MODEL RISK MANAGEMENT DOCUMENT

Credit Line Decrease Model: Version 1

Model Group : Customer Valuation Model

Date Created : 20 Feb 2021

Developer : Harsh Deep Verma, Mayank Jain

Reviewer : Corridor

Last Review Date : 25 March 2023

Table of Content

Model Risk Evaluation

Model risk occurs primarily for two reasons:

- A model may have fundamental errors and produce inaccurate outputs when viewed against its design objective and intended business uses
- A model may be used incorrectly or inappropriately or there may be a misunderstanding about its limitations and assumptions. Model risk increases with greater model complexity, higher uncertainty about inputs and assumptions, broader extent of use, and larger potential impact.

Overview

The Credit Line Decrease (CLD) model is used to identify high risks accounts with the objective to mitigate potential losses by decreasing their credit limit. The model will be used on all the RRB credit card customers.

RRB branded credit card is a core product of RRB and accounts for ~$90 million ANR with ~20 million open accounts till Dec’22.

The model has been built using Dec 2018 development vintage. February 2018 and April 2019 has been used for out-of-time (OOT) validations. The model has been built on entire card customers except for the customers whose past 12 months historical data is not known.

The model estimates the likelihood of an account having a status of 60+ days past due at the end of 18 months (classified as ‘bad’ hereafter in this document).

The model has been built on both internal as well as external data. The external credit bureau data is received monthly through a batch process for the Existing Card Members (ECM). In addition, the model leverages daily data – daily bureau (in addition to batch bureau), daily triggers and a few daily on-us attributes – to allow the identification of high risk accounts before they utilize the remaining open-to-buy amount.

Model Overview

The CLD model is developed as an account level score and scores all the ECM consumer accounts. It is developed using on-us ECM information (daily and cycle-end), batch bureau (monthly) and Bureau trigger (daily) information. The use of daily data in the model development allows the score to adjust according to the changes observed at a daily level and hence capturing the accounts with an early sign of financial distress.

The model development went through four major stages:

- Vintage Selection and Performance Definition
- Data Preparation
- Model Development and Validation
- Documentation

Vintage Selection

The choice of vintages used for model development is governed by the following key factors:

- Data availability: The model makes use of daily CLD extracts as development base, which is available as archives and is retained for 24 months. Further, the model uses daily bureau attributes.
- Recency: Choosing a recent vintage to incorporate recent patterns related to credit risk behavior. Also, recency in vintage ensures better representation of portfolio performance.
- Forward looking: Vintages selected are representative of the forward looking state of the portfolio. This has been demonstrated with a validation on most recent vintage of April 2019.

The final vintages used for the Model development and validations are:

- Development: Dec’2018
- Back-Testing: Feb’2018
- Validation: Apr’2019
- Early Warnings: Dec’2019

Performance Definition:

The model uses performance definition of ‘Bucket 3+ at the end of 18 months’. This definition has been obtained on the basis of a business analysis conducted with a terminal window of 24 months. A greater emphasis has been made on keeping the false positive rate lower given that Credit Line Decrease is a negative action to a customer. The business analysis is also supported by statistical analysis on the choice of target variable definition. In addition, forbearance, re-age and settlements in 18 months has been classified as bads.

To minimize overlap between Good/Bad characteristics, ‘indeterminate has been assigned in the model development. Accounts with Bucket 2 at the end of performance window has been tagged as ‘indeterminate’ as accounts have higher false positive rate (~50%), but belongs to a negative segment.

Performance Definition (in 18 months performance window)	Good/Bad/Indeterminate
Bucket 3+ at the end of 18 months or Forbearance/Charge-offs/Re-age/Bankrupt/Settlement in 18 months	Bad
Bucket 2 at the end of 18 months	Indeterminate
Remaining	Good

Data Preparation

Dataset Creation

The model has been developed using the extracts from the Card Master File (CMF). Given that the development of the model was priority based for the business based on the impact analysis of quick model developed earlier, a limited set of features were used from a previous model (APD model). This was done to facilitate quick development and implementation to suit business needs.

Given that the APD model uses attributes from varied behavior dimensions, all the dimensions were captured for this model as well. The model also uses on-us data along with the bureau data. This data was used to create trended derived variables and was used in model development. Daily on-us variables and daily bureau trigger were also used to capture daily variations in the customer characteristics.

The common variables include delinquency status, days since last delinquent, months on book, closing balances, recent behavior of accounts, available credit, auto pay enabled, payment method, card type etc.

Data Quality Check

Checks were created at different stages of the model dataset creation in order to ensure the integrity in the final model development as well as model validation datasets. A few of the checks undertaken are described below:

- Check for Volume Consistency: Number of observations pulled for every vintage was cross-checked for consistency at every stage of the data pull process before merging and after roll-up. At each stage of data pull, the number of accounts were reconfirmed to ensure and investigate the reason of drop in the number of accounts, if any
- Univariate Distribution: Variable distribution of all samples have been checked. Each variables percentile information, count of missing, minimum, maximum and standard deviation were checked. Key

and eventually charge-off. Conventionally, CLD program has been leveraging portfolio level behavior risk score, FICO score along with a set of rule based criteria to identify high risk customers. However, machine learning technique captures non-linear complex relationships more effectively as compared to simple linear techniques like logistic regression. The score is computed at an account level and developed using XGBoost (Extreme Gradient Boosting) technique leveraging all the different type of data.

The model output will be used by the policy team along with other rules to screen the high risk customers to decrease their credit limits to the current balances.

Model Methodology

XGBoost and Logistic Regression are two techniques used for creating the model along with GBM trees. The final model is built using XGBoost and this is determined based on the performance benefits observed in using XGBoost. It is observed that XGBoost outperforms both the other models on all the important business metrics and does so consistently across a wide range of data points.

Extreme Gradient Boosting Machines (XGBoost) has been used for model development. XGBoost is a widely used technique for both Classification and Regression outcomes. It is a stage-wise additive modelling technique, which improves upon predictions by building decision trees on residuals remaining at any stage. For a classification outcome, XGBoost is built based on a Bernoulli distribution and optimizes for a log loss function. The residuals in a classification problem are defined by gradients of the log loss function. At each stage, XGBoost uses decision trees to predict the remaining residuals. An extensive hyper-parameter tuning was performed in a Grid search pattern to obtain the optimal set of parameters.

It should be noted that the model does not make use of segmentation since XGBoost is an ensemble of decision trees itself and thus is capable of finding interactions without the explicit use of segmentation.

XGBoost technique allows for multiple hyper-parameters to control and fine-tune the growth of a XGBoost model. An extensive grid search was performen on key hyper-parameters and the performance was analyzed. The hyper-parameters used in tuning process were:

- Number of trees (n trees)
- Depth of a tree (interaction depth)
- Learning rate (shrinkage)

All the metrics suggests satisfactory performance across different business metrics.

Ongoing Model Monitoring and Governance Process

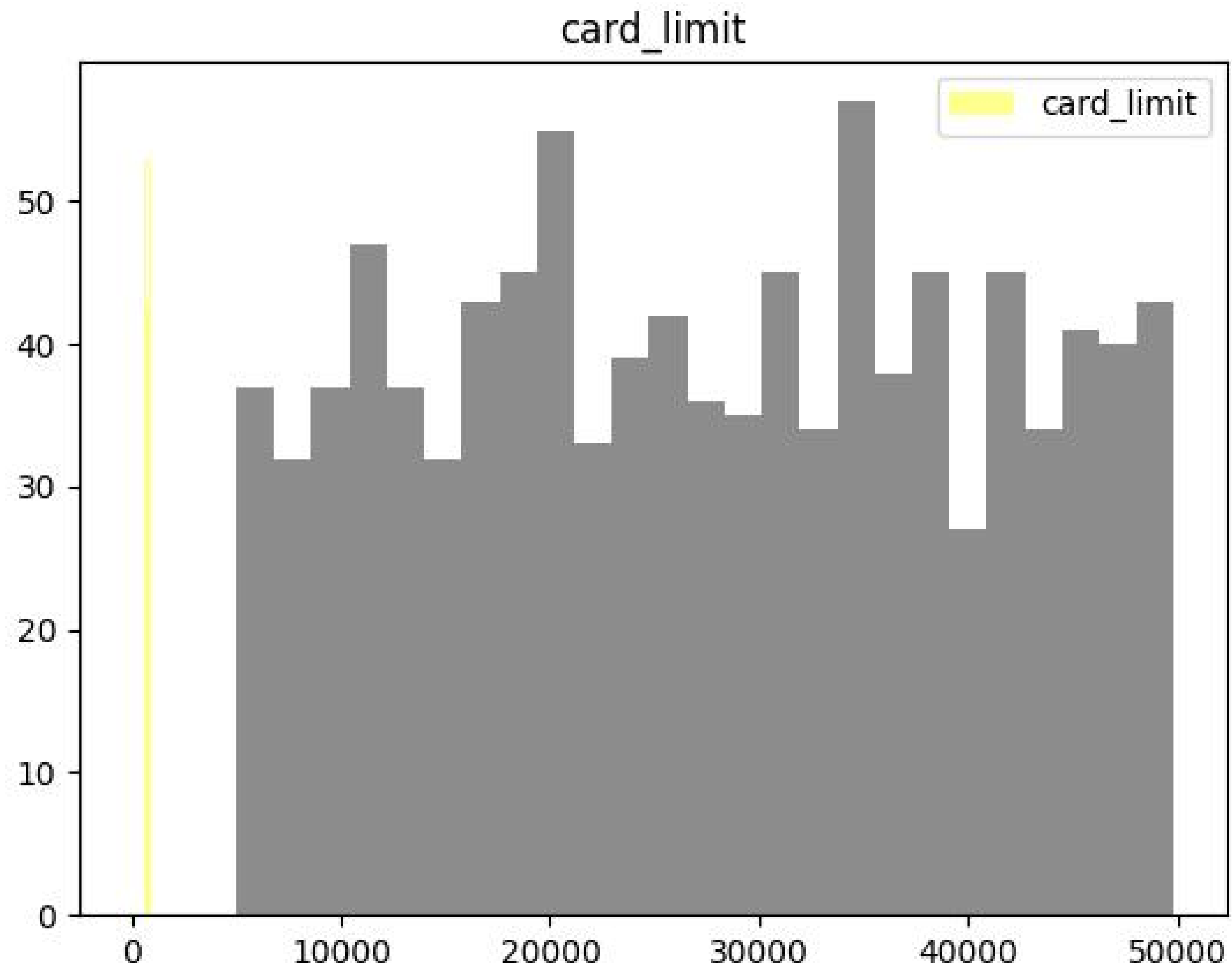
Based on the last annual model review exercise, MRR (Model Risk Rating) of the model is high. Hence, monitoring frequency of the model is Quarterly.

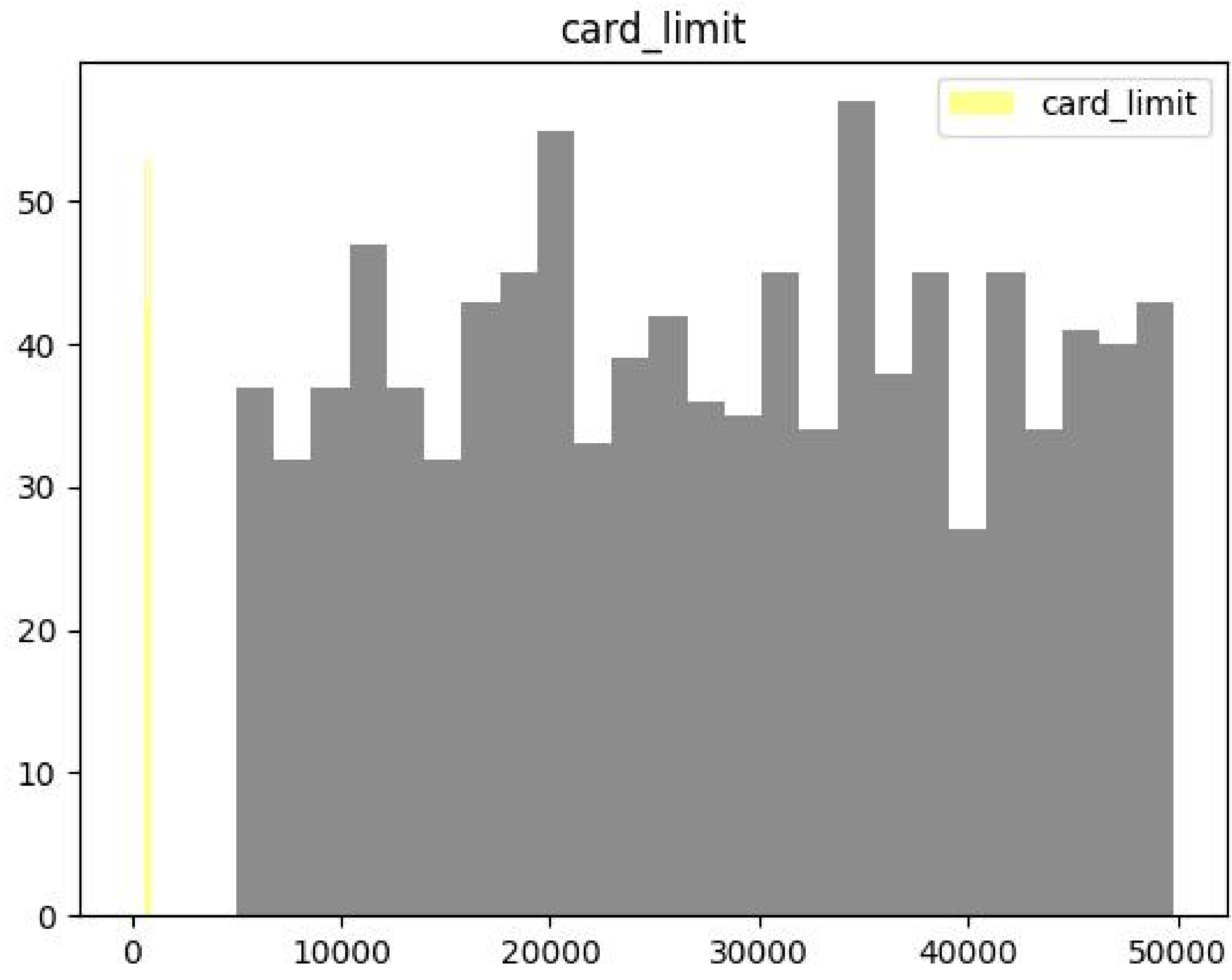
All the model performance metrics defined by MRM for consumer valuation model are monitored based on the frequency defined by MRM to ensure the robustness of the model

Annual Model Review Plan

As per the MRM policy requirement, model should be reviewed annually to ensure the model usage for the next year. This year model was due for revalidation, hence re-validation is performed and submitted at place of Annual Model Review (AMR). An AMR will be submitted for the model in the next annual validation exercise to be held in October this year on the latest available performance results during the validation.

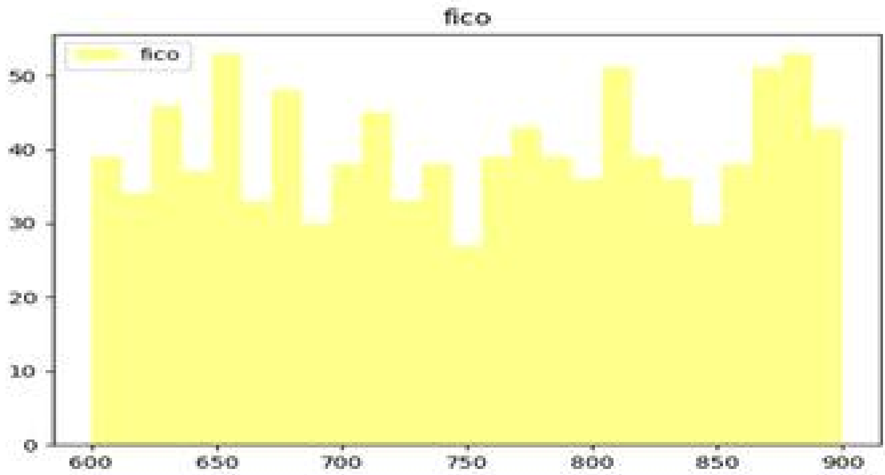
The performance checks would include metrics like PSI, KS, ROB, MAPE etc. and all MRM required metrics for AMR.


card_limit
card_limit
0
10
20
30
40
50
0
10000
20000
30000
40000
50000

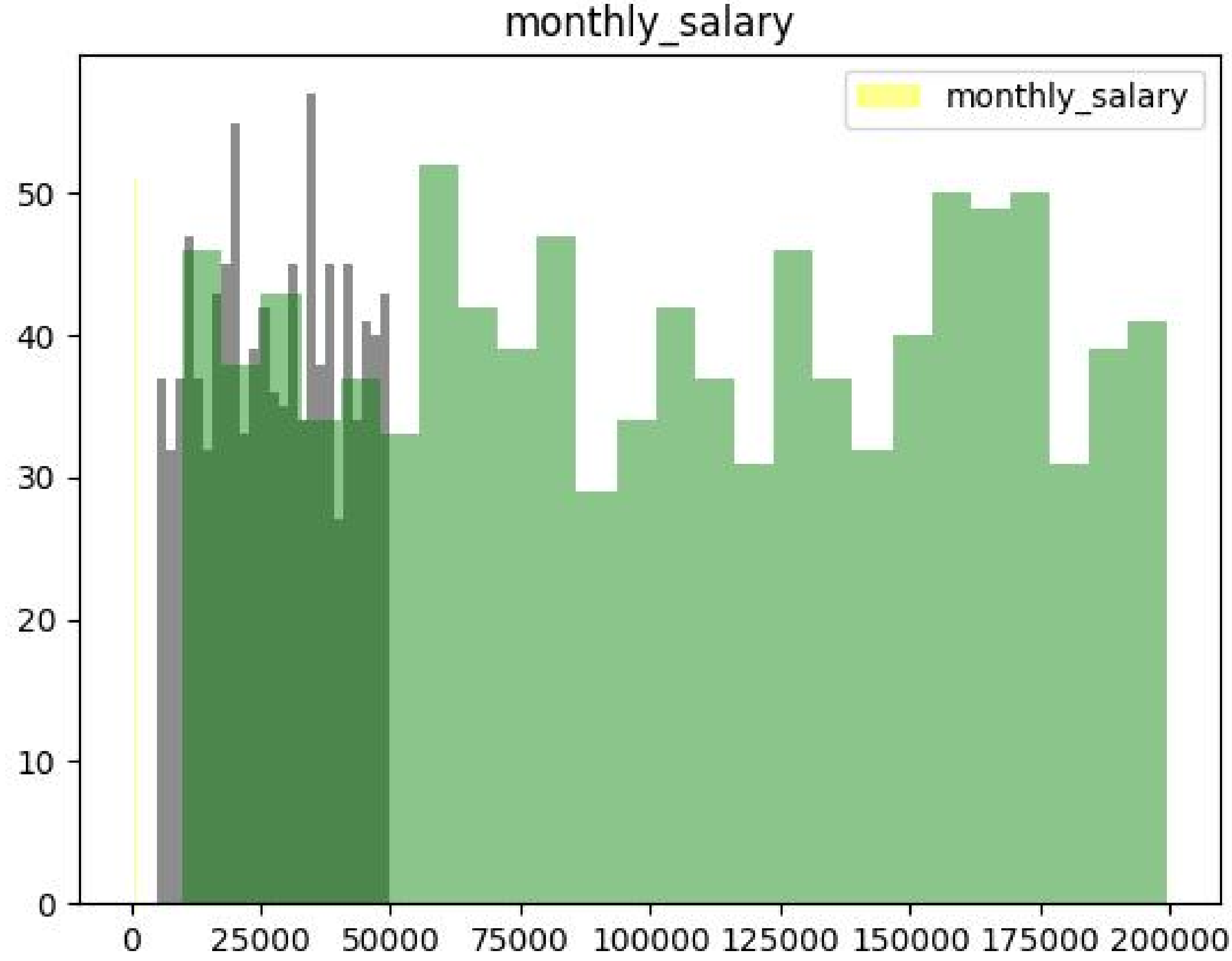


card_limit
card_limit
0
10
20
30
40
50
0
10000
20000
30000
40000
50000

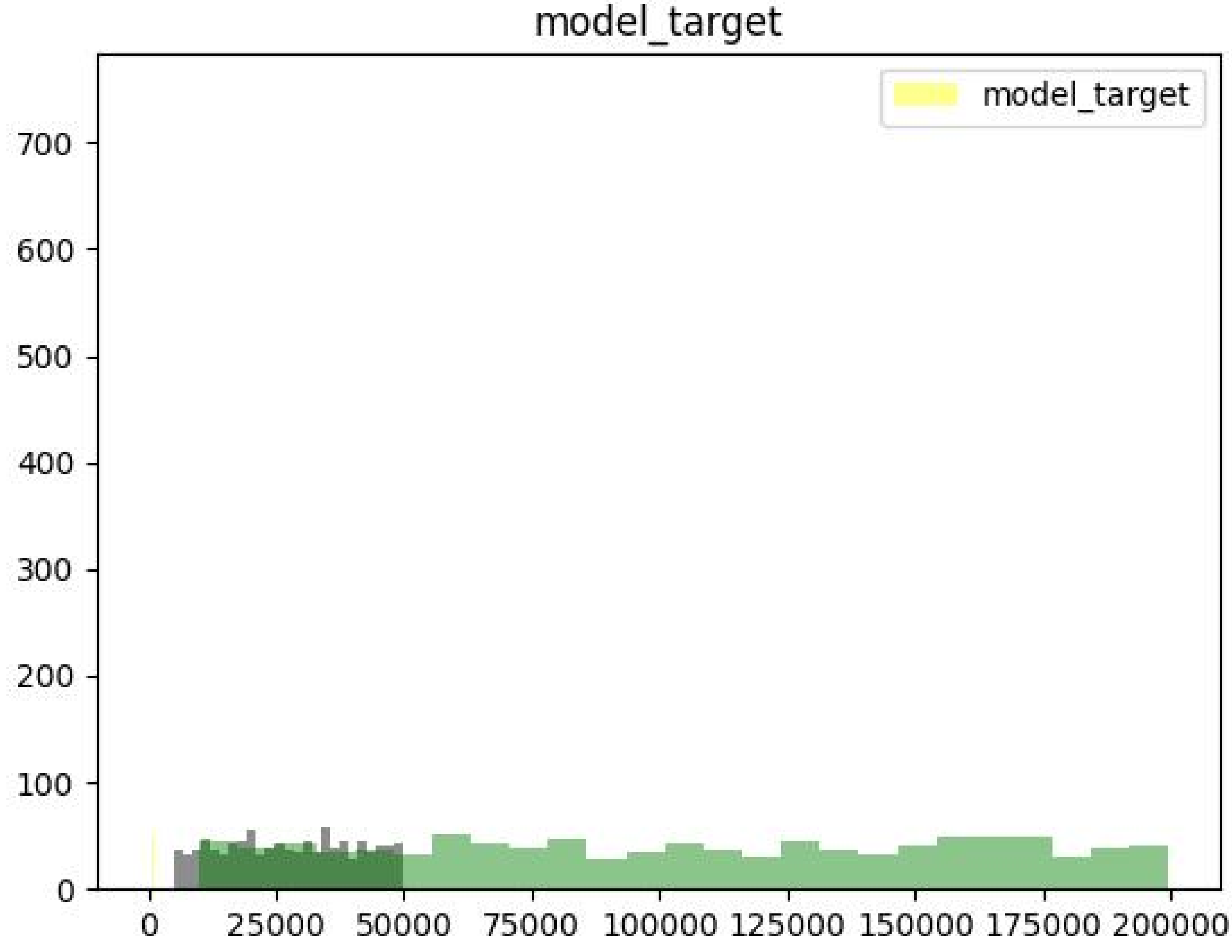
monthly_salary	fico	utilization	card_limit	card_interest_rate	model_output	model_target
78780.0	897.0	0.1316805607934714	31032.0	0.03746681372688547	0.223082181015947	0.0
10065.0	794.0	0.06219744898083074	31697.0	0.03408024180121726	0.9543139003535083	1.0
97434.0	880.0	0.08442595468228653	48998.0	0.09675441759626811	0.990619741843032	0.0
126755.0	729.0	0.0072965409391748115	38827.0	0.04142487710731699	0.023173546930568723	0.0
70696.0	624.0	0.15470927524805642	9082.0	0.05506184729576953	0.10753385389263637	0.0



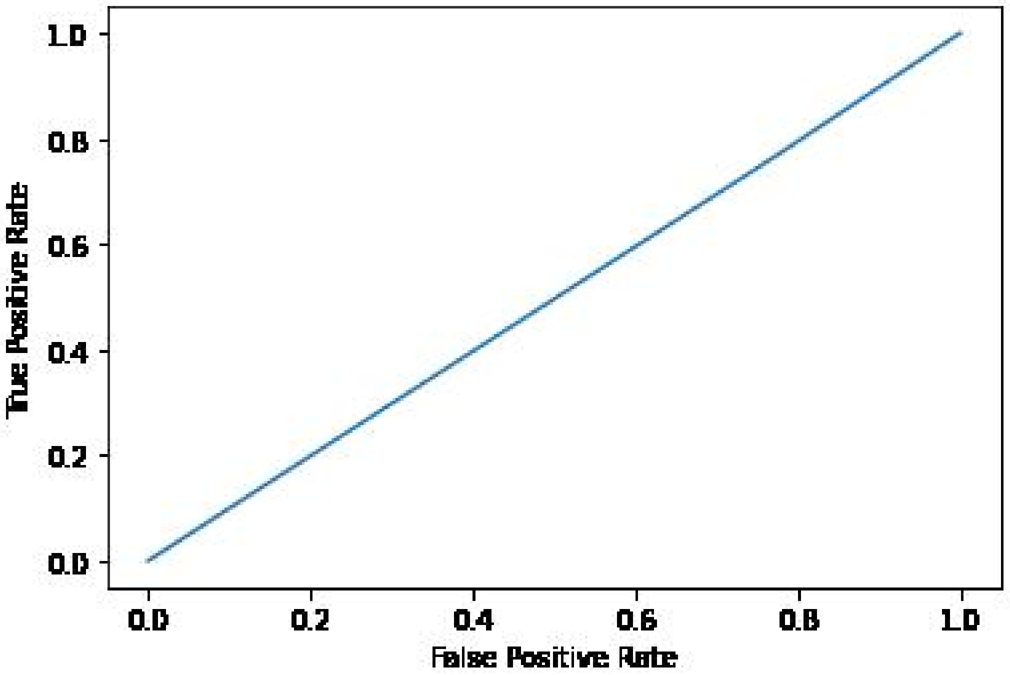
fico
fico
0
10
20
30
40
50
600
650
700
750
800
850
900



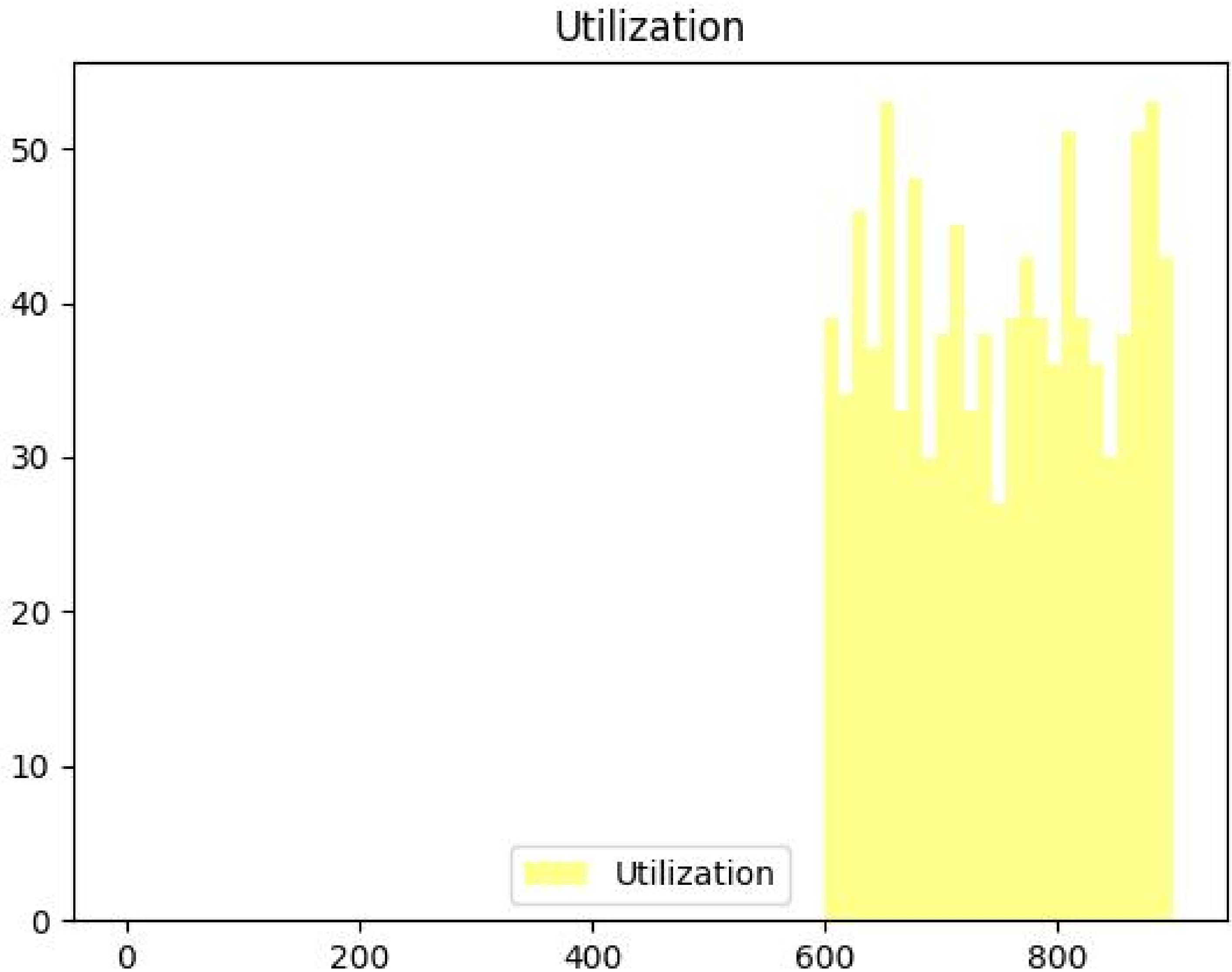
monthly_salary
monthly_salary
0
10
20
30
40
50
0
25000
50000
75000
100000
125000
150000
175000
200000

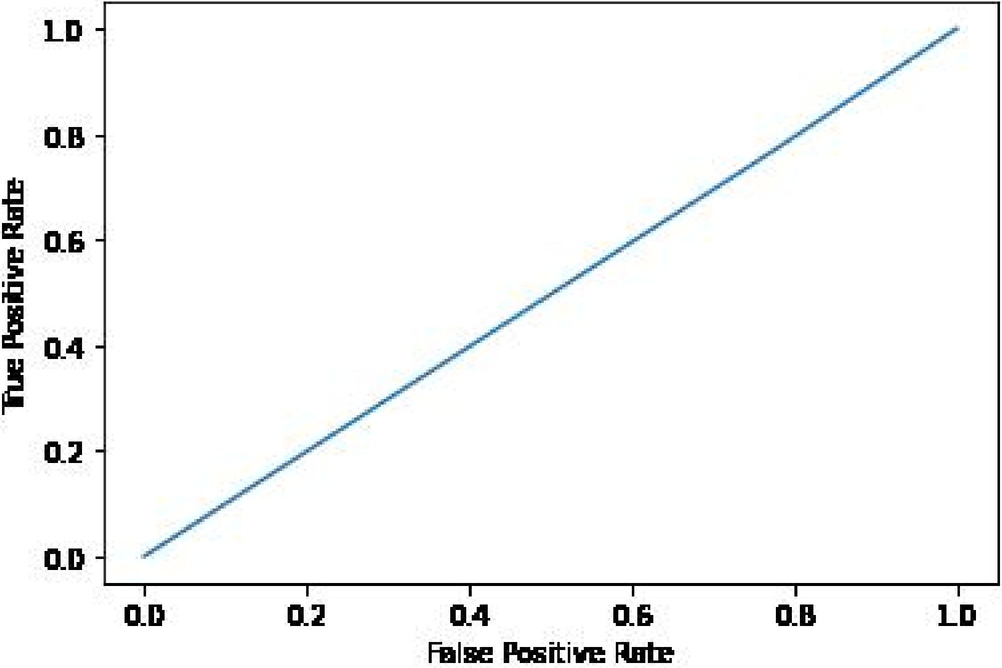

model_target
model_target
700
600
500
400
300
200
100
0
0
25000
50000
75000
100000
125000
150000
175000
200000

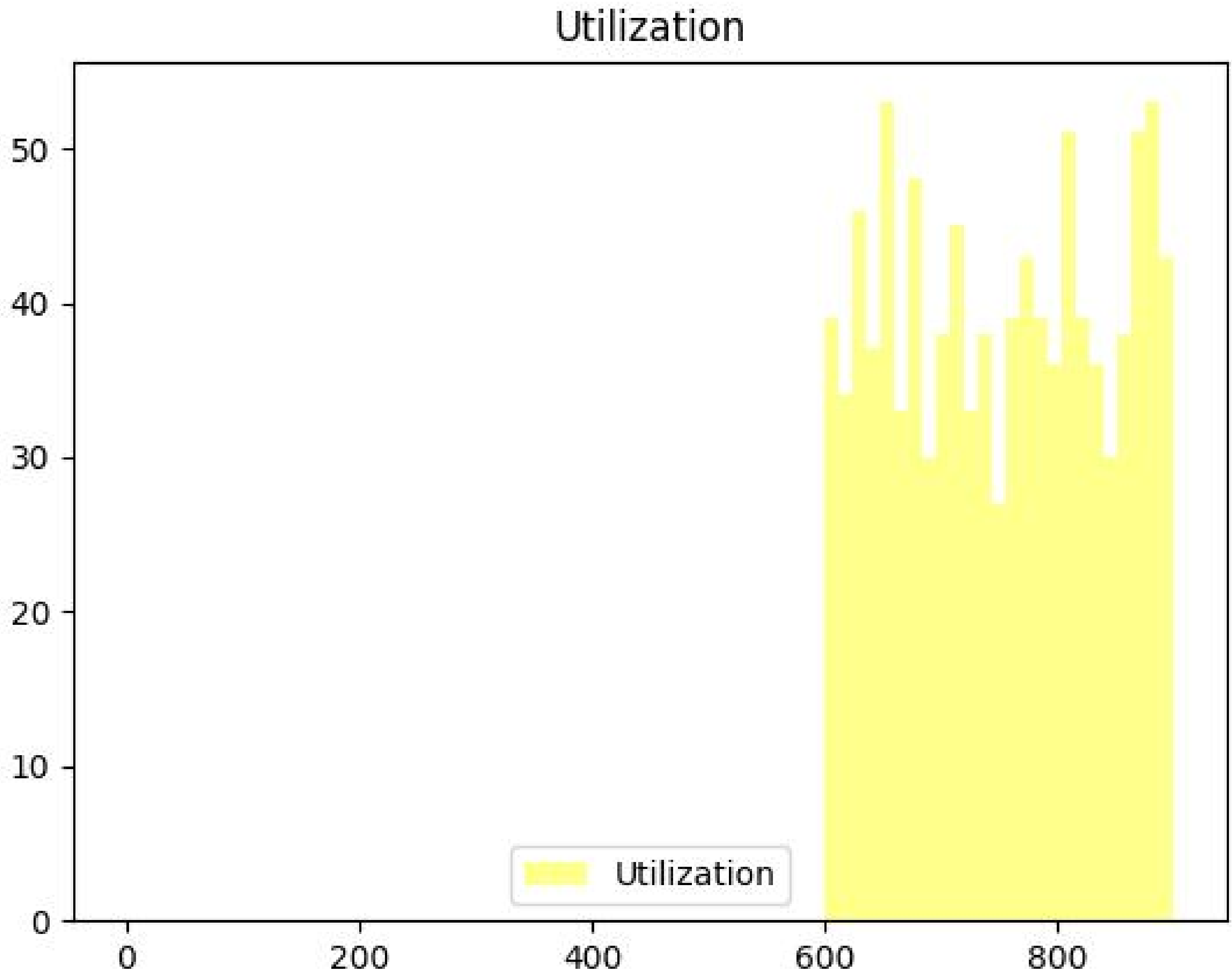
monthly_salary	fico	utilization	card_limit	card_interest_rate	model_output	model_target
999.0	999.0	999.0	999.0	999.0	999.0	999.0
105375.97497497498	751.91991991992	0.08462124902570549	27660.368368368367	0.07176645897347463	0.5063571896126577	0.25225225225225223
55439.852672523186	88.45413422399332	0.046560447191947775	12794.068691930182	0.029301571692248295	0.2889412385488691	0.4345227871326284
10006.0	600.0	9.326195062288785e-05	5033.0	0.02010396802218434	0.0007042672467625355	0.0
58333.5	673.5	0.0460609759511848	17269.5	0.04640339898806236	0.2580230214839624	0.0
106353.0	755.0	0.08527530840981518	27829.0	0.072112804761293	0.5113109220415197	0.0
155904.5	828.0	0.12321036798455527	38435.0	0.09762728049641894	0.7615762953311813	1.0
199659.0	900.0	0.16655351678655295	49867.0	0.11999063141494118	0.9999091598340564	1.0



True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0


Utilization
50
40
30
20
10
0
0
200
400
600
800
Utilization


True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0


Utilization
50
40
30
20
10
0
0
200
400
600
800
Utilization