Model Risk Management

credit_adjudication

Executive summary

This section reports the description of the model as specified by the model owner and reported on mlflow registry. The narrative should clearly articulate the business motivations behind this initiative, the desired business outcomes as well as its associated risk for the bank and / or its customers. This section will be used to assess the materiality of this model and may trigger different review processes and compliance requirements accordingly.

Model name	credit_adjudication
Model catalog	fsgtm
Model schema	mrm
Model creation date	2024-03-18
Model owner	antoine.amend@databricks.com
Model materiality	HIGH
Model review	REQUESTED

description from mlflow registry

Co-developped with EY, This model is a simple example of how organisations could standardize their approach to AI by defining a series of steps that any data science team ought to address prior to a model validation. Although not exhaustive, this shows that most of the questions required by IVU process for a given use case (Credit adjudication) could be addressed upfront to reduce the friction between regulatory imposed silos, increase model validation success rate and drammatically reduce time from exploration to productionization of AI use cases.

Model history

This section reports all previous model versions, the name of the submitter and date of registration.

Version	Timestamp	Owner	Alias
1	2024-03-18	antoine.amend@databricks.com	@production

Model submission request

This section reports the description of the version of model submitted by the model owner and reported on mlflow registry. The narrative should clearly articulate the business motivations behind this new submission, and the desired benefits relative to previous model versions. Please ensure markdown is attached to your model submission on mlflow registry.

Model submission date	2024-03-18T10:38:03.438000
Model version owner	antoine.amend@databricks.com
Model version	1
Model run Id	824164f19fc0419c9edd129011fa8d9b
Model complexity	MEDIUM
Model explainability	MEDIUM
Model selection	HYPEROPT
Model type	XGBClassifier

description from mlflow registry

This version of credit adjudication model was built for the purpose of unity catalog demo. Model was co-developped between EY and Databricks, finding XGBClassifier as best fit model trained against 50 different experiments. All experiments are tracked and available on MLFlow experiment tracker.

Developmental history and conceptual soundness

This section reports the technical approach taken during the implementation of this particular model version. The narrative should clearly articulate the consideration behind the use of specific libraries and frameworks, the relevance of the data used throughout this exercise as well as the assumptions, constraints, dependencies, limitations and risks (ACDLR) as identified at the start of this project. Using empirical evidence, this section should clearly indicate why this particular experiment was proved to be the best model candidate and why other experiments or approaches were discarded. Finally, when applicable, the practitioner should be able to explain their strategies to ensure an explainable, fair and ethical use of data / AI. Please ensure markdown is attached to your model experiment on mlflow.

Execution time	2024-03-18T10:37:33.753000
Execution user	antoine.amend@databricks.com
Execution workspace	e2-demo-west.cloud.databricks.com
Execution type	NOTEBOOK
Execution code	/Repos/antoine.amend@databricks.com/mrm-generation/templates/Credit Adjudication - Example
Execution code url	https://github.com/databricks-industry-solutions/fsi-mrm-generation.git
Execution code revision	406bbd7c659b55fed347e25c0bf75a6509dca1d9

description from mlflow experiment

A 10 fold cross-validation procedure was used to select the best model and hyperparameters across multiple techniques. Our model selection included XGBoost and K nearest neighbors and selected XGBClassifier as best fit. This run was evaluated as our best run that maximizes `cross_val_score`.

Model development, implementation and testing

This section dynamically pulls all the technical context around the implementation of the model itself. A given model registered on mlflow should have an associated experiment that can be linked to actual code at a given version. The goal is to document the approach taken by the model developer in the implementation of the model. We report all the technical metadata and specification of the artifact(s) logged on mlflow, the parameters used and output metrics.

Submitted artefacts

In this section, we report all binary artefacts that were stored alongside this model. Since a model may have multiple 'flavors' (or interpreter), we report each binary and their respective version.

Logged time	Artifact	Interpreter version
2024-03-18 17:37:35.377333	python_function	3.10.12
2024-03-18 17:37:35.377333	sklearn	1.1.1

Developmental overview

This section will automatically retrieve the code associated with the model experiment. We report a databricks JOB output or a databricks NOTEBOOK markdown and their respective output cells. This becomes the responsibility of the model developer to document their approach with distinct sections and headers, from data sourcing and transformation, exploratory data analysis, feature selection, model selection and validation as well as model explainability when applicable. We recommend organizations to create template notebooks covering internal policies and external compliance requirements to ensure consistency and relevance of this documentation. Such policies will be seamlessly reported here.

markdown cell #0

Credit Risk Adjudication Model

This notebook is a simple example of how organisations could standardize their approach to AI by defining a series of steps that any data science team ought to address prior to a model validation. Although not exhaustive, this shows that most of the questions required by IVU process could be addressed upfront (at model development phase) to reduce the friction between regulatory imposed silos, increase model validation success rate and drammatically reduce time from exploration to productionization of AI use cases.

EY
Building a better working world

markdown cell #1

Executive Summary

This notebook demonstrates the use of Machine Learning for credit risk adjudication model. We will be loading a publicly available dataset and evaluate multiple modelling techniques and parameter tuning using hyperopts in order to select the best approach balancing between model explainability and model accuracy.

markdown cell #2

1 Introduction

markdown cell #3

1.1 Model Background and Initiation

The motivations behind this modeling effort is to showcase Lakehouse capabilities combined with EY expertise as it relates to model risk management. The goal is not to build the best model nor to showcase latest state of the art AI capabilities.

markdown cell #4

1.2 Model Purpose

The purpose of this document is to provide a detailed description of the new retail credit adjudication model.

markdown cell #5

1.3 Model Portfolio

The MLflow Model Registry component is a centralized model store, set of APIs, and UI, to collaboratively manage the full lifecycle of an MLflow Model. It provides model lineage (which MLflow experiment and run produced the model), model versioning, stage transitions (for example from staging to production), and annotations. Used as a backbone of our model risk management solution accelerator, this becomes the de facto place to register both machine learning and non machine learning models.

markdown cell #6

1.4 Model Risk Rating

Credit Risk Model would give creditors, analysts, and portfolio managers a way of ranking borrowers based on their creditworthiness and default risk. Any issue on the model output would have financial consequences, leading to a relative `HIGH` model materiality.

markdown cell #7

1.5 Model Log of Changes

We captured all different models and previous versions using MLFlow. Model development history is available through the MLFlow registry UI / API.

markdown cell #8

1.6 Business-Driven Risk Considerations

Note: Explain the business risks that are explored and assessed during the model development process, and how they are accounted for in the final model (outputs). Describe and justify any mitigation action (plan) that helps reduce the business-driven risk.

markdown cell #9

1.7 Economic and Market Outlook

Note: Explain how the current and forward-looking overall economic conditions may impact the business line and subsequently the model outcome.

markdown cell #10

1.8 Model Development Process

Note: Describe the overall model development process, the different milestones of the process, along with the roles and responsibilities of the stakeholders involved at each of these key steps.

markdown cell #11

1.9 Economic and Market Outlook

Note: Explain how the current and forward-looking overall economic conditions may impact the business line and subsequently the model outcome.

markdown cell #12

2 Data

markdown cell #13

2.1 Borrower Definition

Note: Describe the borrowers' categories of the model portfolio/population. For instance, whether the model applies to borrowers with a certain range of exposure, within a geographical area, or with a minimum/maximum of total asset (e.g., when the model also applies to SMEs). It outlines the borrower identification process in the data bases as well.

markdown cell #14

2.2 Data Sources

markdown cell #15

2.2.1 Internal Data Sources

Note: Describe the internal data sources, as well as their appropriateness with the model purpose and model population.

markdown cell #21

2.2.2 External Data Sources

The external data contains personal information on clients with saving and/or checking accounts. Overall, 1,000 observations are included in the dataset. The following describes the different variables along with the features of the dataset. We display a few records below as well as table statistics.

- `AGE` (numeric)
- `SEX` (text: male, female)
- `JOB` (numeric: 0 - unskilled and non-resident, 1 - unskilled and resident, 2 - skilled, 3 - highly skilled)
- `HOUSING` (text: own, rent, or free)
- `SAVING_ACCOUNT` (text - little, moderate, quite rich, rich)
- `CHECKING_ACCOUNT` (numeric, in DM - Deutsch Mark)
- `CREDIT_AMOUNT` (numeric, in DM)
- `DURATION` (numeric, in month)
- `PURPOSE` (text: car, furniture/equipment, radio/TV, domestic appliances, repairs, education, business, vacation/others)

markdown cell #29

2.3 Data Historical Coverage and Suitability

Note: Describe the data extraction process, along with the period spanned by the data and the statistics on the extracted observations. The section should not only evidence that the extracted data reflects the business practices and experiences, but is also suitable for the model purpose, modeling methodology and modeling assumptions.

markdown cell #30

2.4 Modeling Timeframes

markdown cell #31

2.4.1 Timeframe Concepts

Note: Explain the different concepts of the modeling timeframes used for the model development, specifically the observation period, the lag period, along with the performance period.

markdown cell #32

2.4.2 Determination of the Performance and Lag Periods

Note: Describe the determination process of the lag and performance periods, including the judgemental considerations that were used. Provide a justification of the selections and their consistency with the model product and the observed borrowers' experience. Explain the different concepts of the modeling timeframes used for the model development, specifically the observation period, the lag period, along with the performance period.

markdown cell #33

2.4.3 Modeling Timeframes

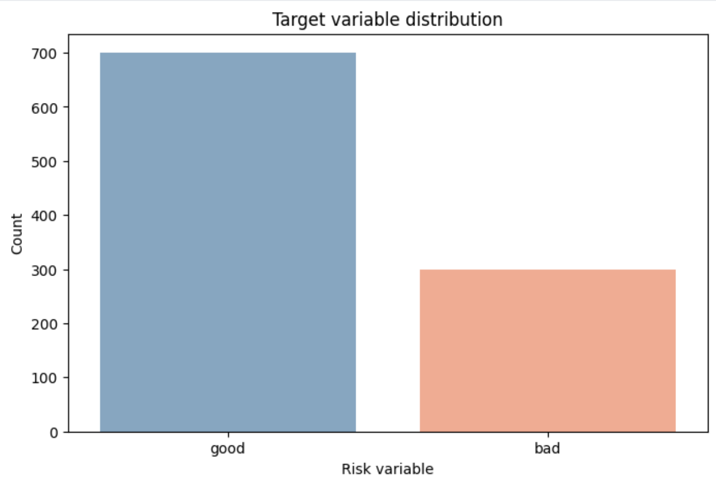
Note: Describe the different modeling timeframes that were finally selected (i.e., the corresponding periods to the concepts explained in Section 2.4.1) for the model development and validation.

markdown cell #34

2.5 Target Variable Definition

The model is designed to predict the likelihood of a loan defaulting. The target variable `RISK` (good/bad) is defined using the information in the extracted dataset. The target variable defines the loans status as 'good' or 'bad'. A 'good' status means a good credit performance, i.e., the client did not default during the observation period, whereas a 'bad' status means a default occurred during the observation period. In the modeling code, 'good' is identified as '0', and 'bad' is identified as '1' and encoded as our `RISK_EN` column. The following figures depict the target variable distribution (percentage of good/bad), according to the different variables.

output cell #35



markdown cell #36

2.6 Modeling Populations

markdown cell #37

2.6.1 Eligible Population

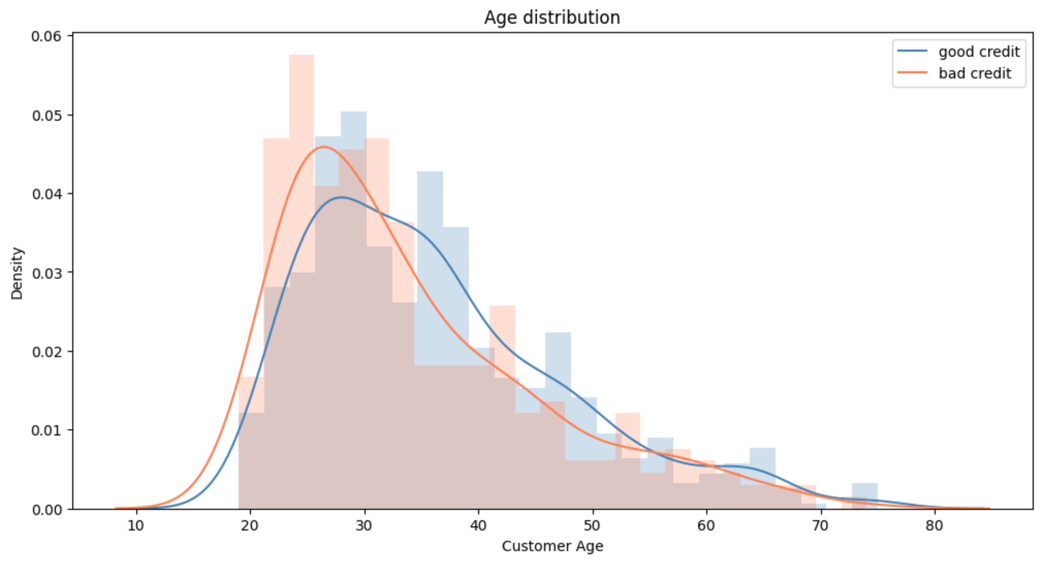
The following table provides descriptive statistics on the eligible population for the model development, which includes 1,000 observations, in total. Descriptive statistics apply to the overall population, without any data treatment such as exclusion or sampling. 'NaN' mostly appears when trying to compute statistics on categorical variables; hence, they may be ignored.

markdown cell #39

2.6.2 Good-Bad Observations

The following provides statistics on the 'good' and 'bad' observations. Overall, 700 'good' and 300 'bad' observations are found in the dataset. Histograms of 'good' and 'bad' observations are plotted below.

output cell #40



markdown cell #41

2.6.3 Indeterminate Observations

Note: Describe and provide statistics on observations that cannot be classified as good or bad observations.

markdown cell #42

2.6.4 Statistically Inferred Performance Data

Note: Describe the observations whose performance could not be observed (e.g.,indeterminate observations), the reject inference technique used to infer the performance. The reason supporting the selected technique, along with the considered population should be described as well.

markdown cell #43

2.7 Data Exclusions and Treatment

Note: Describe exclusions and any treatments (e.g., outlier and missing value treatment, and application of floors and caps) applied to the data, along with the supporting justification.

markdown cell #44

2.8 Sampling Methodology

Two different datasets, training and validation, were created for the modeling purpose. More specifically, a stratified random sampling methodology was used to sample the original dataset: About 80% was used to train the model, and the remaining 20% was considered for the model performance assessment. The tables below present descriptive statistics on the datasets.

markdown cell #47

2.9 Modeling Data Assessment

Note: Describe the final dataset that will be used for the model development. Describe the data quality, using statistics and graphs, describe any data limitations and their potential impact on the model output.

markdown cell #48

databricks

3 Model Development

markdown cell #49

3.1 Methodology Selection

Note: Describe the modeling methodology selection process. More specifically, first present and compare the different alternatives through the literature and industry practice review, and then explain the rationale behind the selected approach. In addition, outline the mathematical definitions and equations, along with the assumptions and limitations of the selected modeling methodology.

markdown cell #50

3.2 Model Segmentation

Note: Describe the model segmentation process, including the judgemental considerations, the statistical analyses, and the supporting rationale for the selected segments.

markdown cell #51

3.3 Model Variable Selection

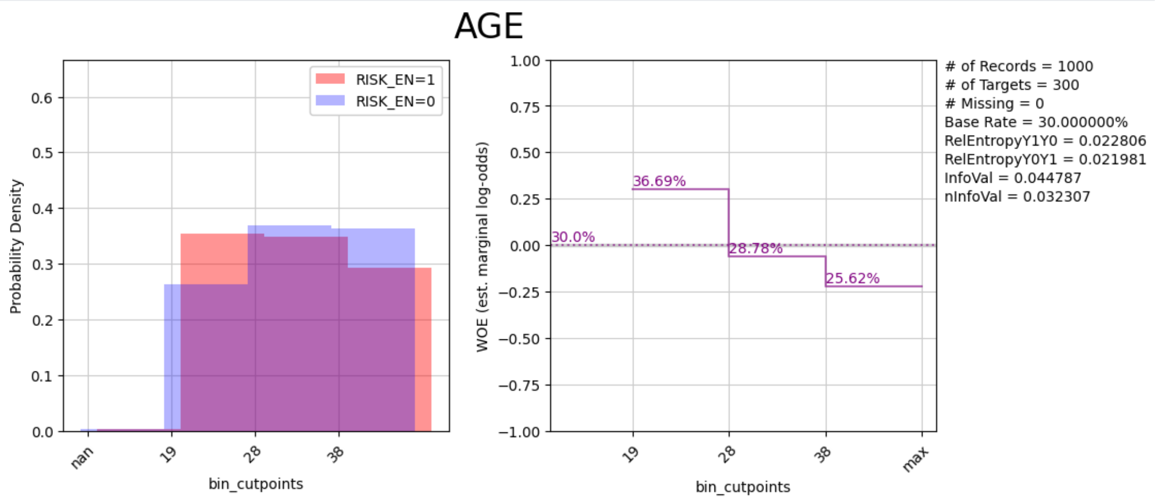
Note: Describe the variable selection process from the initial list until the selected variables. The statistical analyses with their results and the business considerations should be described in the corresponding sub-sections below. Only relevant and applicable sub-sections should documented. Additional analyses or tests may be added.

markdown cell #52

3.3.1 Variable Reduction

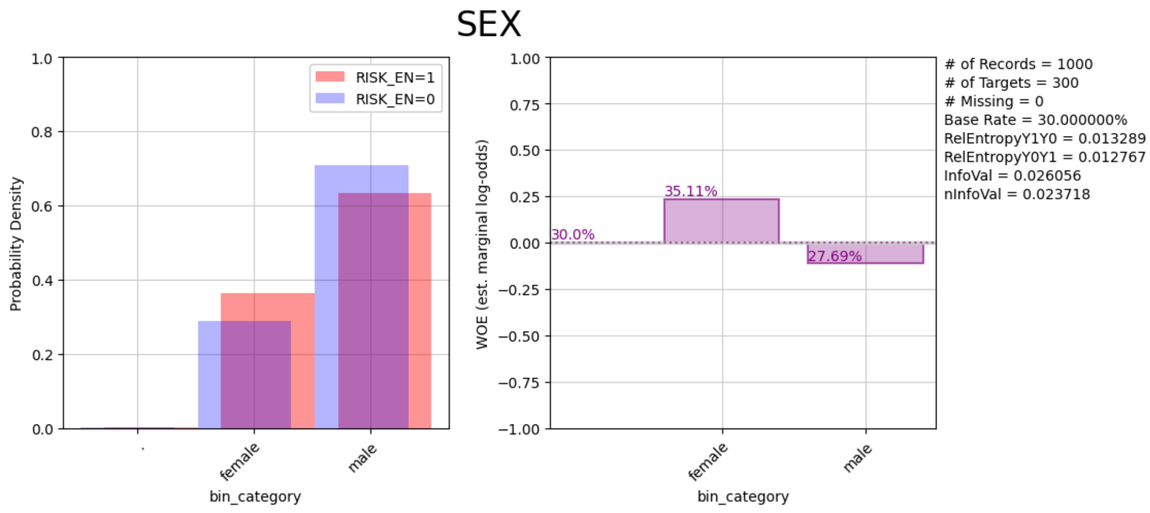
Preliminary analyses were conducted to support the variable selection process. More precisely, the variables were plotted according to different bin categories to assess their probability density. Moreover, weights of evidence (WOEs) which measure the relative risk of each bin within each variable were also calculated and evaluated as part of the variable selection process. The results of the probability density and WOEs for each variable are showed below.

output cell #65

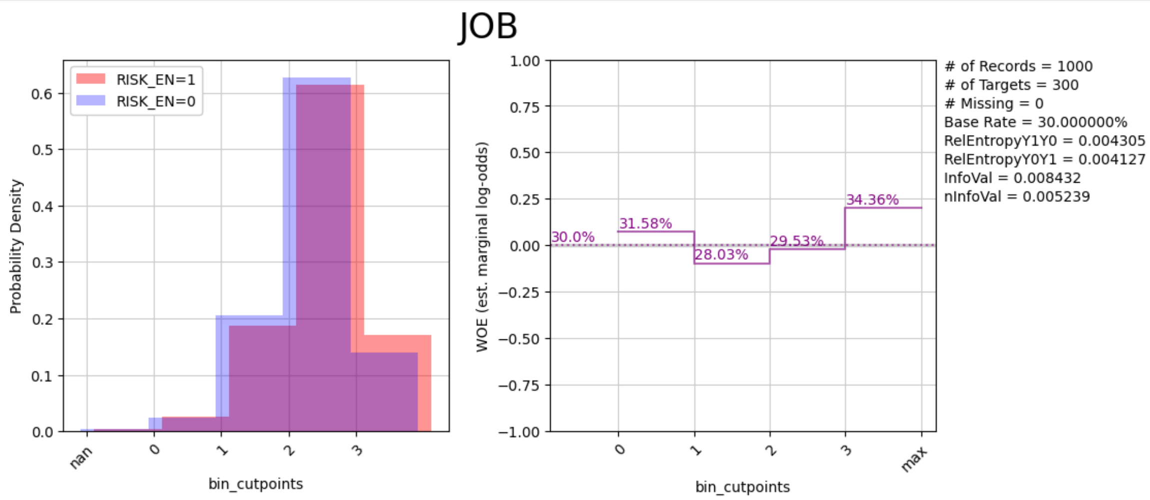


output cell #65

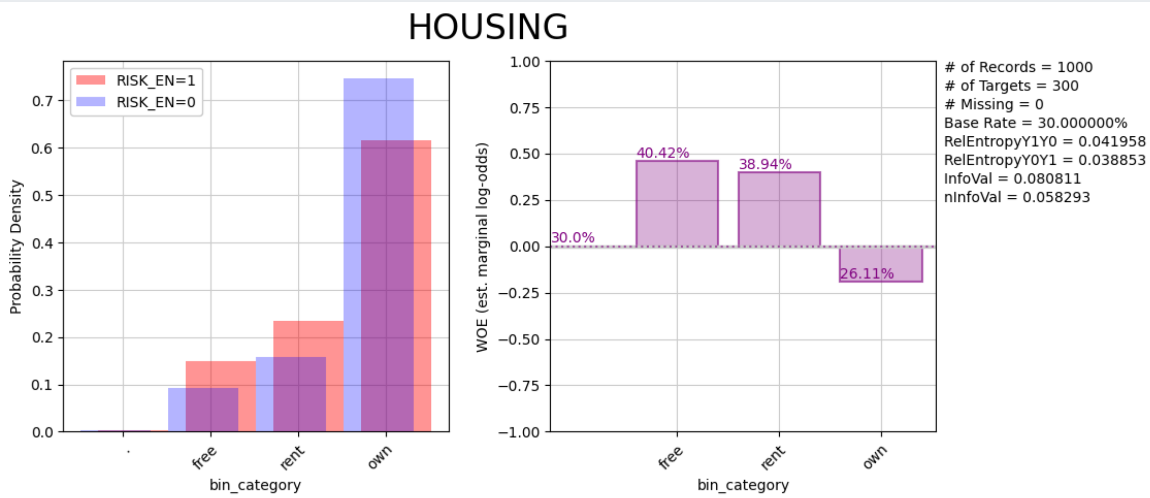
databricks



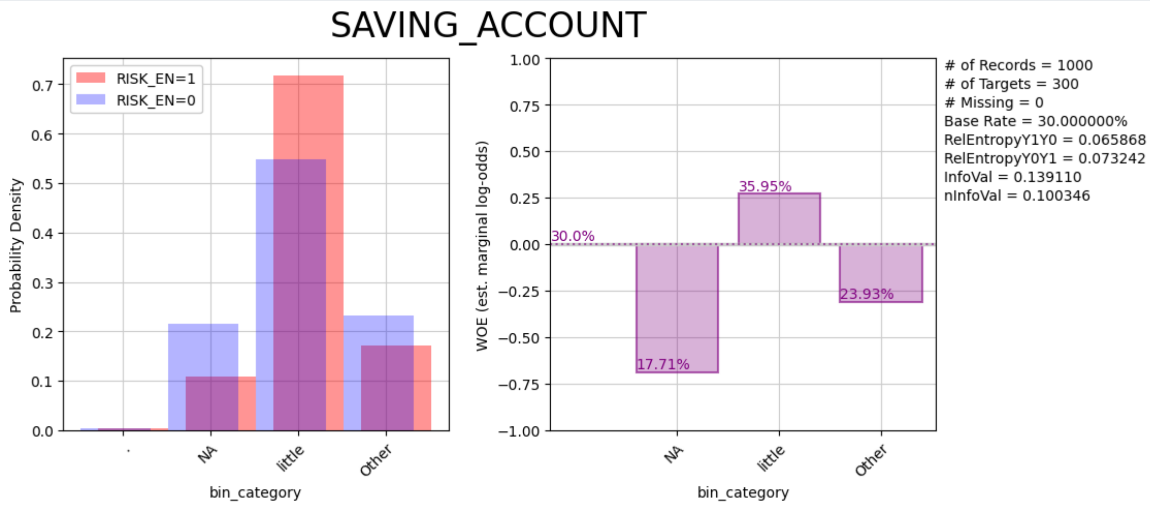
output cell #65



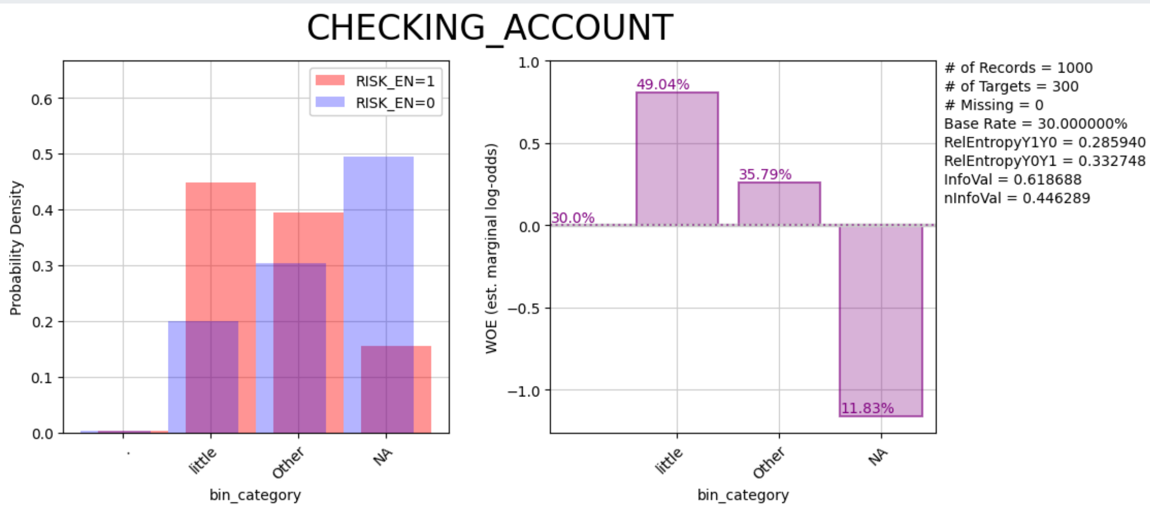
output cell #65



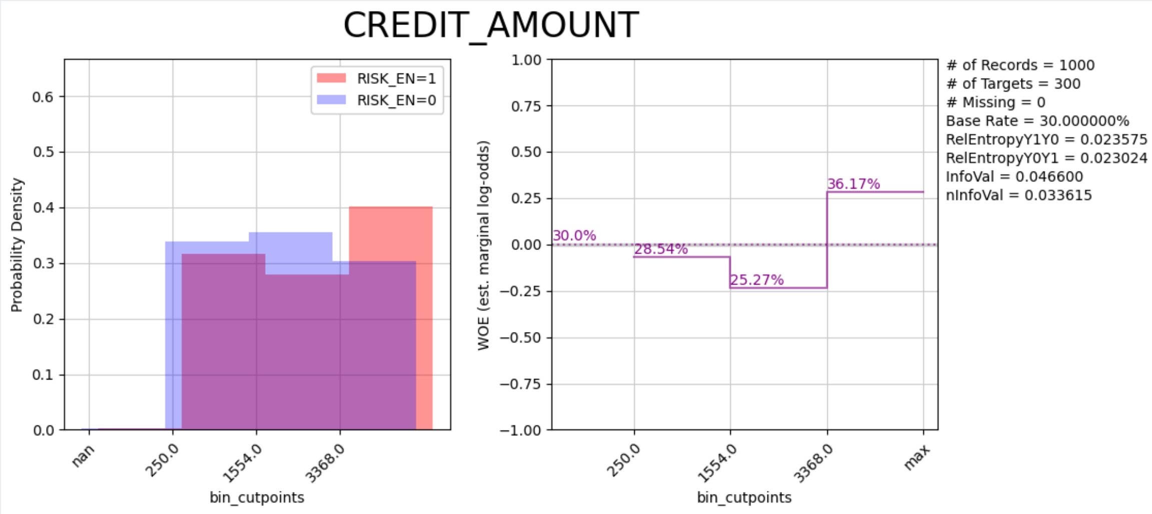
output cell #65



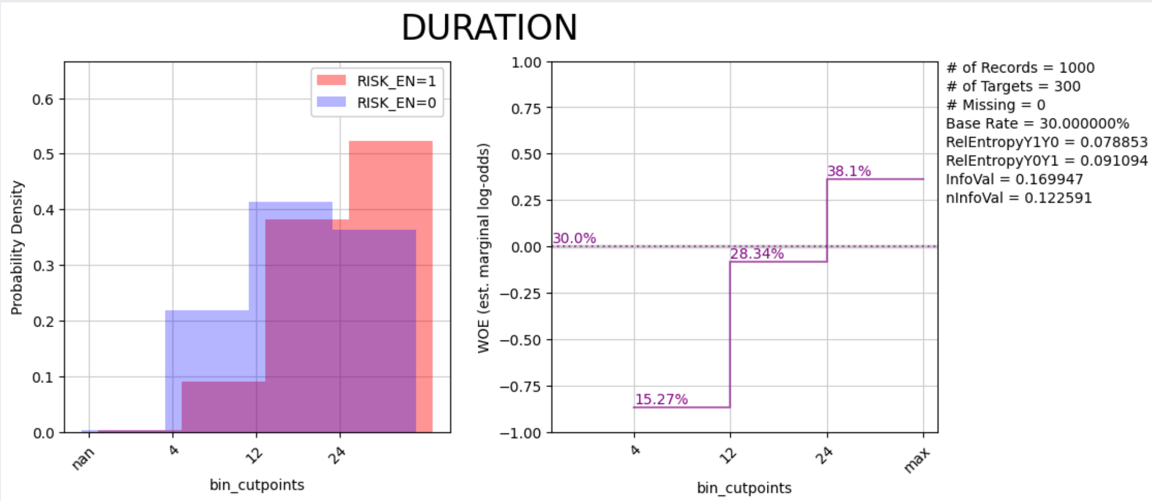
output cell #65



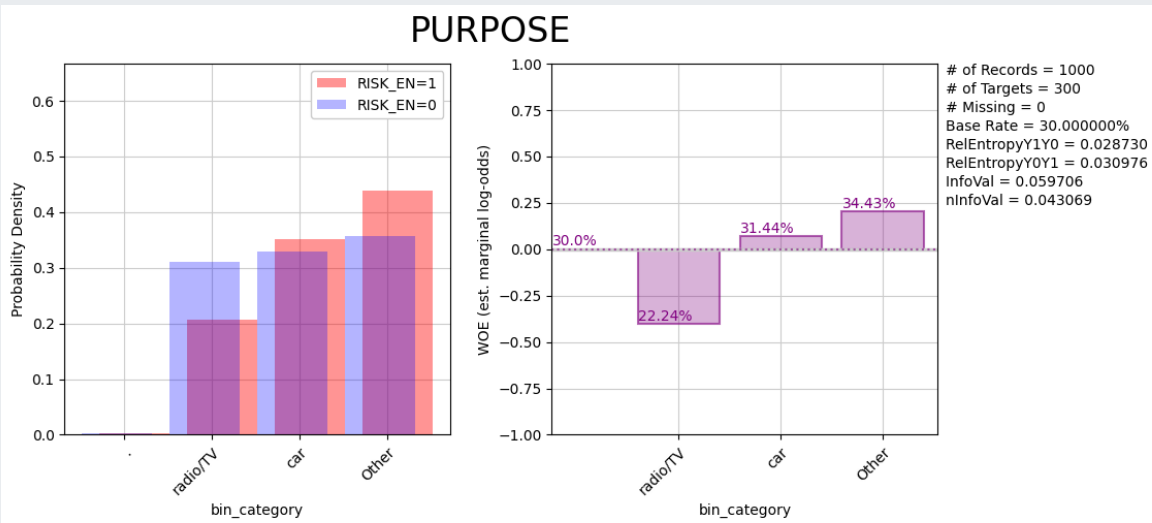
output cell #65



output cell #65



output cell #65



markdown cell #68

3.3.2 Final Variable Reduction

For the final variable reduction, intervals were created for some continuous variables such as the age, whereas dummies were created for categorical variables such the sex, housing, etc. Results of the analyses are presented below.

markdown cell #71

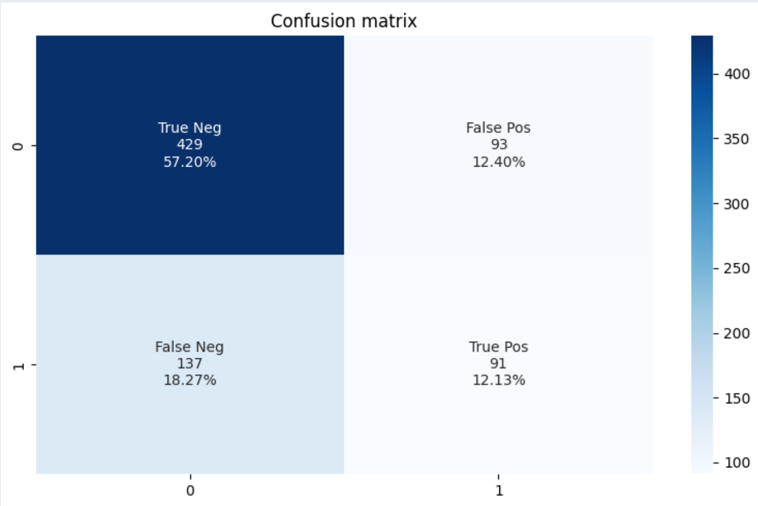
3.4 Model Estimation

For the model selection and estimation, a 10 fold cross-validation procedure is used to compare and select among different alternative models. The following models were trained using hyperopt for hyper parameter tuning.

- K Neighbors
- XGBoost

Confusion matrices (or error matrices) will be produced for model comparison purposes. Indeed, these matrices easily allow the visualization of the performance of the different models, in terms of actual vs. predicted classes.

output cell #81



markdown cell #82

3.5 Model Scaling

Note: Describe the model scaling process. More specifically, cover the selection of the scaling equations and parameters, as well as the expert judgements that were considered. Display and interpret the model final results.

markdown cell #83

4 Model Performance Assessment

Note: Thoroughly assess the model performance in this section. Each sub-section is designed to cover particular dimension that is assessed, outline the analysis or statistical test that is performed and provide the results interpretation. Keep only relevant and applicable sub-sections. Add additional analyses or tests.

markdown cell #84

4.1 Output Analysis

markdown cell #85

4.2 Discriminatory Power Testing

markdown cell #86

4.2.1 Accuracy Ratio Test

To better assess the models' performance, different accuracy tests including, the accuracy ratio, the precision test, the recall test and the F1 test were performed. Results of these tests are showed in the following tables.

markdown cell #89

4.2.2 Kolmogorov-Smirnov Test

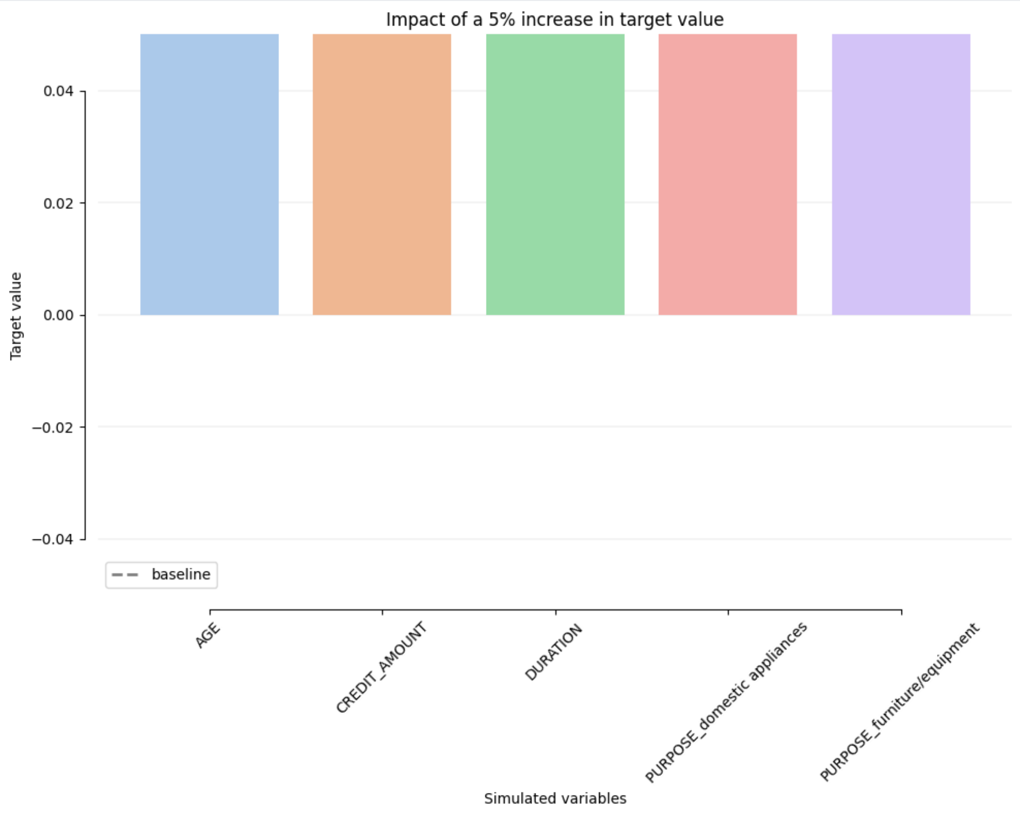
In addition to the aforementioned performance tests, the KS test was also performed, and results are the following.

markdown cell #92

4.3 Sensitivity Analysis

A sensitivity analysis was conducted to identify the key variables that mostly impact the model results. For instance, a 5% increase in the following variables, age, credit amount, duration, purpose (if domestic appliances and furniture/equipment) was performed, and the impact reasonableness was assessed. Sensitivity analyses results are showed below.

output cell #96



markdown cell #97

4.4 Population Stability Analysis

markdown cell #98

4.5 Benchmarking

For the benchmarking, please refer to the section of the model estimation results, where different models were trained, and the results were compared using confusion matrices.

markdown cell #99

5 Model Assumptions and Limitations

markdown cell #100

5.1 Model Assumptions

Note: Describe the key assumptions made throughout the model development process and provide evidence to support their reasonableness and soundness.

markdown cell #101

5.2 Model Limitations

Note: Describe the key model limitations, their potential impact on the model, as well as the corresponding mitigation action plan(s) to reduce the model risk.

markdown cell #102

6 Model Ongoing Monitoring

markdown cell #103

6.1 Ongoing Performance Assessment

Note: Describe the ongoing model performance monitoring plan. Cover the statistical tests (including e.g., the frequency and acceptance thresholds) that will be performed on an ongoing basis to ensure the model is still performing adequately.

markdown cell #104

6.2 Documentation Review

Note: Describe the conditions or types of model changes that trigger the model documentation review, as well as the key components that need to be reviewed.

markdown cell #105

7 References

markdown cell #106

8 Model registry

Finally, we will log all evidence required to trigger an independant review of our modeling approach. We show how to do so programmatically, though this process could be done manually from the MLFlow UI.

markdown cell #111

EY
Building a better working world

Disclaimer: The views and opinions expressed in this blog are those of the authors and do not necessarily reflect the policy or position of EY.

Model parameters

In this section, we report all the parameters used in the creation of this model. We highly recommend the use of mlflow auto-logging capability to ensure consistency of information reported from different teams and across different frameworks.

Parameter	Value
enable_categorical	False
eval_metric	logloss
missing	nan
n_estimators	100
objective	binary:logistic
verbosity	0

Model calibration

In this section, we report all the metrics logged in the creation of this model. We highly recommend the use of mlflow auto-logging capability to ensure consistency of information reported from different teams and across different frameworks.

Metric	Value
cv_accuracy	0.6933333333333332
cv_f1	0.44125741290000775
cv_precision	0.5016923703119691
cv_recall	0.4025829531092689
ks_pvalue	0.0
ks_statistic	1.0

Model dependencies

This section will retrieve all technical context surrounding the development of the model, the data set used, the input and output features, as well as infrastructure requirements and external libraries. This section will ensure model output can be reproduced under same conditions.

Infrastructure dependency

This section will programmatically retrieve the specification of the infrastructure used for the creation of the model. What environment was created, how many nodes were leveraged for distributed computing, what databricks runtime was used. We highly encourage users to leverage LTS versions of our runtimes.

Cluster name	antoine.amend@databricks.com's Personal Compute Cluster
Cluster runtime	14.3.x-cpu-ml-scala2.12
Cluster instance type	i3.xlarge
Cluster number workers	None

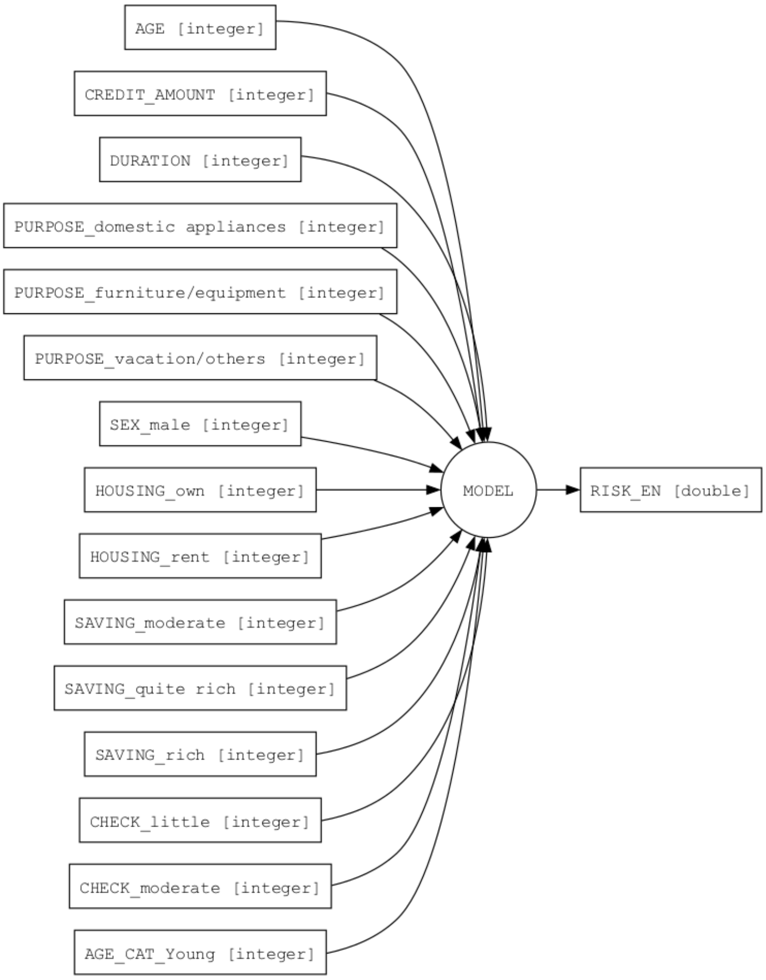
Libraries dependencies

Beside the cluster and infrastructure used for the creation of the model, specific libraries (open source or proprietary) may have been leveraged. This section will report all external dependencies (maven, pypi, custom packages) and their respective versions. We highly encourage users to install libraries at an infrastructure level rather than at a notebook level to ensure each library is properly tracked and reported here.

Could not find any associated libraries. Make sure dependencies are captured and installed as a cluster level (linked to an infrastructure rather than a notebook).

Input and output signatures

This section will programmatically represent the input features of the model and expected output signature. The transformations applied upfront should be documented as part of the developmental overview reported earlier.



Data dependencies

This section will report the different data sources used throughout this exercise. Using mlflow coupled with databricks notebooks, we should be able to track all data sources loaded through spark alongside their versions whenever possible. We highly encourage users to leverage delta format whenever possible to lock an experiment on a given data version we can easily time travel to.

Path	Format	Version
fsgtm.mrm.german_credit_data_risk	delta	v0

Model lineage

Whenever applicable, we will track the associated data lineage for every data dependency tracked in this experiments. Please refer to unity catalog to ensure lineage is captured end to end and reported here as a graphical representation.