



German Credit Risk Factsheet Report

Table of contents

Business

● Overview	3
------------	---

Data

● Dataset	4
● Data Quality	5

Model Lifecycle

● Summary	6
● Metrics	7
● Metrics Configuration	8
● Develop	9
● Validate	10
● Monitor Configuration	11
● Monitor	12

Appendix

● Documentation	13
● Additional Information	14

Business

Overview

Purpose	Predict mortgage approval
Risk Level	High
Model Policy	<ul style="list-style-type: none">• Datasets must be approved and in data catalog.• Race, ethnicity, and gender of applicant cannot be used in models used to make mortgage related decisions.• Model predictive performance metrics must minimally include accuracy, balanced_accuracy and AUC score.• Models must be checked for bias using Disparate Impact.• Models must be checked for faithfulness of explanations.• Models must be checked for robustness to dataset shift.
Licensing note	MIT

Data

Dataset

Source

Purpose

File

train

german_credit_risk_train_24_months.csv

test

german_credit_risk_test_6_months.csv

Selection conditions

- **loan_purpose** is Home purchase
- **derived_loan_product_type** is Conventional:First Loan
derived_dwelling_category is Single Family (1-4 Units):Site-Built
- Covered loan or application is not for an open-end line
- Covered loan or application is not primarily for a business or commercial purpose
- Covered loan or application is not reverse mortgage
- The **occupancy_type** is a primary residence

Data split

Purpose

Ratio

train

65%

validate

15%

test

20%

Problem type

Binary classification

Label column

Risk

Data Quality

Column	Data Type	Count	Unique	Null
checking_status	Categorical	3,000	8	0
loan_duration	Numeric	3,000	156	0
credit_history	Categorical	3,000	5	0
loan_purpose	Categorical	3,000	5	0
loan_amount	Categorical	3,000	2,451	0
current_savings	Categorical	3,000	8	0
employment_duration	Categorical	3,000	5	0
other_loans	Categorical	3,000	4	0
gender	Categorical	3,000	2	0
age	Numeric	3,000	24	0
housing	Categorical	3,000	24	0
job	Categorical	3,000	7	0
dependents	Numeric	3,000	5	0
telephone	Categorical	3,000	3	0

Model Lifecycle

Summary

Develop

Validate

Monitor

German Credit Risk-SGD

Randy P

2022-02-10 10:00

German Credit Risk-SGD

Nijesh K

2022-02-15 13:00

Rejected

German Credit Risk-RF

Randy P

2022-02-10 10:30

German Credit Risk-RF

Nijesh K

2022-02-15 15:00

Approved

German Credit Risk-RF

Fujioka-san

2022-02-21 11:00

Deployed

Metrics

Standard Metrics	Description
Accuracy	Proportion of correct predictions.
F1	Harmonic mean of precision and recall.
ROC	Area under receiver operating characteristic curve.
False Positive Rate	Proportion of false positives against all positive predictions.
True Positive Rate	Proportion of actual positives was identified correctly.
Average precision	Average precision computes the average precision value for recall value over 0 to 1.
OpenScale Metrics	Description
Quality	Quality evaluation reveals how well your model predicts outcomes.
Drift	Drift is the degradation of predictive performance over time.
Fairness	Fairness evaluates your model for biases to ensure fair outcomes among different groups.
Statistical Parity Difference	The difference of the rate of favorable outcomes received by the unprivileged group to the privileged group.

Metrics Configuration

Fairness	Feature	Majority	Minority	Favourable	Unfavourable
	Age	26 - 75	18 - 25	No Risk	Risk
	Gender	Male	Female	No Risk	Risk
Statistical Parity Difference	Feature	Majority	Minority	Favourable	Unfavourable
	Age	26 - 75	18 - 25	No Risk	Risk
	Gender	Male	Female	No Risk	Risk

Develop

Model #1

German Credit Risk-SGD

Created by Randy P

Created on 2022-02-10 10:00

Metrics	Value
Accuracy	83.8%
F1	80.9%
ROC	81.2%
False Positive Rate	0.5%
True Positive Rate	90.9%
Average Precision	83.9%

Training Parameters	Value
alpha	0.0001
l1_ratio	0.15
max_iter	1,000
penalty	l2
power_t	0.5
sparse_threshold	0.3

Model #2

German Credit Risk-RF

Created by Randy P

Created on 2022-02-10 10:30

Metrics	Value
Accuracy	89.1%
F1	92.8%%
ROC	92.1%
False Positive Rate	0.3%
True Positive Rate	94.5%
Average Precision	90.1%

Training Parameters	Value
criterion	gini
min_samples_leaf	1
min_samples_split	2
n_estimators	100

Validate

Model #1

German Credit Risk-SGD

Validated by	Nijesh K
Validated on	2022-02-15 13:00
Outcome	Rejected

Standard Metrics	Train Value	Validate Value
Accuracy	81.9%	80.1%
F1	80.7%	81.2%
ROC	83.5%	82.9%
False Positive Rate	0.6%	0.6%
True Positive Rate	82.8%	83.1%
Average Precision	83.9%	84.6%

OpenScale Metrics	Train Value	Validate Value
Quality	81.2%	82.9%
Drift	5%	5%
Fairness	57.8%	59.1%
Statistical Parity Difference	0.1	0.1

Model #2

German Credit Risk-RF

Validated by	Nijesh K
Validated on	2022-02-15 15:00
Outcome	Approved

Standard Metrics	Train Value	Validate Value
Accuracy	90.1%	92.7%
F1	93.4%	94.9%
ROC	92.6%	93.8%
False Positive Rate	0.5%	0.5%
True Positive Rate	94.9%	95.1%
Average Precision	85.1%	86.4%

OpenScale Metrics	Train Value	Validate Value
Quality	92.1%	93.8%
Drift	5%	5%
Fairness	60.1%	61.5%
Statistical Parity Difference	0.08	0.08

Monitor Configuration

Drift

Parameter	Value
min_samples	100
enable_model_drift	True
enable_data_drift	True
drift_threshold	0.1

Fairness

Parameter	Value
metric	fairness_value
Age - lower_limit	80
Gender - lower_limit	80

Quality

Parameter	Value
min_feedback_data_size	100
metric	area_under_roc
lower_limit	0.9

Statistical Parity Difference

Parameter	Value
upper_limit	0.01

Monitor

Model #2

German Credit Risk-RF

Deployed (production) by Fujioka-san
 Deployed (production) on 2022-02-21 11:00
 Metrics as of 2022-02-21 16:00

Standard Metrics	Train Value	Validate Value	Current Value
Accuracy	90.1%	92.7%	91.3%
F1	93.4%	94.9%	93.5%
ROC	92.6%	93.8%	93.1%
False Positive Rate	0.5%	0.5%	0.6%
True Positive Rate	94.9%	95.1%	94.8%
Average Precision	85.1%	86.4%	86.2%
OpenScale Metrics	Train Value	Validate Value	Current Value
Quality	92.1%	93.8%	93.1%
Drift	5%	5%	6%
Fairness	60.1%	61.5%	60.9%
Statistical Parity Difference	0.1	0.1	0.1

Appendix

Documentation

Additional Information