

Navigation

Go to

Home

Upload model and data

Add public and model

info

Visualise data and model

output

Export pdf

Complete Generated Report

Download PDF

2023-03-29 22:16:21.017279

Overview

The Credit Line Decrease (CLD) model is a tool used to identify accounts with high risk of potential losses. This model will be applied to all RRB credit card customers in order to mitigate potential losses by decreasing their available credit limit.

Data Preparation

1. **Data Collection:** First, the data was collected from the Card Master File, as well as from other sources such as customer surveys, market research, and financial statements.
2. **Data Cleaning:** The data was then cleaned to remove any irrelevant or incomplete information. This included removing duplicate records, correcting errors, and formatting data.
3. **Data Transformation:** The data was then transformed into a format that could be used in the model. This included normalizing or standardizing numerical variables, creating dummy variables for categorical variables, and performing feature engineering.
4. **Data Splitting:** The data was then split into a training and test set to evaluate the model's performance.

5. Model Building: The model was then built using the training data. This involved selecting the right algorithms, tuning parameters, and evaluating the model’s performance.
6. Model Evaluation: The model was then evaluated on the test data. This involved assessing the model’s accuracy, precision, recall, and other metrics.
7. Model Deployment: Finally, the model was deployed in production and monitored for performance.

Validation data results

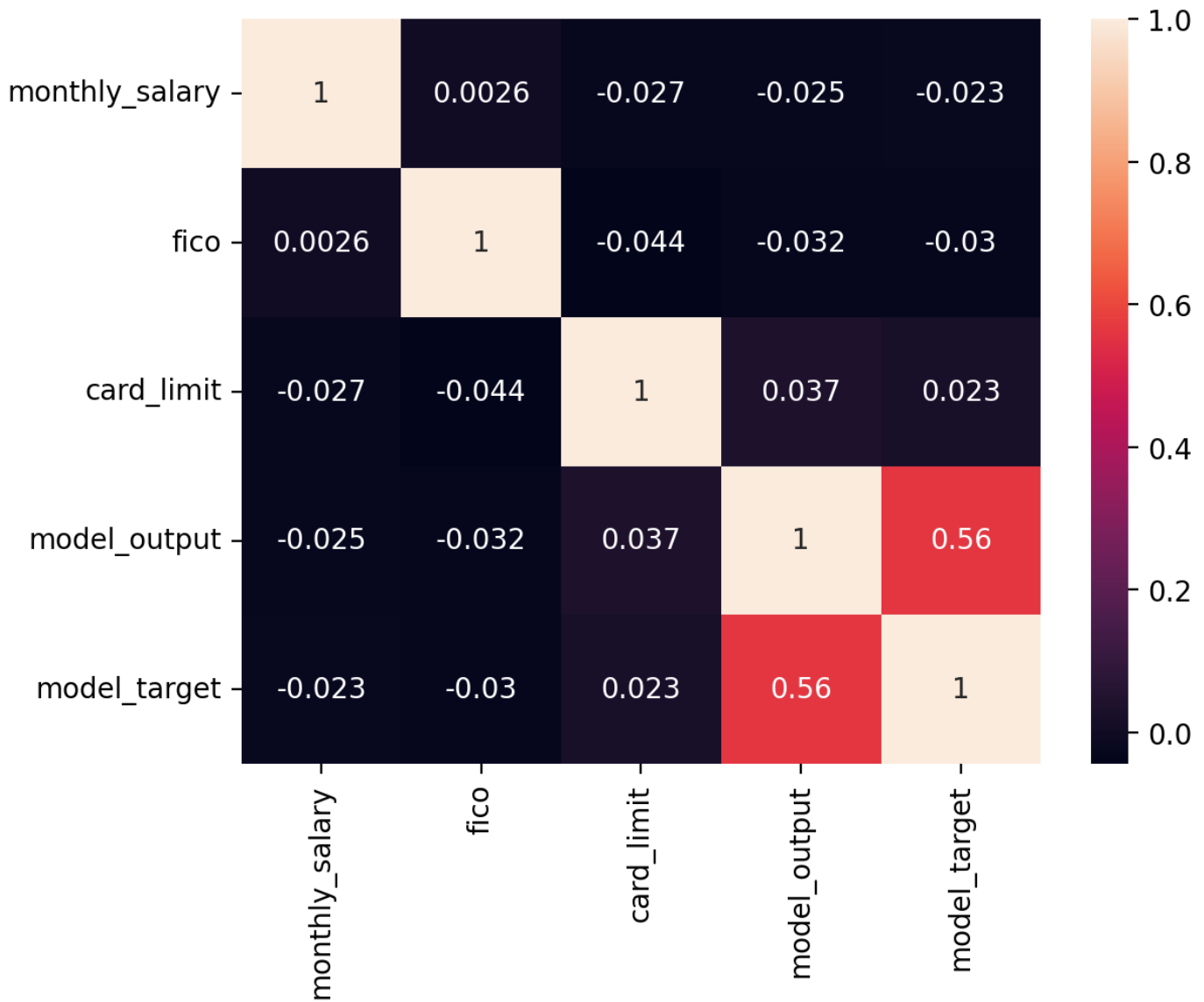
Below is our dataframe that contains the model output along with its confidence score

	monthly_salary	fico	utilization	card_limit	card_interest_rate	model_output	model_target
0	78,780	897	13.17%	31,032	3.75%	0	0
1	10,065	794	6.22%	31,697	3.41%	1	1
2	97,434	880	8.44%	48,998	9.68%	1	0
3	126,755	729	0.73%	38,827	4.14%	0	0
4	70,696	624	15.47%	9,082	5.51%	0	0
5	173,006	889	14.57%	18,039	2.86%	0	0
6	88,528	814	14.53%	9,946	2.40%	0	0
7	40,985	743	11.69%	25,788	9.50%	0	0
8	154,213	864	4.80%	20,240	9.06%	0	0
9	48,492	612	0.15%	45,523	10.22%	0	0

Data stats

	monthly_salary	fico	card_limit	model_output	model_target
count	999	999	999	999	999
mean	105,375.975	751.9199	27,660.3684	0.5155	0.2523
std	55,439.8527	88.4541	12,794.0687	0.5	0.4345
min	10,006	600	5,033	0	0
25%	58,333.5	673.5	17,269.5	0	0
50%	106,353	755	27,829	1	0
75%	155,904.5	828	38,435	1	1
max	199,659	900	49,867	1	1

Correlation Matrix

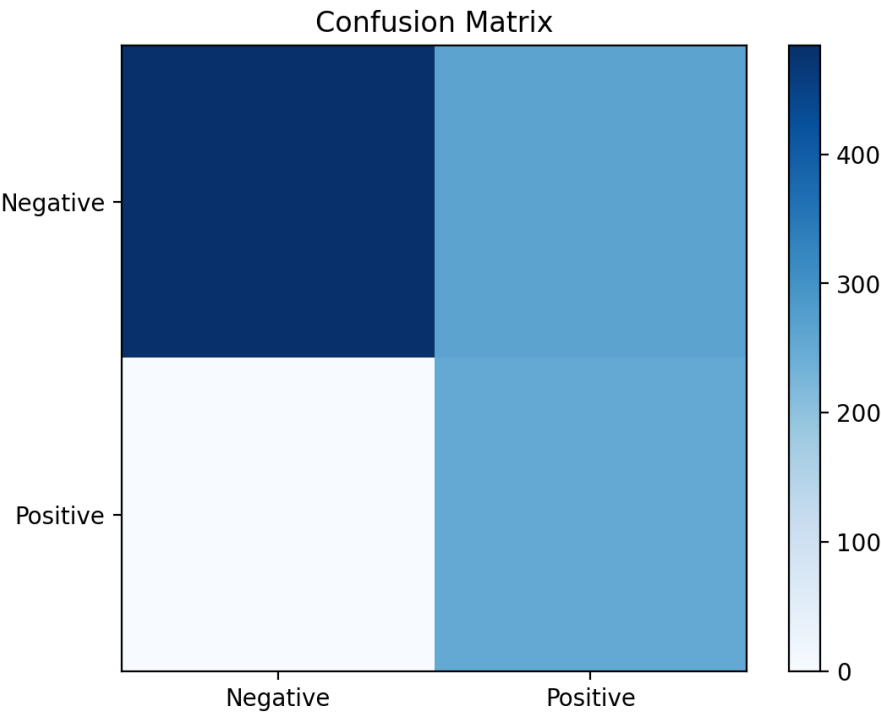


Modelling and results

The model that has been selected is: **XGBoost**

and Logistic Regression are two techniques used for creating a predictive model. GBM trees are also commonly used for this purpose. In the given example, XGBoost is used to build the final model, as it has been observed that it provides the best performance benefits.

Confusion Matrix



accuracy: 0.7367367367367368

precision: 0.4893203883495146

recall: 1.0

f1: 0.6571056062581486

Data descriptions

	accuracy	precision	recall	f1
0	0.7367	0.4893	1	0.6571