# Driverless AI Experiment: Credit Card

Generated by: admin

Generated on: Fri Jan 22 05:08:20 2021

[Driverless AI Experiment: Credit Card 1](#_Toc24631991)

[Experiment Overview 3](#_Toc24631992)

[Data Overview 3](#_Toc24631994)

[Methodology 3](#_Toc24631996)

[Data Sampling 3](#_Toc24631998)

[Validation Strategy 3](#_Toc24632000)

[Model Tuning 3](#_Toc24632002)

[Feature Evolution 3](#_Toc24632004)

[Feature Transformations 3](#_Toc24632006)

[Final Model 3](#_Toc24632008)

[Alternative Models 3](#_Toc24632010)

[Deployment 3](#_Toc24632012)

[Partial Dependence Plots 3](#_Toc24632014)

[Appendix 4](#_Toc24632019)

## Experiment Overview

Driverless AI built 1 LightGBMModel to predict *default.payment.next.month* given 24 original features from the input dataset *UCI\_Credit\_Card.csv*. This classification experiment completed in 5 minutes and 56 seconds (0:05:56), using 22 of the 24 original features, and 0 of the 0 engineered features.

### Performance

|  |  |
| --- | --- |
| **Dataset** | **ACCURACY** |
| Internal Validation | 0.821 |
| Test Data | Test Data not Provided |

### Driverless Settings

|  |  |  |  |
| --- | --- | --- | --- |
| **Dial Settings** | **Description** | **Setting Value** | **Range of Possible Values** |
| **Accuracy** | Controls accuracy needs of the model | 3 | 1-10 |
| **Time** | Controls duration of the experiment | 1 | 1-10 |
| **Interpretability** | Controls complexity of the model | 5 | 1-10 |

### System Specifications

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Address** | **System** | **System Memory** | **CPUs** | **GPUs** |
| http://127.0.0.1:12345 | Docker/Linux | 60 GB | 4 | 1 |

### Versions

|  |  |
| --- | --- |
| **Driverless AI version** | 1.9.0 |
| **h2o4gpu version** | 0.4.1 |
| **h2o\_mli version** | 1.9.117 |
| **mojo2\_runtime version** | 2.4.8 |
| **procsy version** | 0.6.0 |
| **pydatatable version** | 0.11.0a1735 |
| **vis\_data\_server version** | 2.0.5 |

## Data Overview

This section provides information on the datasets used for the experiment.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **data** | **file path** | **file size** | **number of rows** | **number of columns** |
| training | ./tmp/admin/8a699c62-5c6e-11eb-86d5-0242ac110002/UCI\_Credit\_Card.csv.1611291541.0764227.bin | 4.3 MiB | 30,000 | 25 |
| validation | Not provided | None | None | None |
| testing | Not provided | None | None | None |

### Training Data

The training data consists of only numeric columns.

The summary of the columns is shown below :

#### Numeric Columns

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **name** | **data\_type** | **min** | **mean** | **max** | **std** | **unique** | **freq of mode** |
| ID | int | 1.000 | 15,000.500 | 30,000.000 | 8,660.398 | 30,000 | 1 |
| LIMIT\_BAL | real | 10,000.000 | 167,484.323 | 1,000,000.000 | 129,747.662 | 81 | 3,365 |
| SEX | int | 1.000 | 1.604 | 2.000 | 0.489 | 2 | 18,112 |
| EDUCATION | int | 0.000 | 1.853 | 6.000 | 0.790 | 7 | 14,030 |
| MARRIAGE | int | 0.000 | 1.552 | 3.000 | 0.522 | 4 | 15,964 |
| AGE | int | 21.000 | 35.485 | 79.000 | 9.218 | 56 | 1,605 |
| PAY\_0 | int | -2.000 | -0.017 | 8.000 | 1.124 | 11 | 14,737 |
| PAY\_2 | int | -2.000 | -0.134 | 8.000 | 1.197 | 11 | 15,730 |
| PAY\_3 | int | -2.000 | -0.166 | 8.000 | 1.197 | 11 | 15,764 |
| PAY\_4 | int | -2.000 | -0.221 | 8.000 | 1.169 | 11 | 16,455 |
| PAY\_5 | int | -2.000 | -0.266 | 8.000 | 1.133 | 10 | 16,947 |
| PAY\_6 | int | -2.000 | -0.291 | 8.000 | 1.150 | 10 | 16,286 |
| BILL\_AMT1 | real | -165,580.000 | 51,223.331 | 964,511.000 | 73,635.861 | 22,723 | 2,008 |
| BILL\_AMT2 | real | -69,777.000 | 49,179.075 | 983,931.000 | 71,173.769 | 22,346 | 2,506 |
| BILL\_AMT3 | real | -157,264.000 | 47,013.155 | 1,664,089.000 | 69,349.387 | 22,026 | 2,870 |
| BILL\_AMT4 | real | -170,000.000 | 43,262.949 | 891,586.000 | 64,332.856 | 21,548 | 3,195 |
| BILL\_AMT5 | real | -81,334.000 | 40,311.401 | 927,171.000 | 60,797.156 | 21,010 | 3,506 |
| BILL\_AMT6 | real | -339,603.000 | 38,871.760 | 961,664.000 | 59,554.108 | 20,604 | 4,020 |
| PAY\_AMT1 | real | 0.000 | 5,663.581 | 873,552.000 | 16,563.280 | 7,943 | 5,249 |
| PAY\_AMT2 | real | 0.000 | 5,921.163 | 1,684,259.000 | 23,040.870 | 7,899 | 5,396 |
| PAY\_AMT3 | real | 0.000 | 5,225.681 | 896,040.000 | 17,606.961 | 7,518 | 5,968 |
| PAY\_AMT4 | real | 0.000 | 4,826.077 | 621,000.000 | 15,666.160 | 6,937 | 6,408 |
| PAY\_AMT5 | real | 0.000 | 4,799.388 | 426,529.000 | 15,278.306 | 6,897 | 6,703 |
| PAY\_AMT6 | real | 0.000 | 5,215.503 | 528,666.000 | 17,777.466 | 6,939 | 7,173 |

#### Boolean Columns

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **name** | **data\_type** | **min** | **mean** | **max** | **std** | **freq of max value** |
| default.payment.next.month | bool | False | 0.2212 | True | 0.4151 | 6,636 |

### Shifts Detected

Driverless AI can perform shift detection between the training, validation, and testing datasets. It does this by training a binomial model to predict which dataset a record belongs to. For example, it may find that it is able to separate the training and testing data with an AUC of 0.8 using only the column: C1 as the predictor. This indicates that there is some sort of drift in the distribution of C1 between the training and testing data.

For this experiment, Driverless AI was not able to check for distribution shifts because only the training dataset was supplied by the user.

## Methodology

This section describes the experiment methodology.

### Assumptions and Limitations

Driverless AI trains all models based on the training data provided (in this case: *UCI\_Credit\_Card.csv*). It is the assumption of Driverless AI that this dataset is representative of the data that will be seen when scoring.

Driverless AI may perform shift detection between the train data and another dataset. If a shift in distribution is detected, this may indicate that the data that will be used for scoring may have distributions not represented in the training data.

For this experiment, Driverless AI was not able to detect any shift in distribution between train data and another dataset because no validation or test data was provided.

### Experiment Pipeline

For this experiment, Driverless AI performed the following steps to find the optimal final model:



The steps in this pipeline are described in more detail below:

* **Ingest Data**
* detected column types
* **Feature Preprocessing**
* turned raw features into numeric
* **Model and Feature Tuning**

This stage combines random hyperparameter tuning with feature selection and generation. Features in each iteration are updated using variable importance from the previous iteration as a probabilistic prior to decide what new features to create. The best performing model and features are then passed to the feature evolution stage.

* found the optimal parameters for constant, xgboost and lightgbm models by training models with different parameters
* the best parameters are those that generate the largest **ACCURACY** on the internal validation data
* 6 models trained and scored to evaluate features and model parameters
* **Feature Evolution**

This stage uses a genetic algorithm to find the best set of model parameters and feature transformations to be used in the final model. This experiment did not perform the Feature Evolution stage due to the experiment's configurations.

* **Final Model**
* created the best model from the feature engineering iterations
* no stacked ensemble is done due to accuracy or ensemble level settings (consider increasing accuracy or the ensemble\_level)
* **Create Scoring Pipeline**
* created and exported the MOJO and Python scoring pipeline
* MOJO Scoring Pipeline: admin/h2oai\_experiment\_fd34098a-5c6e-11eb-86d5-0242ac110002/mojo\_pipeline/mojo.zip
* Python Scoring Pipeline: admin/h2oai\_experiment\_fd34098a-5c6e-11eb-86d5-0242ac110002/scoring\_pipeline/scorer.zip

Driverless AI trained models throughout the experiment in an effort to determine the best parameters, model dataset, and optimal final model. The stages are described below:

|  |  |  |
| --- | --- | --- |
| **Driverless AI Stage** | **Timing (seconds)** | **Number of Models** |
| **Data Preparation** | 19.27 | 0 |
| **Model and Feature Tuning** | 61.82 | 6 |
| **Feature Evolution** | 2.34 | 0 |
| **Final Pipeline Training** | 54.89 | 9 |

### Experiment Settings

Below are the settings selected for the experiment by admin. The Defined Parameters represent the high-level parameters.

**Defined Parameters**

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| is\_classification | True |
| enable\_gpus | True |
| seed | False |
| accuracy | 3 |
| time | 1 |
| interpretability | 5 |
| time\_groups\_columns | None |
| num\_prediction\_periods | None |
| num\_gap\_periods | None |
| is\_timeseries | False |
| is\_image | False |

These Accuracy, Time, and Interpretability settings map to the following internal configuration of the Driverless AI experiment:

|  |  |
| --- | --- |
| **Internal Parameter** | **Value** |
| data filtered | False |
| number of feature engineering iterations | 1 |
| number of models trained per iteration | 2 |
| early stopping rounds | 0 |
| monotonicity constraint | False |
| number of model tuning model combinations | 5 |
| number of base learners in ensemble | 0 |
| time column | [OFF] |

#### Details

* **data filtered:** Driverless AI may filter the training data depending on the number of rows and the Accuracy setting.
* for this experiment, the training data was not filtered.
* **number of feature engineering iterations**: the number of iterations performed of feature engineering.
* **number of models evaluated per iteration:** for each feature engineering iteration, Driverless AI trains multiple models. Each model is trained with a different set of predictors or features. The goal of this step is to determine which types of features lead to the largest ACCURACY.
* **early stopping rounds:** if Driverless AI does not see any improvement after 0 iterations of feature engineering, the feature engineering step is automatically stopped.
* **monotonicity constraint:** if enabled, the models will only have monotone relationships between the predictors and target variable.
* **number of model tuning combinations:** the number of model tuning combinations evaluated to determine the optimal model settings for the models.
* **number of base learners in ensemble:** the number of base models used to create the final ensemble.
* **time column:** the column that provides the time column. If a time column is provided, feature engineering and model validation will respect the causality of time. If the time column is turned off, no time order is used for modeling and data may be shuffled randomly (any potential temporal causality will be ignored).

## Data Sampling

Driverless AI did not perform any down sampling of the data.

## Validation Strategy

Driverless AI automatically split the training data to determine the performance of the model parameter tuning and feature engineering steps. For the experiment, Driverless AI randomly split the data into **3/4 training** and **1/4 validation**.

## Model Tuning

The table below shows the score and training time of the constant, xgboost and lightgbm models evaluated by Driverless AI. The table shows the parameter tuning models evaluated, ordered based on a combination of largest score and lowest training time.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **job order** | **booster** | **nfeatures** | **scores** | **training times** |
| 0 | lightgbm | 22 | 0.8244 | 1.8575 |
| 1 | gbtree | 22 | 0.8227 | 24.7819 |
| 3 | lightgbm | 22 | 0.8203 | 3.3474 |
| 2 | gbtree | 22 | 0.8196 | 4.9178 |
| 4 | constant | 1 | 0.218 | 0.3013 |

More detailed information on the parameters evaluated for each algorithm is shown below.

### constant tuning

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **job order** | **booster** | **nfeatures** | **scores** | **training times** |
| 4 | constant | 1 | 0.218 | 0.3013 |

### gbtree tuning

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **tree method** | **grow policy** | **max depth** | **max leaves** | **colsample bytree** | **subsample** | **nfeatures** | **scores** | **training times** |
| gpu\_hist | depthwise | 6.0 | 0.0 | 0.8 | 0.7 | 22 | 0.8227 | 24.7819 |
| gpu\_hist | lossguide | 0.0 | 256.0 | 0.55 | 0.9 | 22 | 0.8196 | 4.9178 |

### lightgbm tuning

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **tree method** | **grow policy** | **max depth** | **max leaves** | **colsample bytree** | **subsample** | **nfeatures** | **scores** | **training times** |
| gpu\_hist | depthwise | 6.0 | 0.0 | 0.8 | 0.7 | 22 | 0.8244 | 1.8575 |
| gpu\_hist | lossguide | 0.0 | 256.0 | 0.4 | 0.6 | 22 | 0.8203 | 3.3474 |

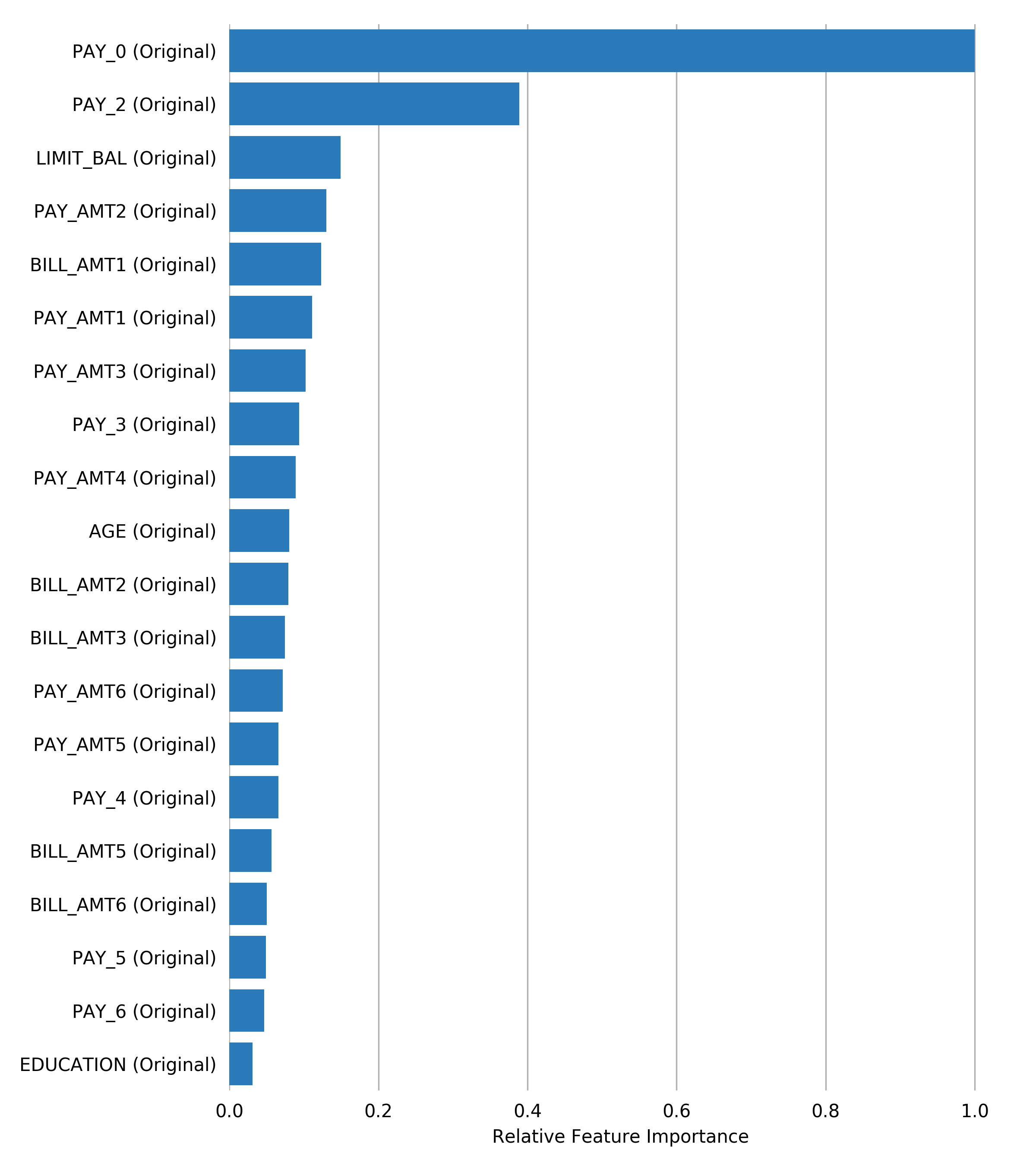
## Feature Evolution

The goal of the Feature Evolution stage is to determine the best features to use for the final model. This experiment did not perform the Feature Evolution stage due to the experiment's configurations.

## Feature Transformations

The result of the Feature Evolution Stage is a set of features to use for the final model. The top features used in the final model are shown below, ordered by importance. The features in the table are limited to the top 50, restricted to those with relative importance greater than or equal to 0.003. If no transformer was applied, the feature is an original column.

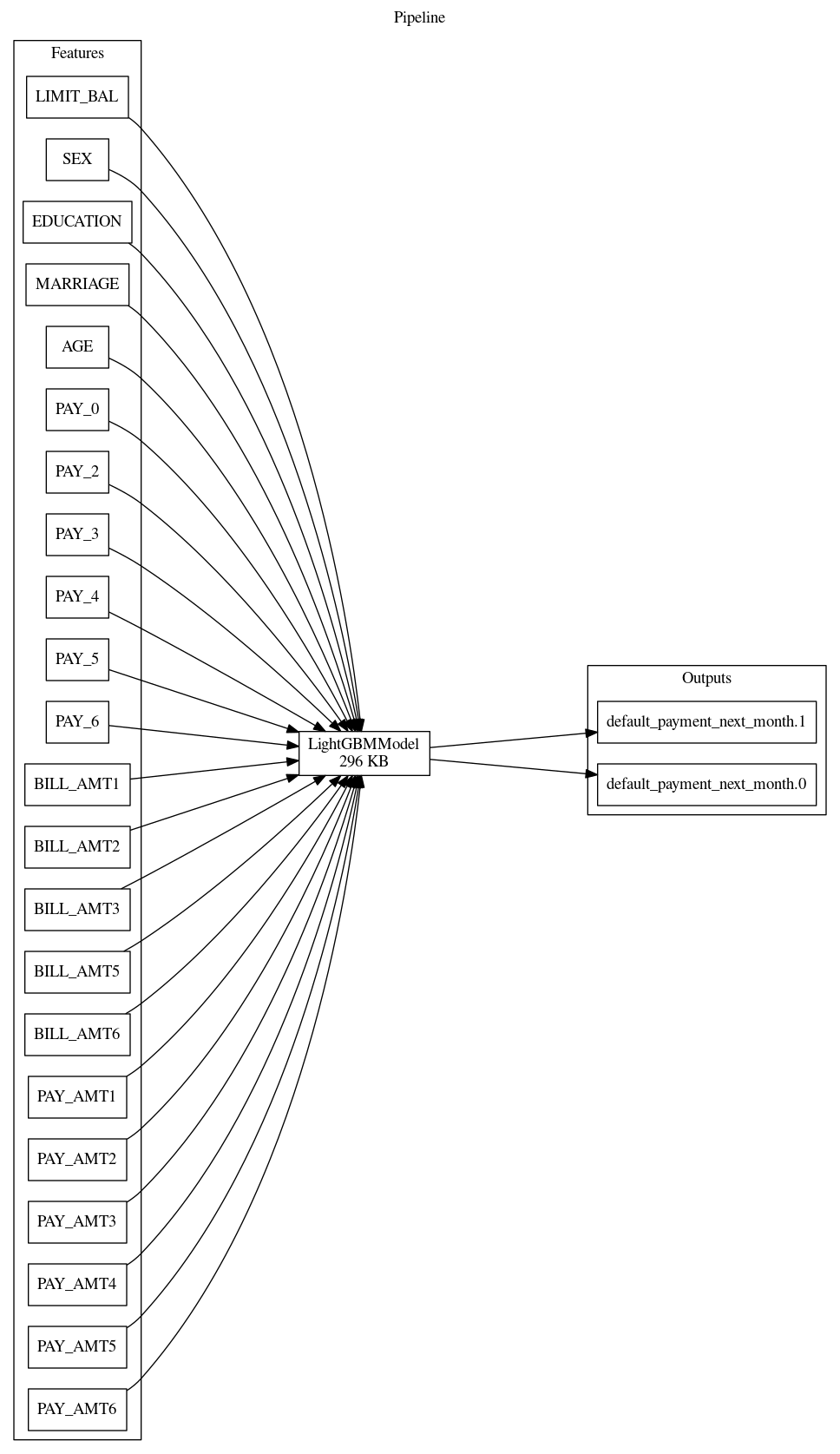
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Feature** | **Description** | **Transformer** | **Relative Importance** |
| 1 | 10\_PAY\_0 | PAY\_0 (Original) | None | 1.0 |
| 2 | 11\_PAY\_2 | PAY\_2 (Original) | None | 0.3891 |
| 3 | 8\_LIMIT\_BAL | LIMIT\_BAL (Original) | None | 0.149 |
| 4 | 17\_PAY\_AMT2 | PAY\_AMT2 (Original) | None | 0.1299 |
| 5 | 1\_BILL\_AMT1 | BILL\_AMT1 (Original) | None | 0.1234 |
| 6 | 16\_PAY\_AMT1 | PAY\_AMT1 (Original) | None | 0.1108 |
| 7 | 18\_PAY\_AMT3 | PAY\_AMT3 (Original) | None | 0.1022 |
| 8 | 12\_PAY\_3 | PAY\_3 (Original) | None | 0.0937 |
| 9 | 19\_PAY\_AMT4 | PAY\_AMT4 (Original) | None | 0.0892 |
| 10 | 0\_AGE | AGE (Original) | None | 0.0801 |
| 11 | 2\_BILL\_AMT2 | BILL\_AMT2 (Original) | None | 0.0791 |
| 12 | 3\_BILL\_AMT3 | BILL\_AMT3 (Original) | None | 0.0746 |
| 13 | 21\_PAY\_AMT6 | PAY\_AMT6 (Original) | None | 0.0717 |
| 14 | 20\_PAY\_AMT5 | PAY\_AMT5 (Original) | None | 0.066 |
| 15 | 13\_PAY\_4 | PAY\_4 (Original) | None | 0.0659 |
| 16 | 5\_BILL\_AMT5 | BILL\_AMT5 (Original) | None | 0.0566 |
| 17 | 6\_BILL\_AMT6 | BILL\_AMT6 (Original) | None | 0.0502 |
| 18 | 14\_PAY\_5 | PAY\_5 (Original) | None | 0.0492 |
| 19 | 15\_PAY\_6 | PAY\_6 (Original) | None | 0.0465 |
| 20 | 7\_EDUCATION | EDUCATION (Original) | None | 0.0312 |
| 21 | 9\_MARRIAGE | MARRIAGE (Original) | None | 0.0176 |
| 22 | 22\_SEX | SEX (Original) | None | 0.0107 |



## Final Model

**Pipeline**

Final LightGBMModel pipeline with ensemble\_level=0 transforming 22 original features -> 22 features in each of 1 models each of 4 fold hyperparameters averaged and re-fit as single model.:



**Details**

* The fitted features of the final model are the best features found during the feature engineering iterations.
* The target transformer indicates the type of transformation applied to the target column.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model Index** | **Type** | **Model Weight** | **Num Folds** | **Fitted features** | **Target Transformer** |
| 0 | LightGBMModel | 1 | 4 | 22 | LabelEncoder |

* Model Index: 0 has a weight of 1 in the final ensemble

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Type** | **subsample** | **grow policy** | **tree method** | **model class name** | **learning rate** | **index** | **max depth** | **colsample bytree** | **max leaves** |
| LightGBMModel | 0.7 | depthwise | gpu\_hist | LightGBMModel | 0.05 | 0 | 6 | 0.8 | 64 |

For a complete list of the parameters of the final model, see the Appendix.

**Performance of Final Model**

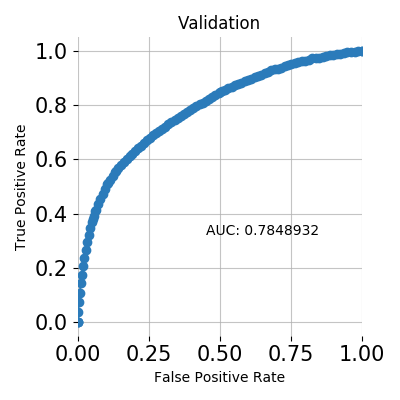
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Scorer** | **Optimized** | **Better score is** | **Final ensemble scores on validation (internal or external holdout(s)) data** | **Final ensemble standard deviation on validation (internal or external holdout(s)) data** |
| ACCURACY | \* | higher | 0.8209333 | 0.001615531 |
| AUC |  | higher | 0.7848932 | 0.002849556 |
| AUCPR |  | higher | 0.5615535 | 0.006989505 |
| F05 |  | higher | 0.5811366 | 0.005684444 |
| F1 |  | higher | 0.5487904 | 0.005021418 |
| F2 |  | higher | 0.6415051 | 0.003496438 |
| FDR |  | lower | 0.4609795 | 0.01148195 |
| FNR |  | lower | 0.441079 | 0.01169147 |
| FOR |  | lower | 0.1266058 | 0.002657329 |
| FPR |  | lower | 0.1357644 | 0.008099831 |
| GINI |  | higher | 0.5697864 | 0.005699111 |
| LOGLOSS |  | lower | 0.4256409 | 0.002667523 |
| MACROAUC |  | higher | 0.7848932 | 0.002849556 |
| MCC |  | higher | 0.4236971 | 0.005185404 |
| NPV |  | higher | 0.8733942 | 0.002657329 |
| PRECISION |  | higher | 0.5390205 | 0.01148195 |
| RECALL |  | higher | 0.558921 | 0.01169147 |
| TNR |  | higher | 0.8642356 | 0.008099831 |

**Validation Confusion Matrix**

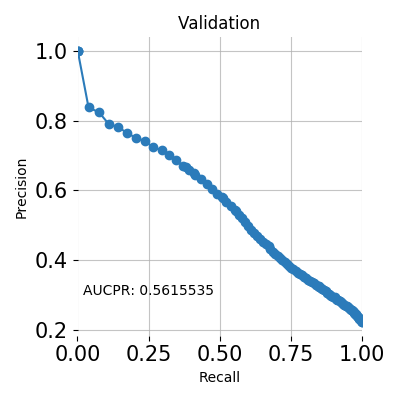
*Threshold 0.28476849200000004*

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Predicted: 0** | **Predicted: 1** | **error** |
| Actual: 0 | 20,192 | 3,172 | 14% |
| Actual: 1 | 2,927 | 3,709 | 44% |

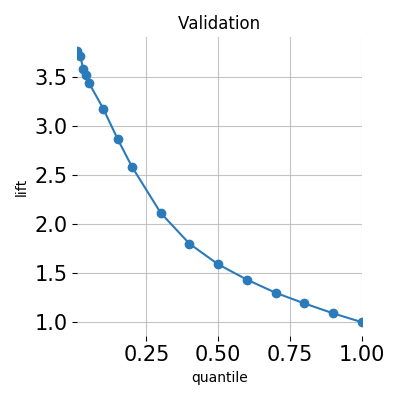
*Receiver Operating Characteristic Curve*



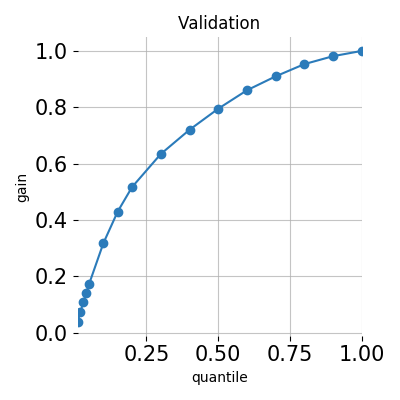
*Precision Recall Curve*



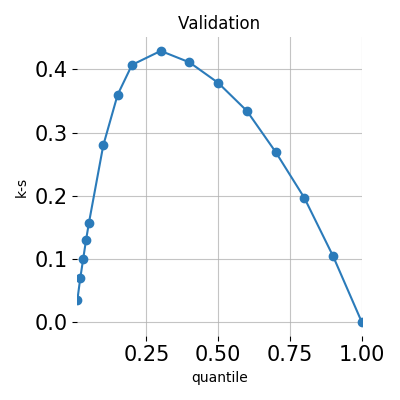
*Cumulative Lift*



*Cumulative Gains*



*Kolmogorov–Smirnov*



## Alternative Models

During the experiment, Driverless AI trained 6 alternative models. The following algorithms were evaluated during the Driverless AI experiment:

|  |  |  |  |
| --- | --- | --- | --- |
| **algorithm** | **package** | **version** | **documentation** |
| constant | custom package | 1.9.0 | reference model that predicts a constant aimed at minimizing the given scorer |
| gbtree | xgboost | 1.1.0 | XGBoost: eXtreme Gradient Boosting library. Contributors: https://github.com/dmlc/xgboost/blob/master/CONTRIBUTORS.md |
| lightgbm | lightgbm | 2.2.4 | LightGBM, Light Gradient Boosting Machine. Contributors: https://github.com/microsoft/LightGBM/graphs/contributors. |

Driverless AI can evaluate an array of algorithms, including but not limited to XGBoost GBM, XGBoost Dart, XGBoost GLM, LightGBM, RuleFit, Tensorflow, and FTRL models. The table below explains why certain algorithms were not selected for the final model, if any.

|  |  |
| --- | --- |
| **algorithm** | **selection** |
| gblinear | algorithm not evaluated due to experiment configuration |
| decision tree | algorithm not evaluated due to experiment configuration |
| rulefit | algorithm not evaluated due to experiment configuration |
| tensorflow | algorithm not evaluated due to experiment configuration |
| ftrl | algorithm not evaluated due to experiment configuration |
| dart | algorithm not evaluated due to experiment configuration |
| gbtree | not selected due to low performance during model tuning stage |
| lightgbm | not selected due to low performance during model tuning stage |

## Deployment

For this experiment, both Python and MOJO Scoring Pipelines are available for productionizing the final model pipeline for a given row of data or table of data.

### Python Scoring Pipeline

This package contains an exported model and Python 3.6 source code examples for productionizing models built using H2O Driverless AI. The Python Scoring Pipeline is located here:

* **admin/h2oai\_experiment\_fd34098a-5c6e-11eb-86d5-0242ac110002/scoring\_pipeline/scorer.zip**

The files in this package allow you to transform and score on new data in a couple of different ways:

* From Python 3.6, you can import a scoring module, then use the module to transform and score on new data.
* From other languages and platforms, you can use the TCP/HTTP scoring service bundled with this package to call into the scoring pipeline module through remote procedure calls (RPC).

### MOJO Scoring Pipeline

Note: The MOJO Scoring Pipeline is currently in a beta state. Updates and improvements will continue to be made in subsequent Driverless AI releases. The MOJO Scoring Pipeline is located here:

* **admin/h2oai\_experiment\_fd34098a-5c6e-11eb-86d5-0242ac110002/mojo\_pipeline/mojo.zip**

For completed experiments, Driverless AI converts models to MOJOs (Model Objects, Optimized). A MOJO is a scoring engine that can be deployed in any Java environment for scoring in real time.

## Partial Dependence Plots

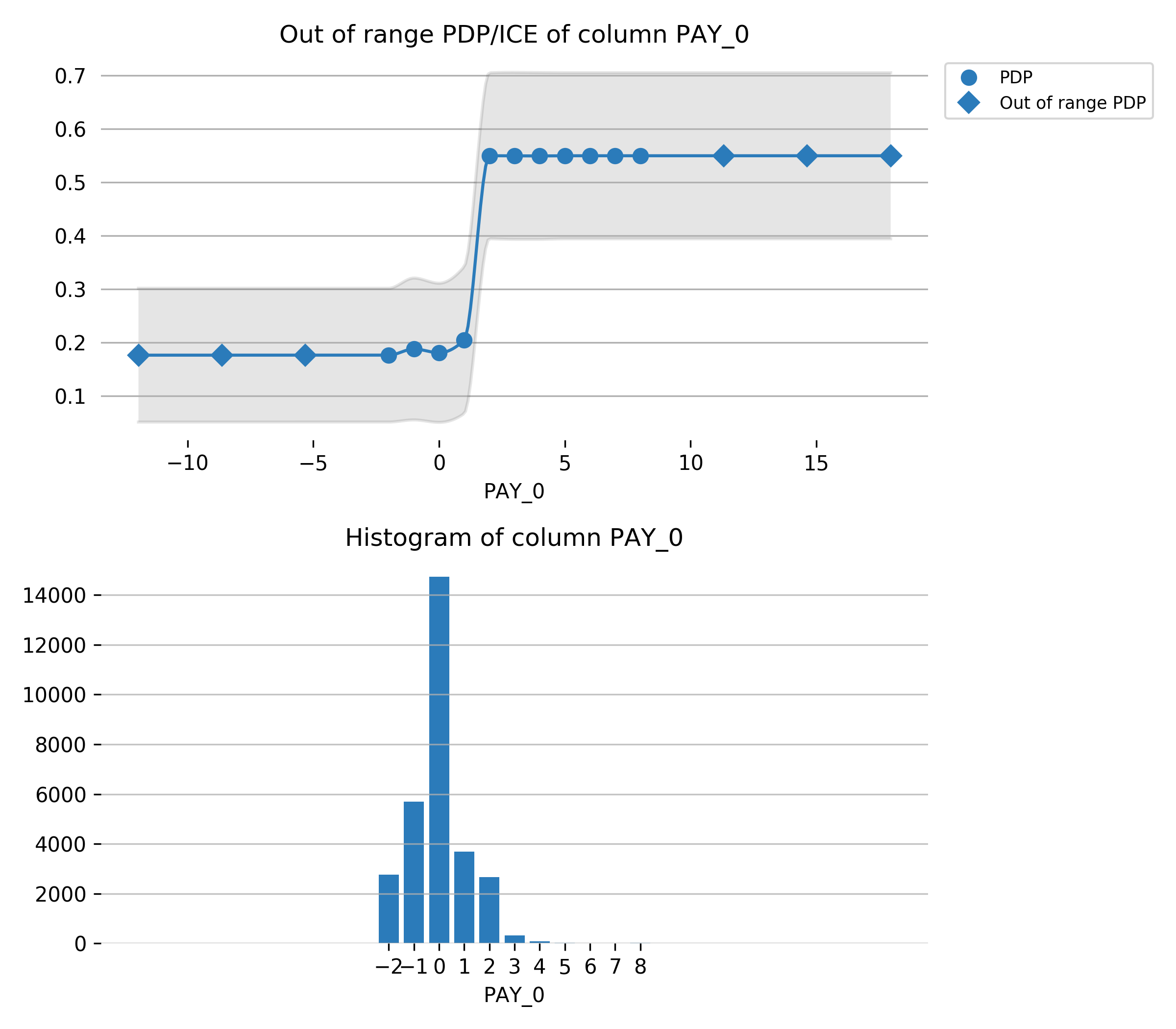
Partial dependence plots show the partial dependence as a function of specific values for a feature subset. The plots show how machine-learned response functions change based on the values of an input feature of interest, while taking nonlinearity into consideration and averaging out the effects of all other input features. Partial dependence plots enable increased transparency in a model and enable the ability to validate and debug a model by comparing a feature's average predictions across its domain to known standards and reasonable expectations.

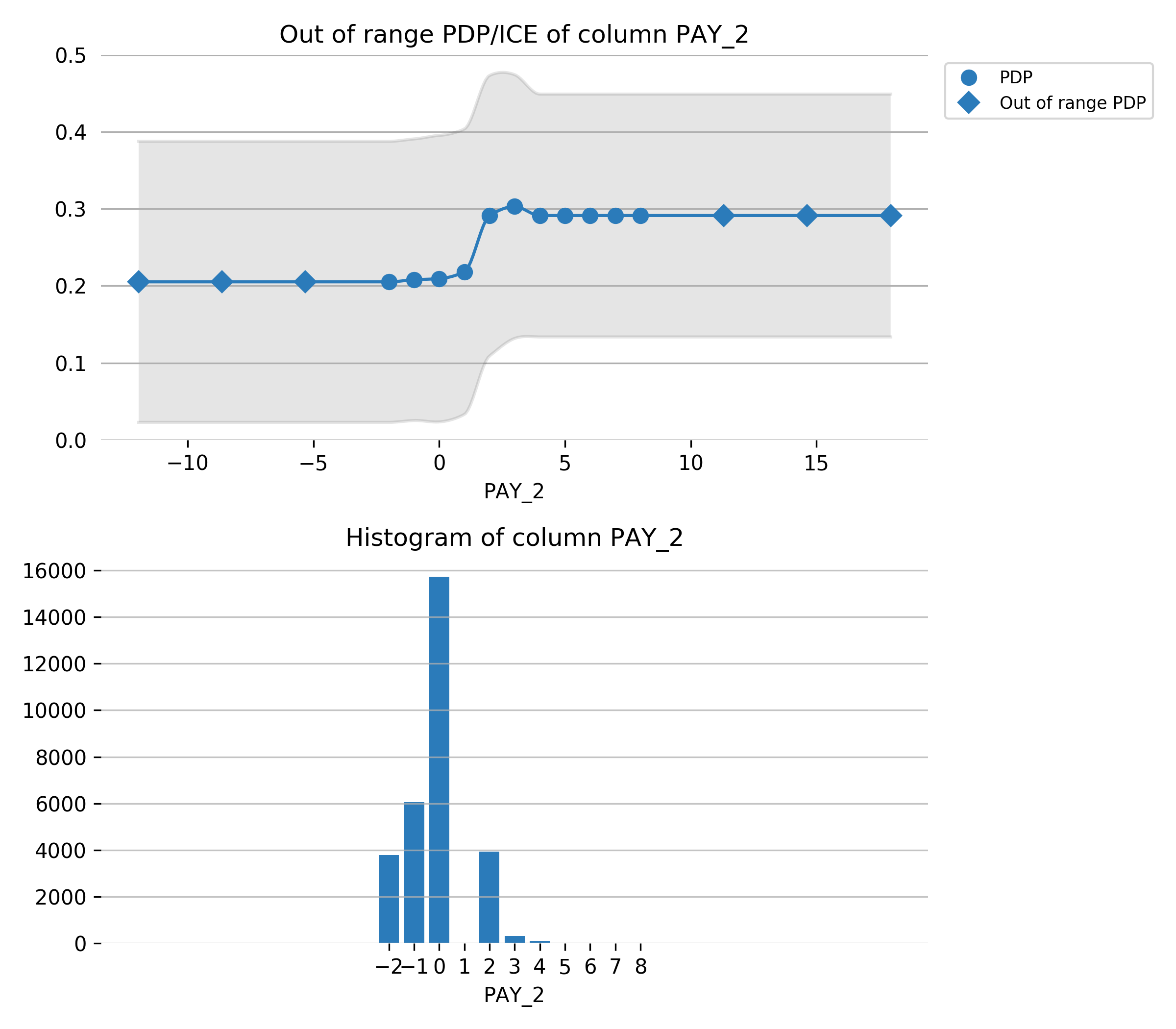
The partial dependence plots are shown for the top 3 original variables. The top 3 original variables are chosen based on their Component Based Variable Importance. Partial Dependence computation reached maximum allowed time 20 seconds.

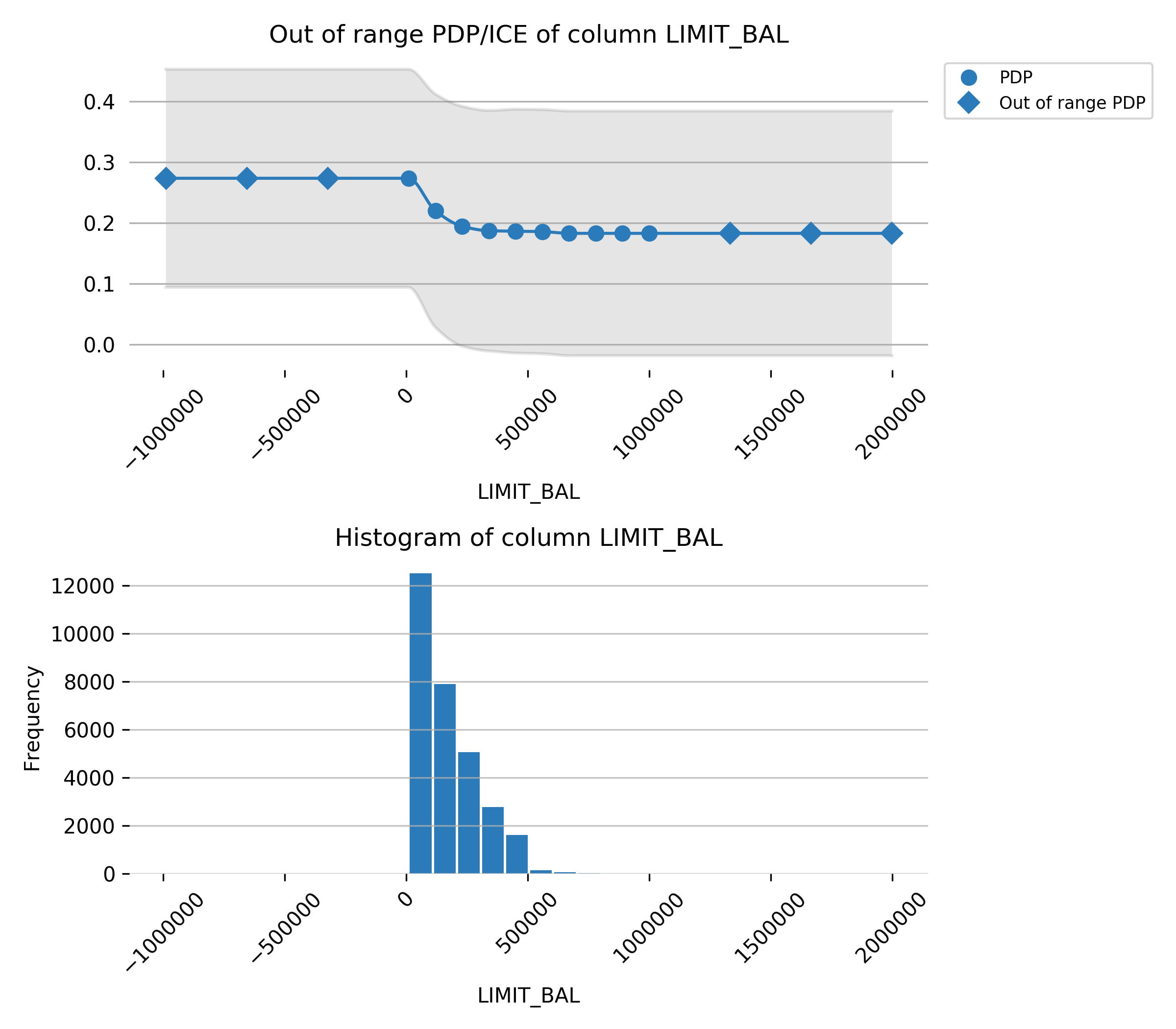
**Plot Details**

In the Driverless AI PDP, the y-axis represents the mean response, and a shaded region (for numeric features) or shaded bar (for categorical features) represents 1 standard deviation. Out-of-range PDP (diamond markers) represent values outside feature intervals seen in the data, unseen categorical values, or missing values.

For continuous features, numeric values up to 3 standard deviations lower than the minimum training value and higher than the maximum training value are feed into the model. For categorical features, an unseen categorical value is feed into the model denoted by UNSEEN (if the categorical value "UNSEEN" already exists in the training data, the out-of-range is done on a value called "UNSEEN\_[x]," where x is some integer).

Feature **PAY\_0**

Feature **PAY\_2**

Feature **LIMIT\_BAL**

## Appendix

### Final Model Details

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model Index** | **Type** | **Model Weight** | **Num Folds** | **Fitted features** | **Target Transformer** |
| 0 | LightGBMModel | 1 | 4 | 22 | LabelEncoder |

**Model Index: Final Model - Single Model**

|  |  |
| --- | --- |
| **parameter** | **value** |
| accuracy | 3 |
| booster | lightgbm |
| boosting\_type | gbdt |
| colsample\_bytree | 0.8 |
| disable\_gpus | False |
| early\_stopping\_rounds |  |
| enable\_early\_stopping\_rounds | False |
| encoder |  |
| ensemble\_level | 1 |
| eval\_metric | logloss |
| gamma | 0 |
| gpu\_id | 0 |
| grow\_policy | depthwise |
| interpretability | 5 |
| labels | [0, 1] |
| learning\_rate | 0.05 |
| lossguide | False |
| max\_bin | 256 |
| max\_delta\_step | 0 |
| max\_depth | 6 |
| max\_leaves | 64 |
| min\_child\_samples | 1 |
| min\_child\_weight | 1 |
| min\_data\_in\_bin | 1 |
| model\_class\_name | LightGBMModel |
| model\_id | Final Model - Single Model |
| model\_origin | SEQUENCE |
| monotonicity\_constraints | False |
| n\_estimators | 103 |
| n\_gpus | 1 |
| n\_jobs | 2 |
| num\_class | 1 |
| num\_classes | 2 |
| objective | binary:logistic |
| pred\_gap |  |
| pred\_periods |  |
| random\_state | 487425905 |
| reg\_alpha | 0.0 |
| reg\_lambda | 1.0 |
| scale\_pos\_weight | 1 |
| score\_f\_name | ACCURACY |
| seed | 487425905 |
| silent | True |
| subsample | 0.7 |
| subsample\_freq | 1 |
| target |  |
| tgc |  |
| time\_column |  |
| time\_tolerance | 1 |
| train\_shape | [30000, 25] |
| tree\_method | gpu\_hist |
| tsp |  |
| valid\_shape |  |
| nfolds | 1 |

### Config Overrides

The Config Overrides represent the fine-control parameters.

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| vis\_server\_port | 12346 |
| procsy\_port | 12347 |
| h2o\_port | 12348 |
| h2o\_recipes\_kwargs | {} |
| redis\_port | 6379 |
| authentication\_method | none |
| supported\_file\_types | ['csv', 'tsv', 'txt', 'dat', 'tgz', 'gz', 'bz2', 'zip', 'xz', 'xls', 'xlsx', 'jay', 'feather', 'bin', 'arff', 'parquet', 'pkl', 'orc'] |
| recipe\_supported\_file\_types | ['py', 'pyc'] |
| enabled\_file\_systems | ['upload', 'file', 'hdfs', 's3', 'recipe\_file', 'recipe\_url'] |
| missing\_values | ['', '?', 'None', 'nan', 'NA', 'N/A', 'unknown', 'inf', '-inf', '1.7976931348623157e+308', '-1.7976931348623157e+308'] |
| monotonicity\_constraints\_dict | {} |
| params\_lightgbm | {} |
| params\_xgboost | {} |
| params\_dart | {} |
| params\_tensorflow | {} |
| params\_gblinear | {} |
| params\_decision\_tree | {} |
| params\_rulefit | {} |
| params\_ftrl | {} |
| params\_tune\_lightgbm | {} |
| params\_tune\_xgboost | {} |
| params\_tune\_dart | {} |
| params\_tune\_tensorflow | {} |
| params\_tune\_gblinear | {} |
| params\_tune\_rulefit | {} |
| params\_tune\_ftrl | {} |
| ts\_target\_trafo\_epidemic\_params\_dict | {} |
| mli\_cloud\_name | H2O-MLI-DAI-18973 |
| recipe\_dict | {} |
| cuda\_visible\_devices | None |
| last\_recipe | auto |
| included\_transformers | ['BERTTransformer', 'CVCatNumEncodeTransformer', 'CVTargetEncodeTransformer', 'CatOriginalTransformer', 'CatTransformer', 'ClusterDistTransformer', 'ClusterTETransformer', 'DateOriginalTransformer', 'DateTimeOriginalTransformer', 'DatesTransformer', 'EwmaLagsTransformer', 'FrequentTransformer', 'ImageOriginalTransformer', 'ImageVectorizerTransformer', 'InteractionsTransformer', 'IsHolidayTransformer', 'IsolationForestAnomalyNumCatAllColsTransformer', 'IsolationForestAnomalyNumCatTransformer', 'IsolationForestAnomalyNumericTransformer', 'LagsAggregatesTransformer', 'LagsInteractionTransformer', 'LagsTransformer', 'LexiLabelEncoderTransformer', 'NumCatTETransformer', 'NumToCatTETransformer', 'NumToCatWoEMonotonicTransformer', 'NumToCatWoETransformer', 'OneHotEncodingTransformer', 'OriginalTransformer', 'TextBiGRUTransformer', 'TextCNNTransformer', 'TextCharCNNTransformer', 'TextLinModelTransformer', 'TextOriginalTransformer', 'TextTransformer', 'TruncSVDNumTransformer', 'WeightOfEvidenceTransformer'] |
| included\_models | ['CONSTANT', 'DECISIONTREE', 'FTRL', 'GLM', 'IMAGEAUTO', 'IMBALANCEDLIGHTGBM', 'IMBALANCEDXGBOOSTGBM', 'LIGHTGBM', 'RULEFIT', 'TENSORFLOW', 'TEXTALBERT', 'TEXTBERT', 'TEXTCAMEMBERT', 'TEXTDISTILBERT', 'TEXTROBERTA', 'TEXTXLM', 'TEXTXLMROBERTA', 'TEXTXLNET', 'XGBOOSTDART', 'XGBOOSTGBM', 'ZEROINFLATEDLIGHTGBM', 'ZEROINFLATEDXGBOOST'] |
| included\_scorers | ['ACCURACY', 'AUC', 'AUCPR', 'F05', 'F1', 'F2', 'FDR', 'FNR', 'FOR', 'FPR', 'GINI', 'LOGLOSS', 'MACROAUC', 'MAE', 'MAPE', 'MCC', 'MER', 'MSE', 'NPV', 'PRECISION', 'R2', 'R2COD', 'RECALL', 'RMSE', 'RMSLE', 'RMSPE', 'SMAPE', 'TNR'] |
| prob\_lag\_non\_targets | 0.1 |
| prob\_default\_lags | 0.2 |
| prob\_lagsinteraction | 0.2 |
| prob\_lagsaggregates | 0.2 |
| included\_transformers\_user | ['BERTTransformer', 'CVCatNumEncodeTransformer', 'CVTargetEncodeTransformer', 'CatOriginalTransformer', 'CatTransformer', 'ClusterDistTransformer', 'ClusterTETransformer', 'DateOriginalTransformer', 'DateTimeOriginalTransformer', 'DatesTransformer', 'EwmaLagsTransformer', 'FrequentTransformer', 'ImageOriginalTransformer', 'ImageVectorizerTransformer', 'InteractionsTransformer', 'IsHolidayTransformer', 'IsolationForestAnomalyNumCatAllColsTransformer', 'IsolationForestAnomalyNumCatTransformer', 'IsolationForestAnomalyNumericTransformer', 'LagsAggregatesTransformer', 'LagsInteractionTransformer', 'LagsTransformer', 'LexiLabelEncoderTransformer', 'NumCatTETransformer', 'NumToCatTETransformer', 'NumToCatWoEMonotonicTransformer', 'NumToCatWoETransformer', 'OneHotEncodingTransformer', 'OriginalTransformer', 'TextBiGRUTransformer', 'TextCNNTransformer', 'TextCharCNNTransformer', 'TextLinModelTransformer', 'TextOriginalTransformer', 'TextTransformer', 'TruncSVDNumTransformer', 'WeightOfEvidenceTransformer'] |