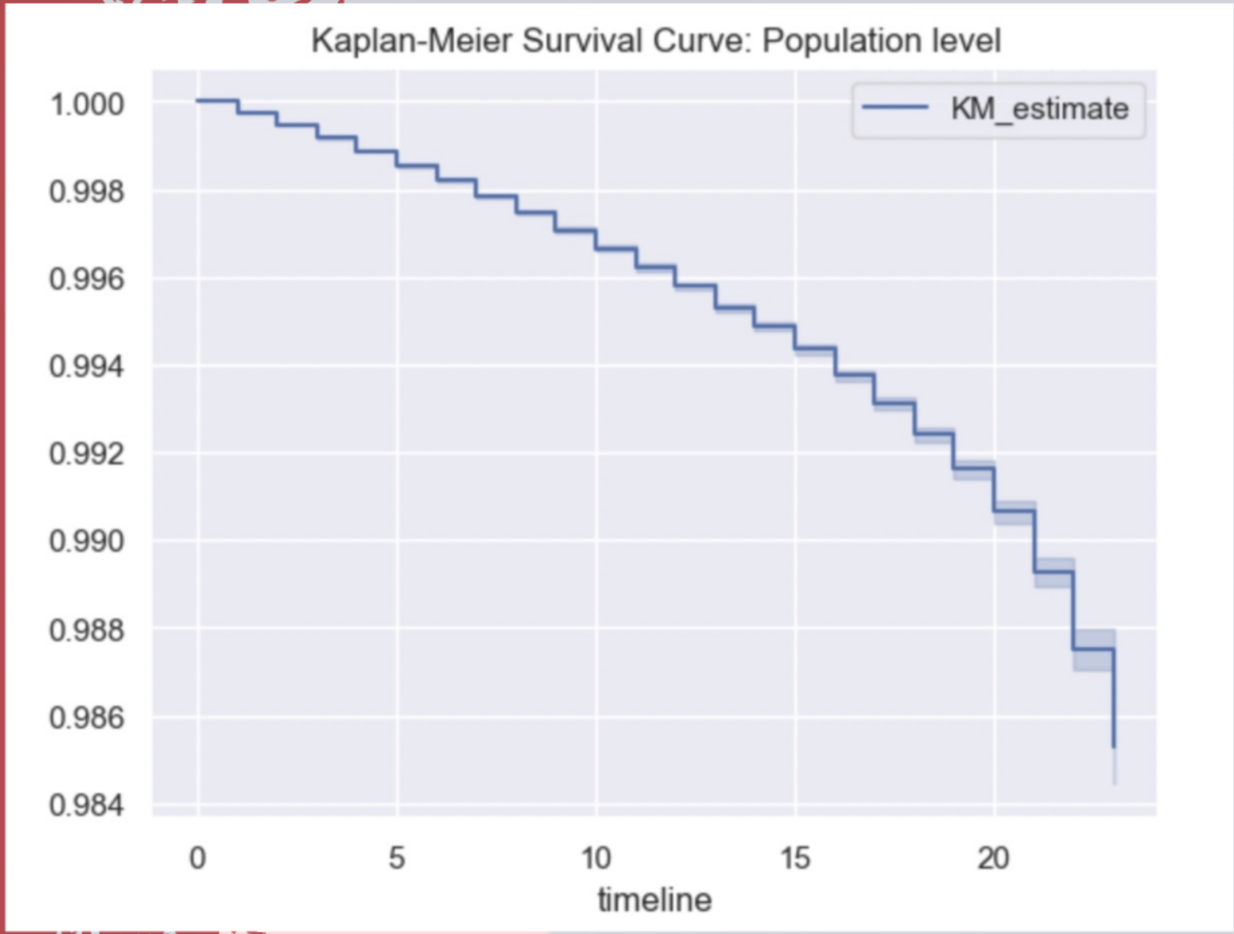
KM Plot

KM Plot

According to the KM plot drawn based on the charge off status and timeline, the charge off rate decreases with the time flatly

KM Plot

According to the KM plot drawn based on the charge off status and timeline, the charge off rate decreases with the time flatly



KM Plot

According to the KM plot drawn based on the charge off status and timeline, the charge off rate decreases with the time flatly

Proportion of Charge Off

KM Plot

Proportion of Charge Off

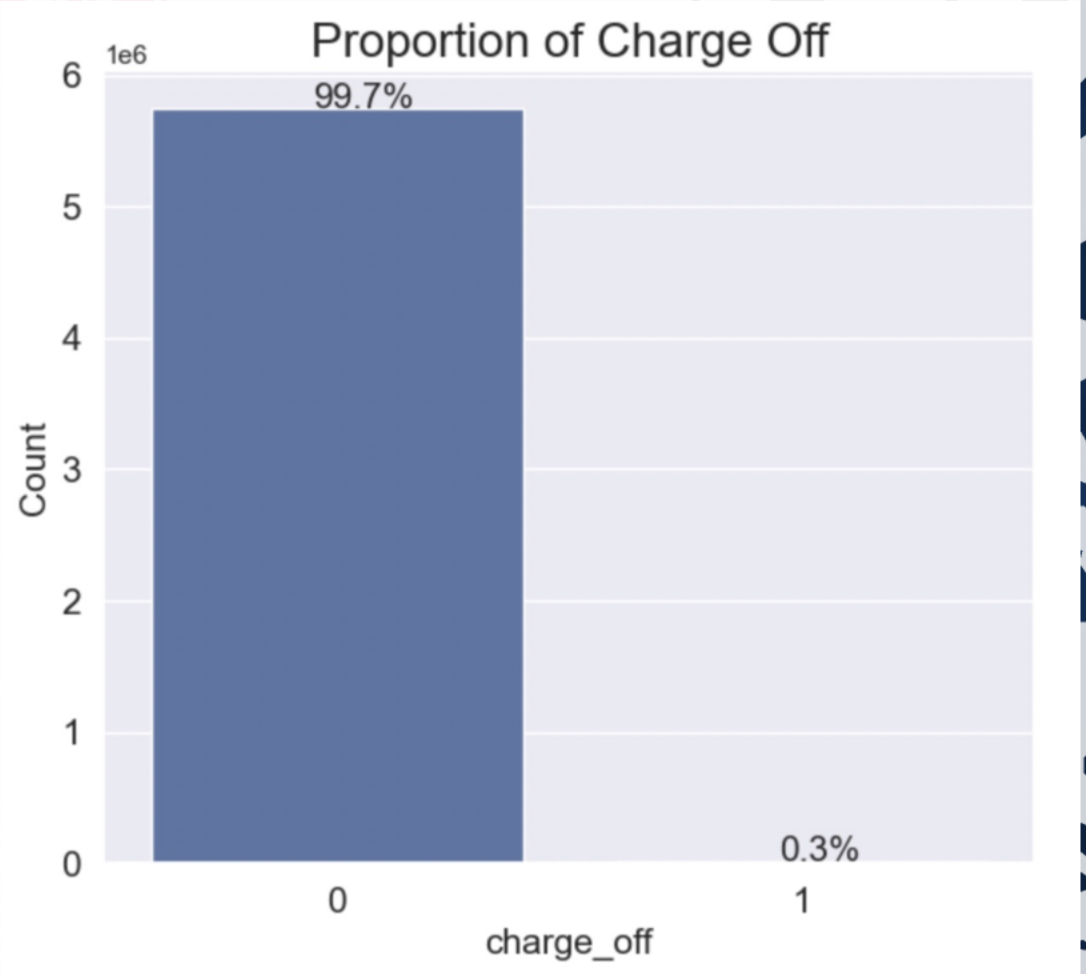
In the "charge off" colomn, 0 accounts for 99.7% percentage, which means most of the consumers are not charged off at the given point of time

Proportion of Charge Off

In the "charge off" colomn, 0 accounts for 99.7% percentage, which means most of the consumers are not charged off at the given point of time

Proportion of Charge Off

In the "charge off" colomn, 0 accounts for 99.7% percentage, which means most of the consumers are not charged off at the given point of time



Proportion of Charge Off

In the "charge off" colomn, 0 accounts for 99.7% percentage, which means most of the consumers are not charged off at the given point of time

Proportion of Charge Off

KM Plot

Credit charge offs is an essential factor to institution's financial stability. It occurs when a customer fails to makepayment and eventually is written off as a loss. Therefore, it is becoming more and more crucial for banks to make precise predictions about their charge offs to make informed decisions for the future.

Problem Setting

Dataset

EDA

Problem Setting

edit charge offs is an
essential factor to
nstitution's financial
bility. It occurs when a
customer fails to
epayment and eventually
written off as a loss.
fore, it is becoming more

Credit charge offs is an essential factor to institution's financial stability. It occurs when a customer fails to makepayment and eventually is written off as a loss. Therefore, it is becoming more and more crucial for banks to make precise predictions about their charge offs to make informed decisions for the future.

Problem Setting

Dataset

EDA

Mitigating Credit Risk: A Predictive Model for Charge Off

Introduction

Discoveries

by Random Seed

Xin Guo, Xinya Dan, Xinkai Zhao, Yumeng Li

Gradient Boosted Model

Perks

A gradient boosted model is similar to a Random Survival Forest, in the sense that it relies on multiple base learners to produce an overall prediction

Strengths:

- CoxPH regression makes assumptions about the shape of the hazard function, such as the proportional hazards assumption, which may not hold in some cases. GBMs, on the other hand, are non-parametric and can capture complex, non-linear relationships between predictors and the outcome.
- GBMs can handle complex interactions between predictors, which CoxPH cannot. This makes GBMs more suitable for analyzing high-dimensional data with many interacting predictors.

Mitigating Credit Risk: A Predictive Model for Charge Off

by Random Seed
Xin Guo, Xinya Dan, Xinkai Zhao, Yumeng Li

Introduction

Discoveries

Gradient Boosted Model

Perks

Perks

Improve data quality

Improve efficiency

Hyper-parameter Tuning

Adapted to skewed data

Missing Values

- delete columns with 90%+ missing
- impute one column with mode
- delete obs containing 4 missing values in one column

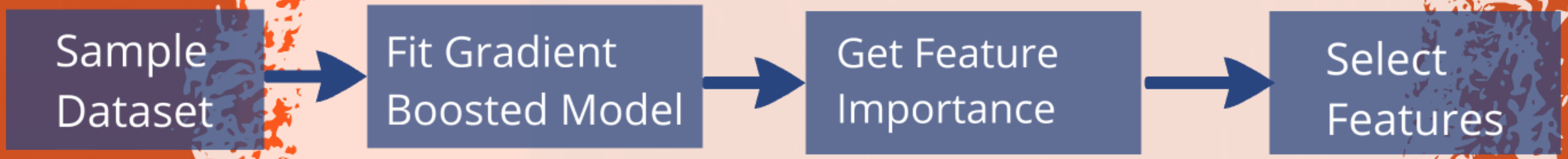
Encode Categorical

- encode with one hot encoder
- check for new/ missing levels in forecast data (account_status_code miss MonthEnd & Clchange--> create columns with 0)

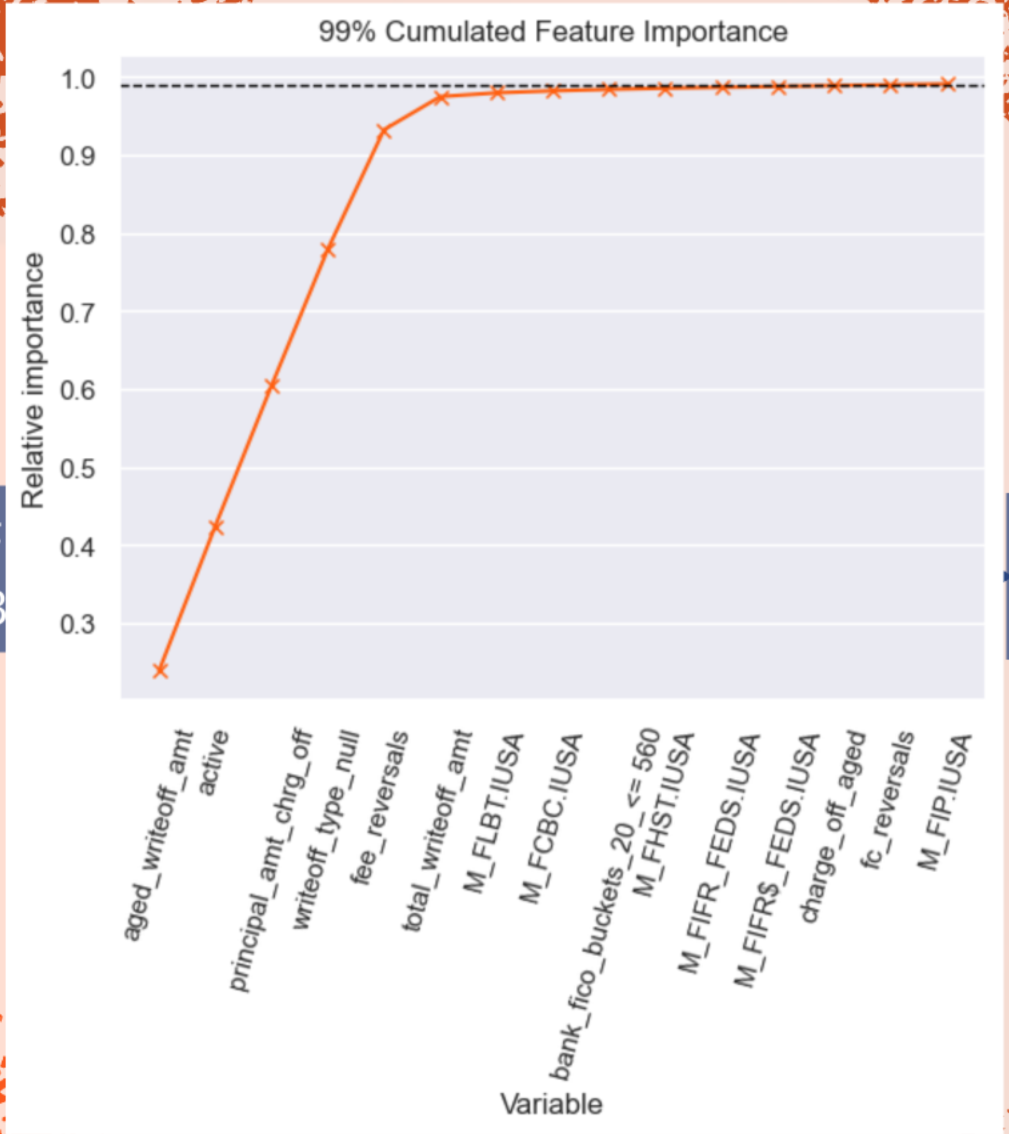
Perks

- Improve data quality
- Improve efficiency
- Hyper-parameter Tuning
- Adapted to skewed data

Feature Selection



Sample Dataset



Select Features

Perks

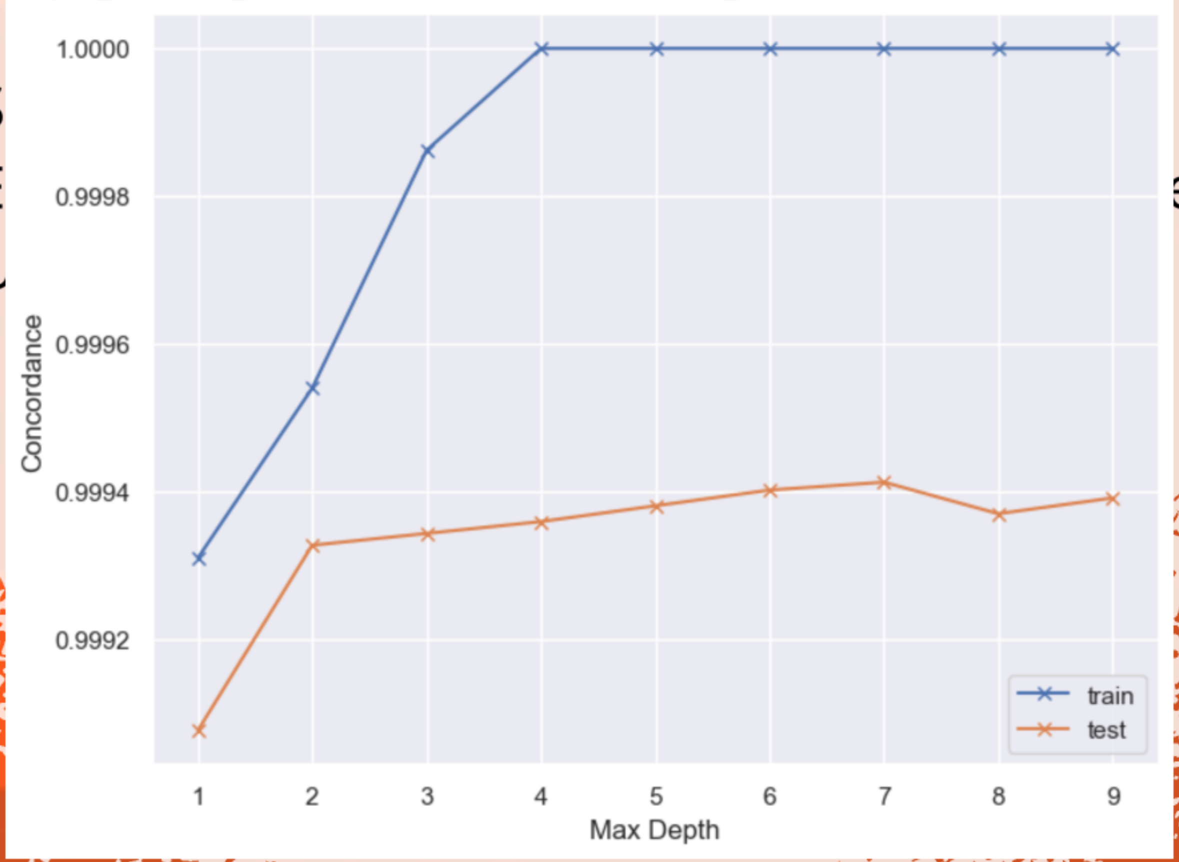
- Improve data quality
- Improve efficiency
- Hyper-parameter Tuning
- Adapted to skewed data

Hyperopt package for hyperparameter optimization

- Select which model parameters to fine-tune
- Explore reasonable range for each parameter
- Use Hyperopt to tune and record history
  - Advantage over grid/random search: informed way to find parameters using Bayesian approach
  - Built-in function to minimize negative concordance index

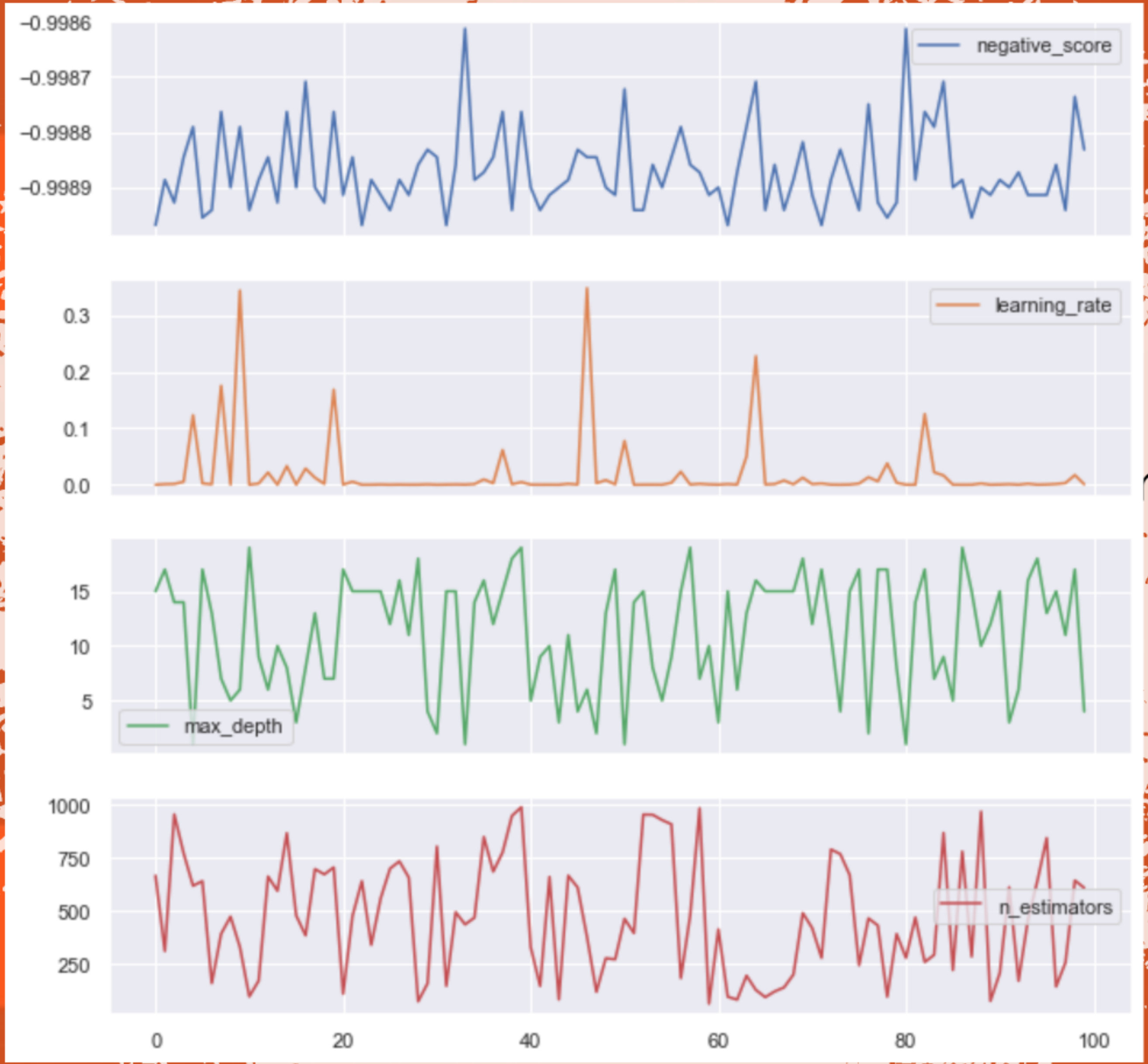
Hyperopt package for hyperparameter optimization

- S
- E … er
- U



Hyperopt package for hyperparameter optimization

- Select which model parameters to fine-tune
- Explore reasonable range for each parameter
- Use Hyperopt to tune and record history
    - Advantage over grid/random search: informed way to find parameters using Bayesian approach
    - Built-in function to minimize negative concordance index



−0.9986
−0.9987
−0.9988
−0.9989
negative_score
0.3
0.2
0.1
0.0
learning_rate
15
10
5
max_depth
1000
750
500
250
n_estimators
0
20
40
60
80
100

Hyperopt package for hyperparameter optimization

- Select which model parameters to fine-tune
- Explore reasonable range for each parameter
- Use Hyperopt to tune and record history
  - Advantage over grid/random search: informed way to find parameters using Bayesian approach
  - Built-in function to minimize negative concordance index

Perks

- Improve data quality
- Improve efficiency
- Hyper-parameter Tuning
- Adapted to skewed data

Use "stratify" in train-test split to keep the proportion of charge offs the same in both training and testing sets

Perks

Improve data quality

Improve efficiency

Hyper-parameter Tuning

Adapted to skewed data

Mitigating Credit Risk: A Predictive Model for Charge Off

by Random Seed

Xin Guo, Xinya Dan, Xinkai Zhao, Yumeng Li

Introduction

Discoveries

Gradient Boosted Model

Perks

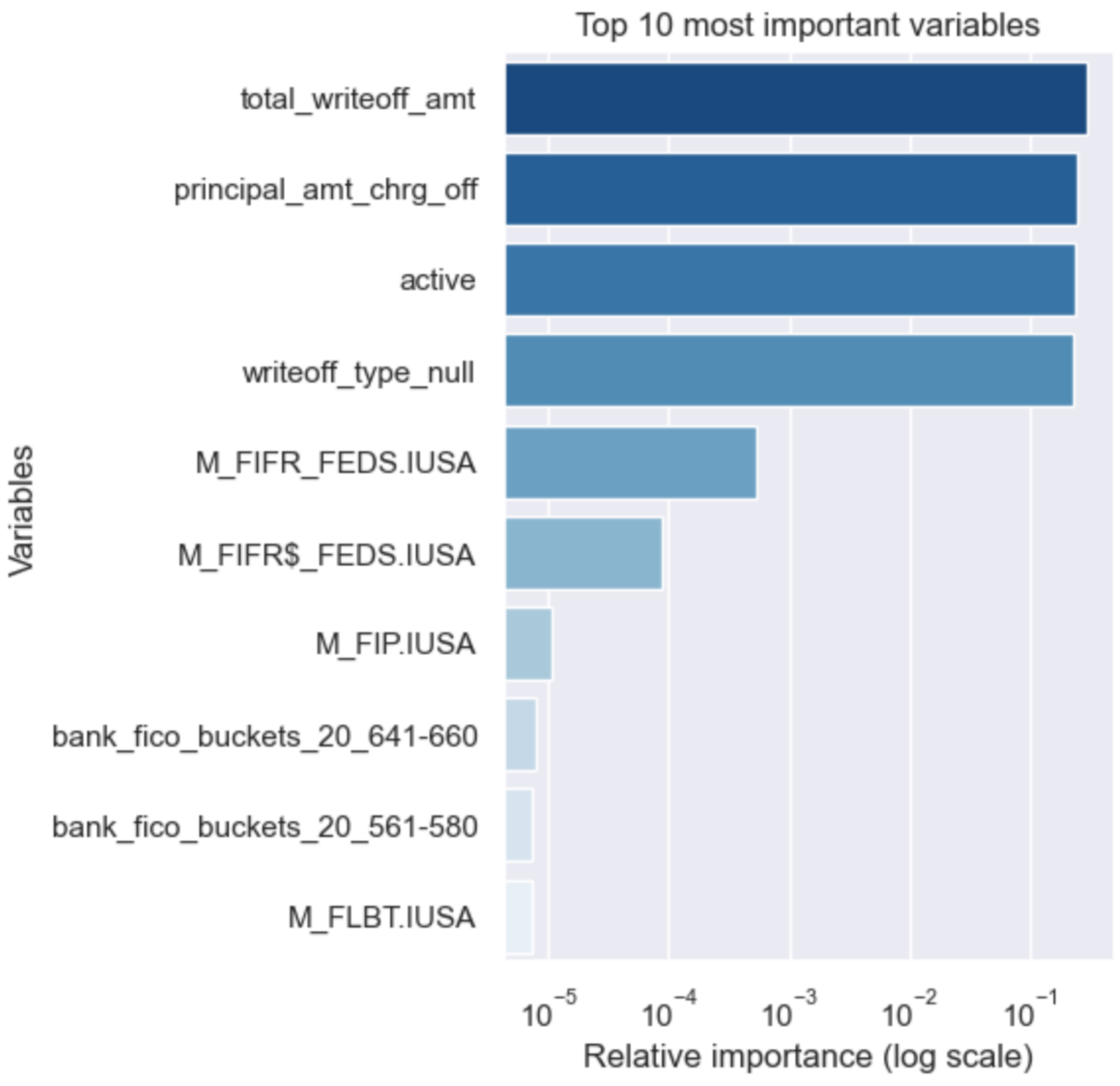
Feature Importance

Probability Prediction

Number of Charge offs Prediction

The probability of being charged off over time depends more heavily on customer's financial status than macroeconomic factors.

However the standard deviation of these macroeconomic factors is relatively low during 2018 and 2019, the impact of macroeconomic factor may be underestimated using the current training data.

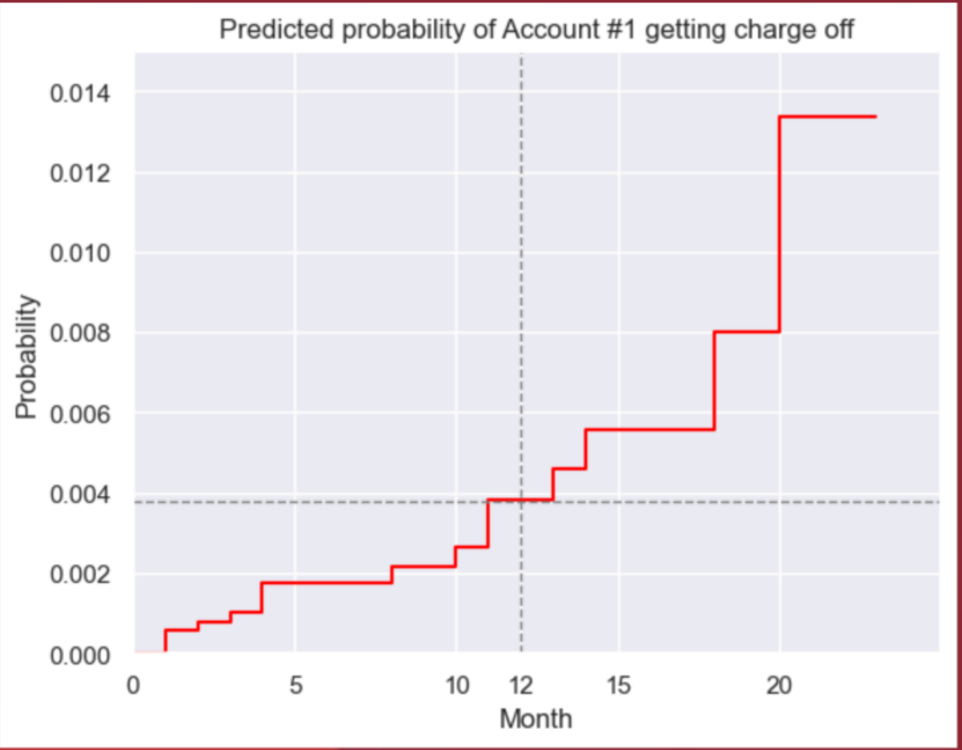


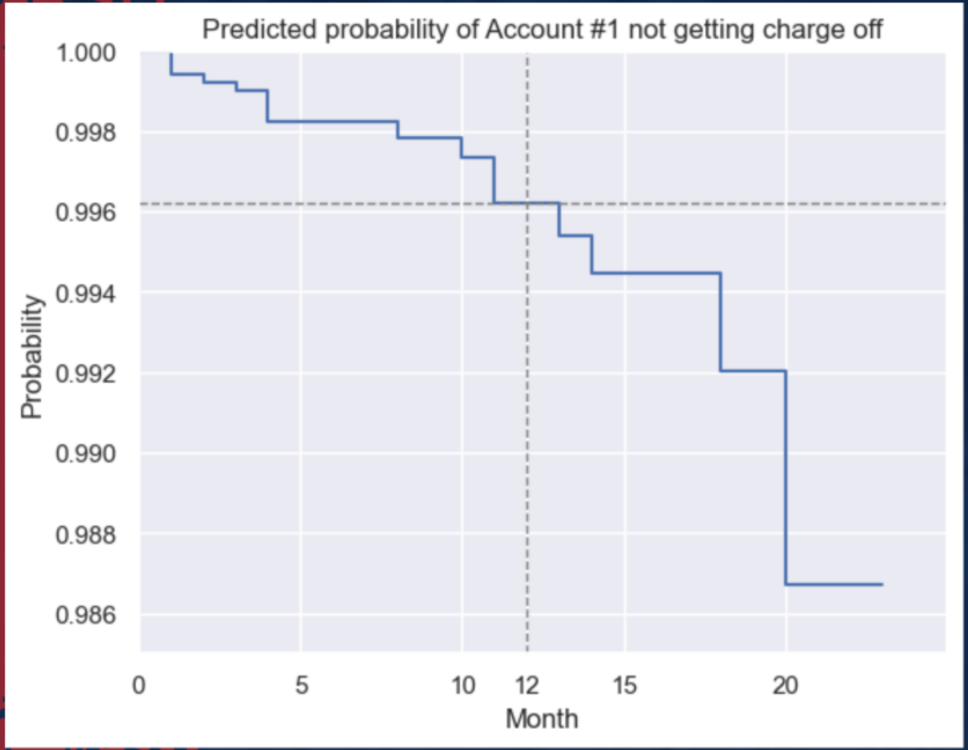
Feature Importance

Probability Prediction

Number of Charge offs Prediction

Model explainability

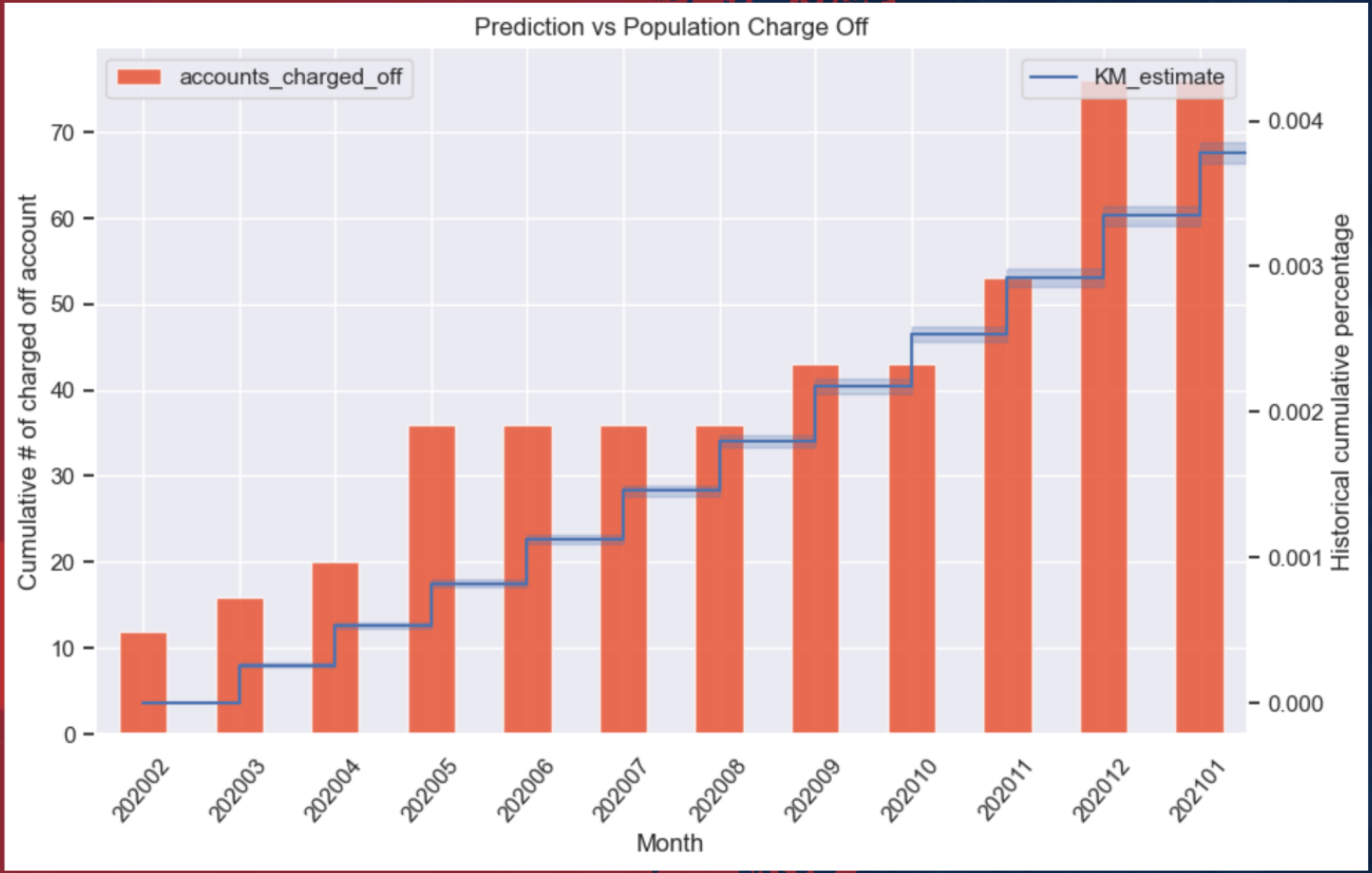




Feature Importance

Probability Prediction

Number of Charge offs Prediction



Prediction vs Population Charge Off
accounts_charged_off
KM_estimate
Cumulative # of charged off account
Historical cumulative percentage
0
10
20
30
40
50
60
70
0.000
0.001
0.002
0.003
0.004
202002
202003
202004
202005
202006
202007
202008
202009
202010
202011
202012
202101
Month


Feature Importance
Probability Prediction
Number of Charge offs Prediction

Mitigating Credit Risk: A Predictive Model for Charge Off

by Random Seed

Xin Guo, Xinya Dan, Xinkai Zhao, Yumeng Li

Introduction

Discoveries

Gradient Boosted Model

Perks