H2O-3 Experiment

Generated by: AutoDoc Team

Generated on: 2020-11-10 09:54

Experiment Overview

H2O-3 built a Gradient Boosting Machine to predict *Churn* given 18 original features from the input dataset. This classification experiment completed in 684 milliseconds (0:00:00.684000).

Performance

Dataset	**auc**
Validation Data	0.871
Test Data	0.857

System Specifications

Attribute	**Value**
H2O cluster uptime:	1 min 16 secs
H2O cluster timezone:	America/Denver

H2O data parsing timezone:	UTC
H2O cluster version:	3.32.0.1
H2O cluster version age:	1 month and 1 day
H2O cluster name:	AutoDoc Team
H2O cluster total nodes:	1
H2O cluster free memory:	3.515 Gb
H2O cluster total cores:	16
H2O cluster allowed cores:	16
H2O cluster status:	locked, healthy
H2O connection url:	http://127.0.0.1:54321
H2O connection proxy:	{'http': None, 'https': None}
H2O internal security:	False
H2O API Extensions:	Amazon S3, XGBoost, Algos, AutoML, Core V3, TargetEncoder, Core V4
Python version:	3.7.4 final

Versions

H2O-3	3.32.0.1

Data Overview

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

data	rows	cols
train	1,735	19
validation	1,598	19
test	780	19

Training Data

The training data consists of both numeric and categorical columns.

The summary of the columns is shown below:

Numeric Columns

name	min	mean	max	std
Account length	1	101.2	225	40.04
Area code	408	437.1	510	42.23
Number vmail messages	0	8.186	51	13.73
Total day minutes	176.6	221.5	350.8	33.05
Total day calls	30	100.5	160	20.06
Total day charge	30.02	37.66	59.64	5.618

Total eve minutes	0	200.9	363.7	50.49
Total eve calls	0	100.2	168	19.92
Total eve charge	0	17.08	30.91	4.291
Total night minutes	43.7	200.7	377.5	50.73
Total night calls	36	100.7	166	19.69
Total night charge	1.97	9.03	16.99	2.283
Total intl minutes	0	10.23	20	2.806
Total intl calls	0	4.489	19	2.463
Total intl charge	0	2.761	5.4	0.7575
Customer service calls	0	1.534	9	1.278

Categorical Columns

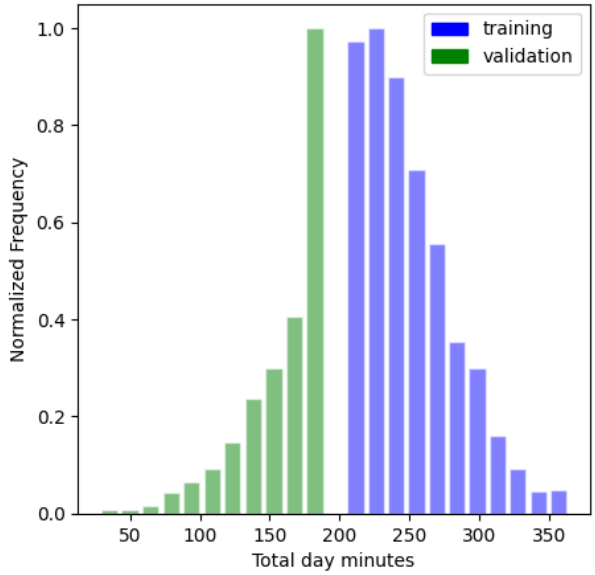
name	unique	top	freq of top value
International plan	2	No	1555
Voice mail plan	2	No	1251
Churn	2	False	1438

Shifts Detected

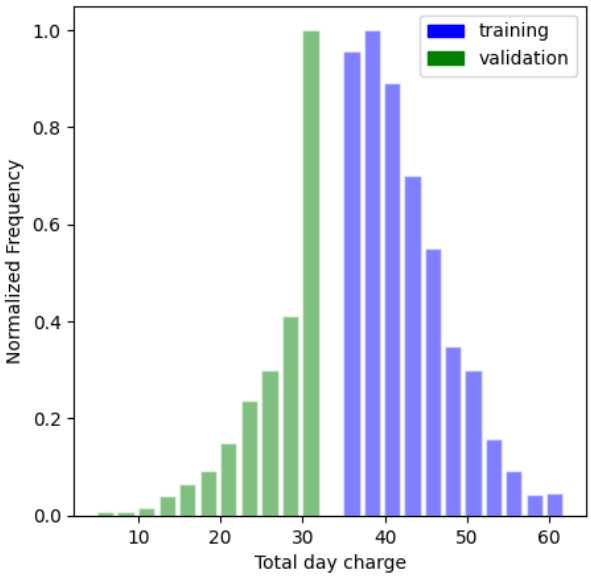
H2O-3 can perform shift detection between the training, validation, and testing datasets. It does this by training a binomial model to predict which dataset a record belongs to. For example, it may find that it is able to separate the training and testing data with an AUC of 0.8 using only the column: 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, H2O-3 checked the train, validation, and test data for any shift in distribution and found the following significant differences:

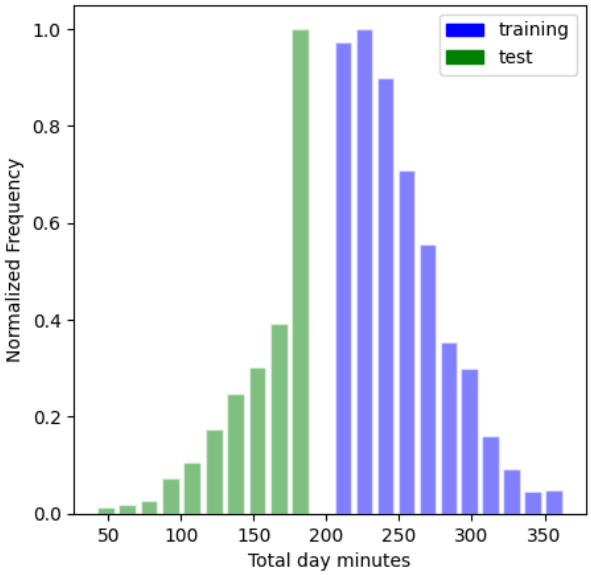
1. Significant difference detected between training and validation data distribution for feature <<<Total day minutes>>> (AUC: 1.0).



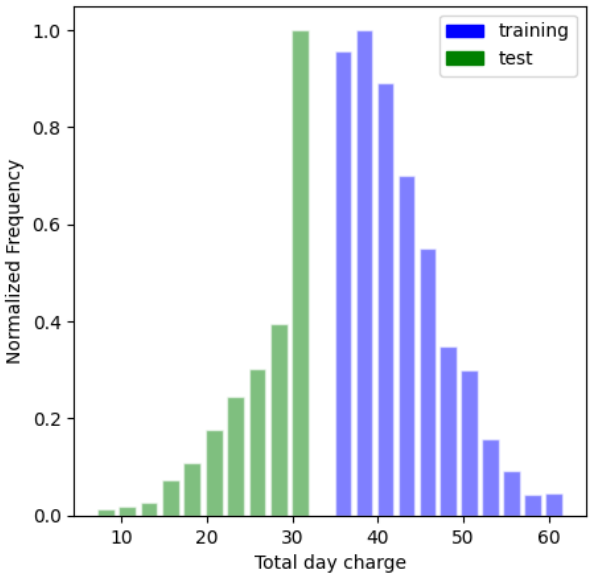
2. Significant difference detected between training and validation data distribution for feature <<<Total day charge>>> (AUC: 1.0).



3. Significant difference detected between training and test data distribution for feature <<<Total day minutes>>> (AUC: 1.0).



4. Significant difference detected between training and test data distribution for feature <<<Total day charge>>> (AUC: 1.0).



Validation Strategy

The model's performance is evaluated using a validation dataset with shape (1598, 19).

Early stopping ends model training when the selected "stopping metric" does not improve for a specified number of training rounds, based on a simple moving average. For example, this model has the following early stopping configurations:

- `stopping_rounds: 3`
- `stopping_metric: logloss`
- `stopping_tolerance: 0.01`

This means the moving average for the last ***4*** stopping rounds is calculated (the first moving average is a reference value to which the other ***3*** moving averages are compared).

The model is set to stop building if the ***logloss*** doesn't improve by ***0.01*** after ***3*** stopping rounds. Specifically, this model stops building, if the ratio between the best moving average and the reference moving average is greater than or equal to ***1 + 0.01***. These stopping configurations restrict the number of model iterations in order to increase the model's performance.

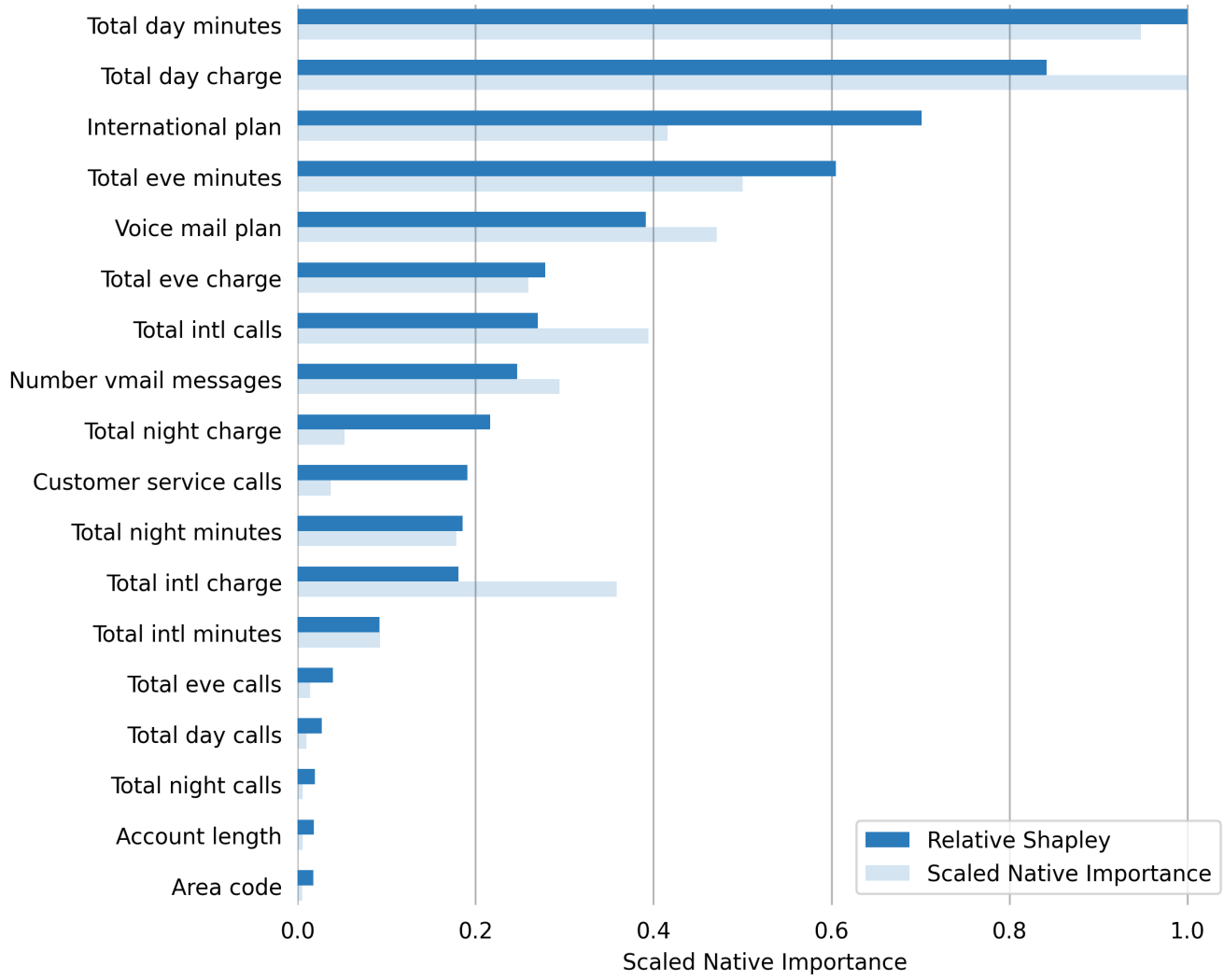
Feature Importance

H2O-3 models provide built-in variable importance (Native Importance) and can provide Shapley Importance for supported algorithms.

- **Native Importance:** Model-specific variable importance calculated with H2O-3's *varimp()* function (H2O-3 documentation details here).
- **Scaled Native Importance:** Native Importance scaled between 0 and 1.
- **Shapley:** The mean absolute Shapley value of a feature, using TreeSHAP (SHAP documentation details here).
- **Relative Shapley:** The feature's mean absolute Shapley value divided by the largest Shapley value.

	Feature	Native Importance	Scaled Native Importance	Shapley	Relative Shapley
0	Total day minutes	164.9468	0.9479	0.3423	1.0
1	Total day charge	174.0203	1.0	0.2881	0.8417
2	International plan	72.3065	0.4155	0.2401	0.7014
3	Total eve minutes	87.0817	0.5004	0.2071	0.6051
4	Voice mail plan	82.0109	0.4713	0.134	0.3915
5	Total eve charge	45.1503	0.2595	0.0953	0.2785
6	Total intl calls	68.5843	0.3941	0.0924	0.2698
7	Number vmail messages	51.2437	0.2945	0.0845	0.247

8	Total night charge	9.2126	0.0529	0.0741	0.2164
9	Customer service calls	6.5248	0.0375	0.0653	0.1907
10	Total night minutes	31.0586	0.1785	0.0635	0.1856
11	Total intl charge	62.4649	0.359	0.062	0.181
12	Total intl minutes	16.1612	0.0929	0.0315	0.092
13	Total eve calls	2.4341	0.014	0.0137	0.0399
14	Total day calls	1.7746	0.0102	0.0093	0.0273
15	Total night calls	1.0209	0.0059	0.0066	0.0194
16	Account length	1.0264	0.0059	0.0063	0.0183
17	Area code	0.9242	0.0053	0.0062	0.018



Final Model

Parameters	Values
model_id	gbm
balance_classes	False
categorical_encoding	Enum

class_sampling_factors	None
col_sample_rate	1.0
col_sample_rate_change_per_level	1.0
col_sample_rate_per_tree	1.0
distribution	bernoulli
fold_assignment	None
fold_column	None
histogram_type	UniformAdaptive
learn_rate	0.1
learn_rate_annealing	1.0
max_abs_leafnode_pred	1.7976931348623157e+308
max_after_balance_size	5.0
max_depth	5
max_runtime_secs	0.0
min_rows	10.0
min_split_improvement	1e-05
nbins	20
nbins_cats	1024
nbins_top_level	1024

ntrees	50
offset_column	None
pred_noise_bandwidth	0.0
response_column	Churn
sample_rate	1.0
sample_rate_per_class	None
seed	1234
stopping_metric	logloss
stopping_rounds	3
stopping_tolerance	0.01
validation_frame	py_11_sid_8916
weights_column	None

Performance of Final Model

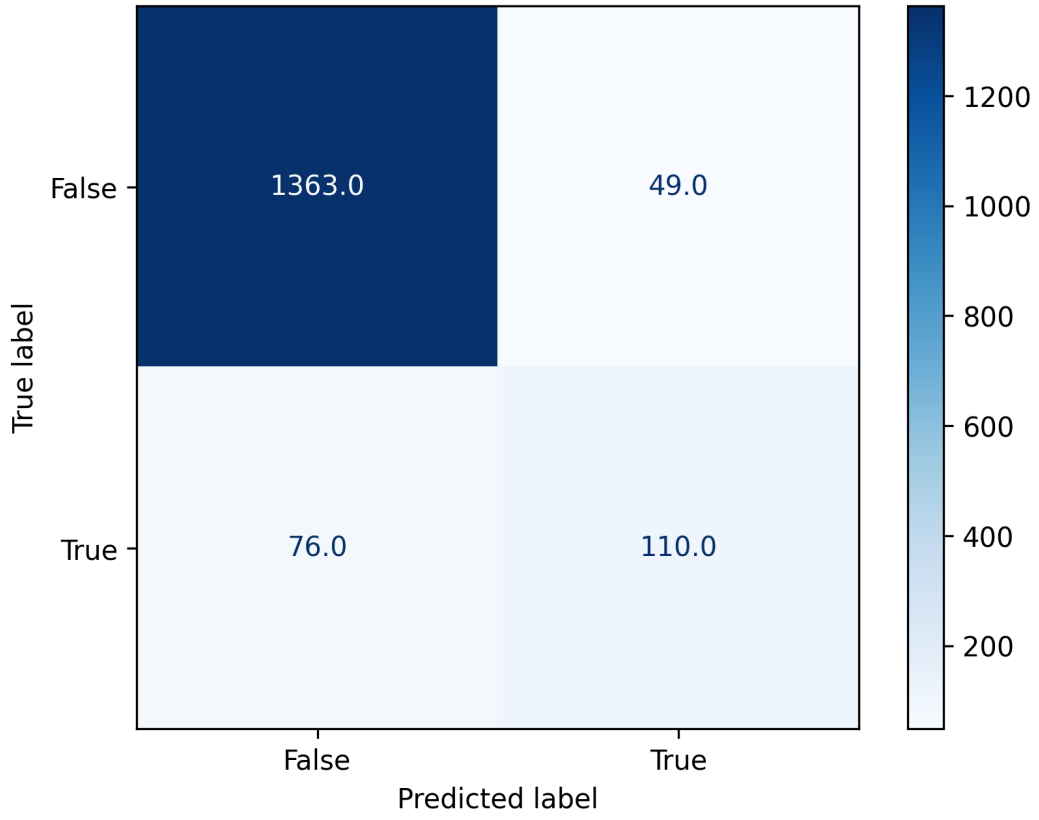
scorer	**training**	**validation**	**test**
AUC	0.9616	0.871	0.8574
ACCURACY	0.9597	0.9237	0.9141
F1	0.878	0.6748	0.6941

MCC	0.8547	0.6311	0.6479
LOGLOSS	0.1738	0.2625	0.2934

Validation Confusion Matrix

The confusion matrix shows how many observations the model correctly classified and misclassified. The first column contains the actual class labels; the first row contains the predicted class labels.

A positive prediction label (e.g., 1, True, or the second label in lexicographical order), is assigned to all observations where the predicted probability is greater than or equal to 0.1017 (the threshold for the highest F1 score on the validation dataset).

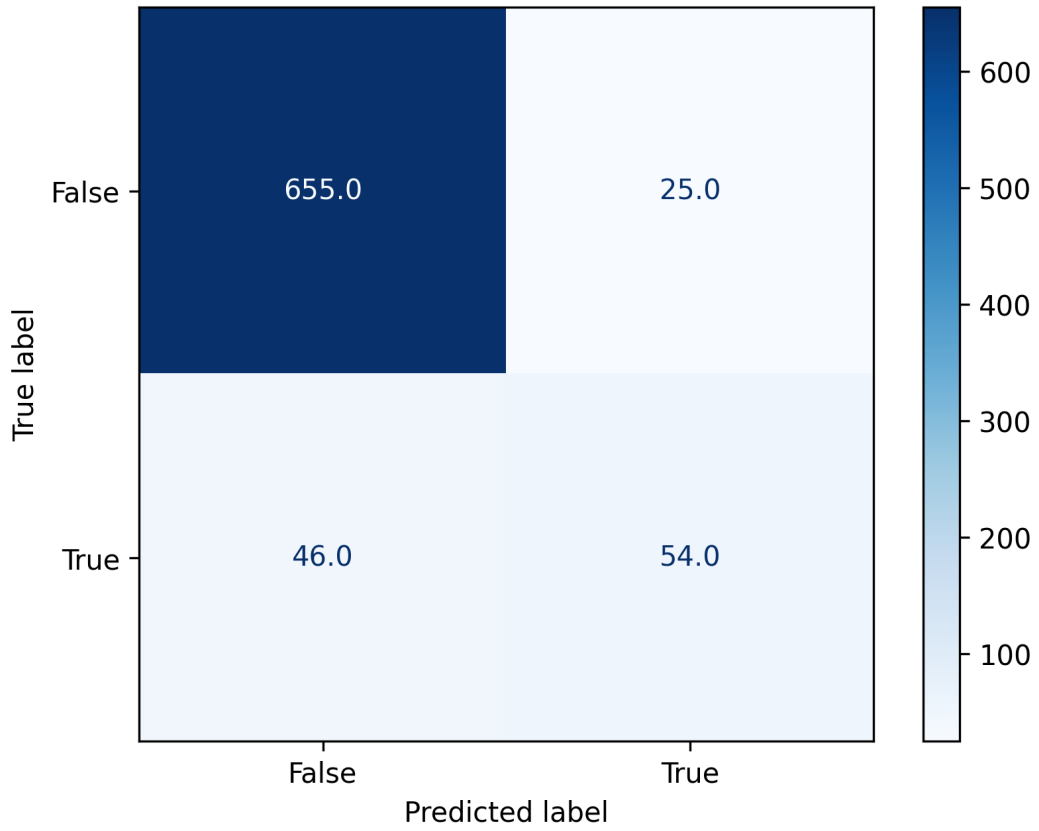


Labels	False	True	Error	Rate
False	1363.0	49.0	0.0347	(49.0/1412.0)
True	76.0	110.0	0.4086	(76.0/186.0)
Total	1439.0	159.0	0.0782	(125.0/1598.0)

Test Confusion Matrix

The confusion matrix shows how many observations the model correctly classified and misclassified. The first column contains the actual class labels; the first row contains the predicted class labels.

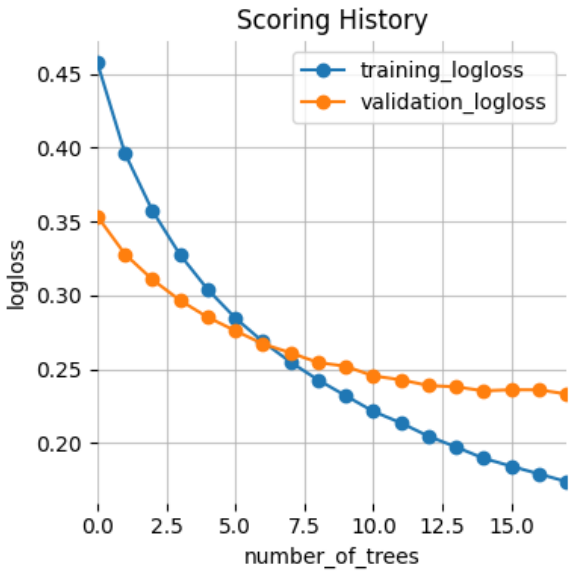
A positive prediction label (e.g., 1, True, or the second label in lexicographical order), is assigned to all observations where the predicted probability is greater than or equal to 0.1017 (the threshold for the highest F1 score on the validation dataset).



Labels	False	True	Error	Rate
False	655.0	25.0	0.0368	(25.0/680.0)
True	46.0	54.0	0.46	(46.0/100.0)
Total	701.0	79.0	0.091	(71.0/780.0)

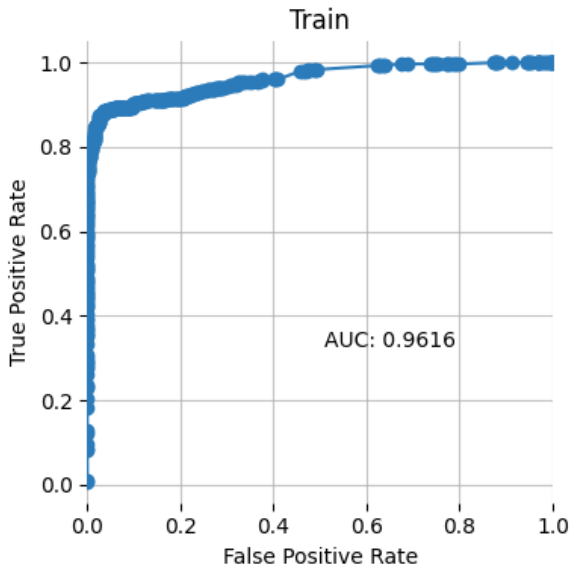
Scoring History

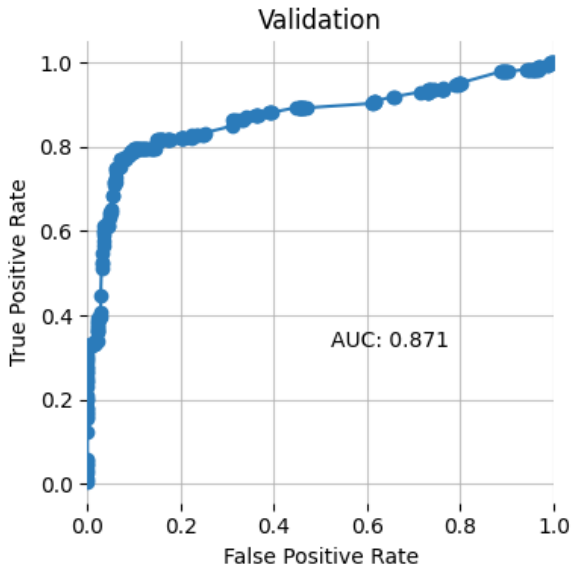
The scoring history plot shows a model's performance at each iteration point. Typically, the performance will be worse at the beginning (the left side of the graph) and then improve as the model training completes and accuracy improves.

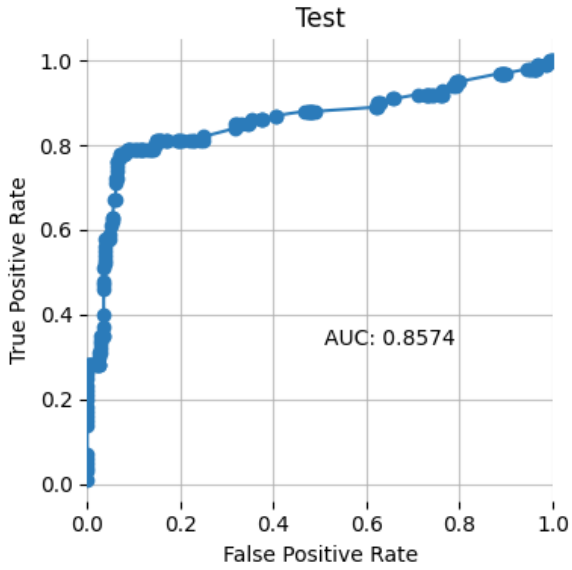


Receiver Operating Characteristic Curve

This plot shows the Receiver Operating Characteristic Curve. The area under this curve is called the AUC. The True Positive Rate (TPR) is the relative fraction of correct positive predictions, and the False Positive Rate (FPR) is the relative fraction of incorrect positive corrections. Each point corresponds to a classification threshold (e.g., YES if probability >= 0.3 else NO). For each threshold, there is a unique confusion matrix that represents the balance between TPR and FPR. In general, the most useful operating points are in the top left corner.

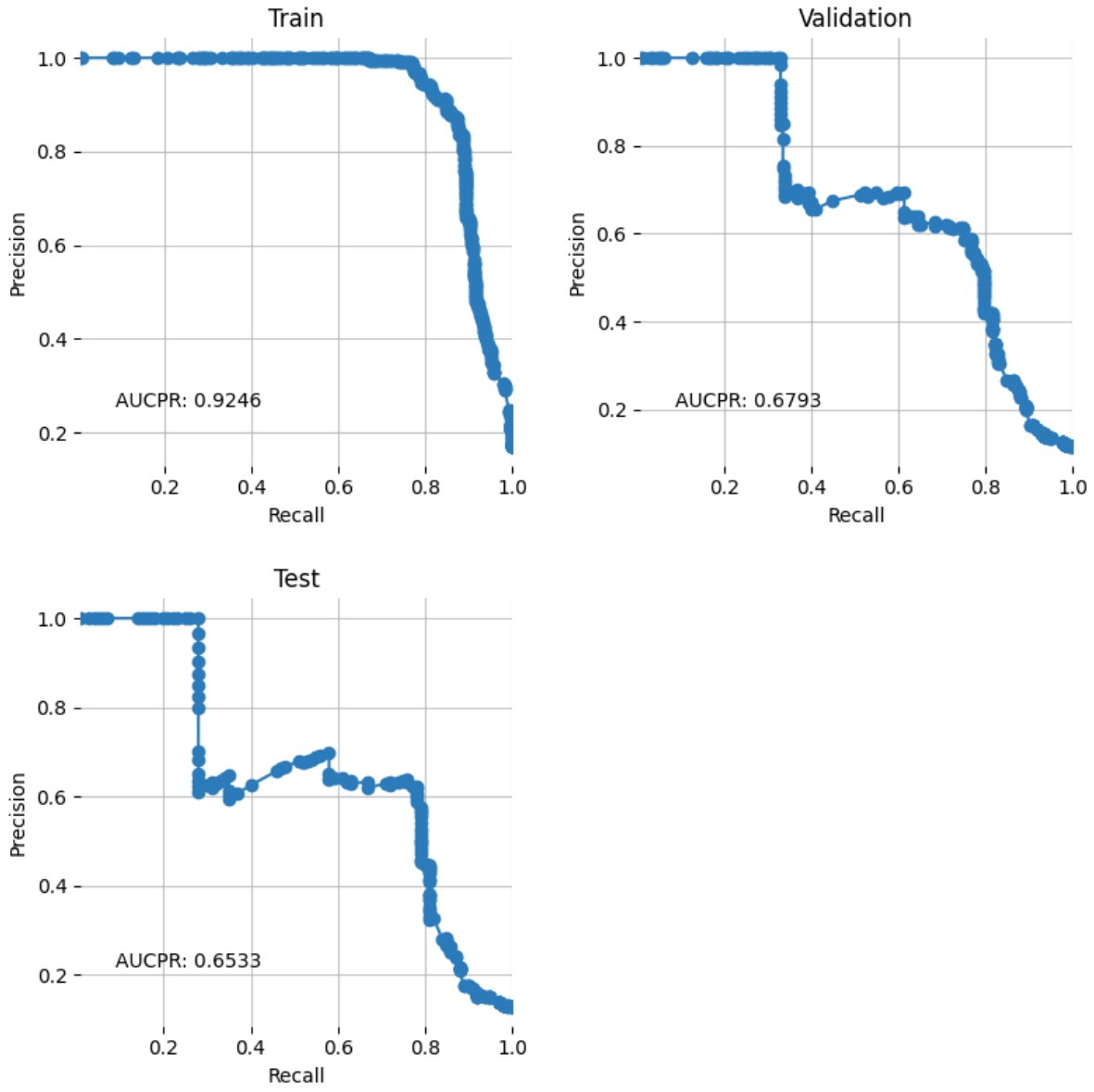






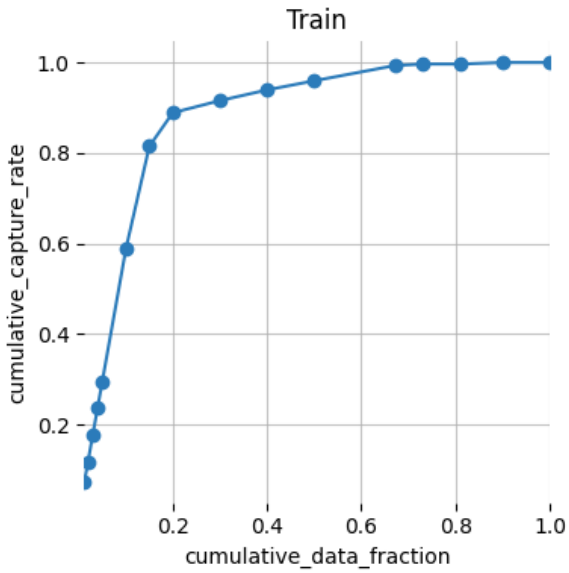
Precision-Recall Curve

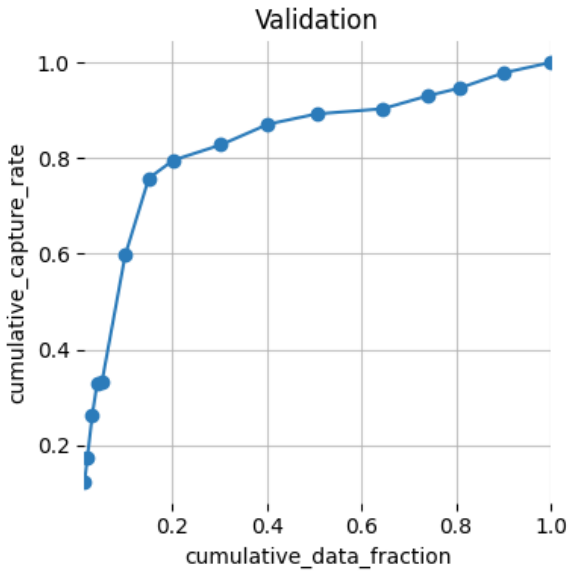
This model metric is used to evaluate how well a binary classification model is able to distinguish between precision recall pairs or points. These values are obtained using different thresholds on a probabilistic or other continuous-output classifier.

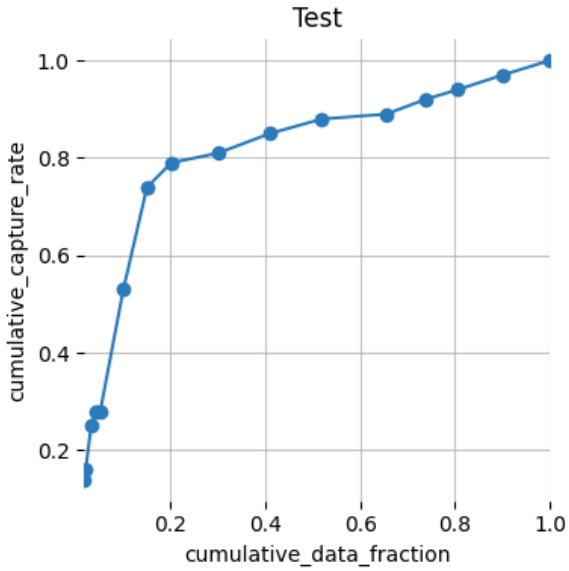


Cumulative Gain

This plot shows the cumulative gains. For example, "What fraction of all observations of the positive target class are in the top predicted 1%, 2%, 10%, etc. (cumulative)?" The gains at 100% are 1.0 by definition.

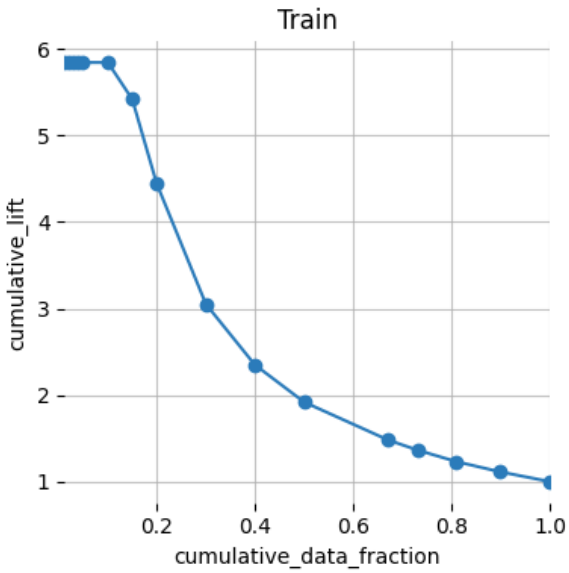


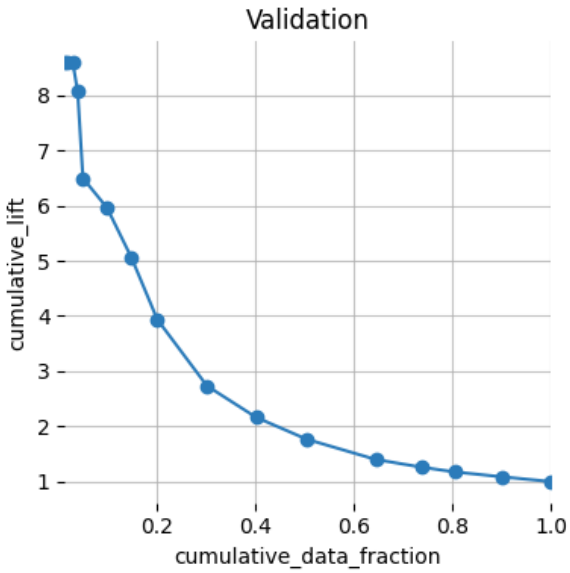


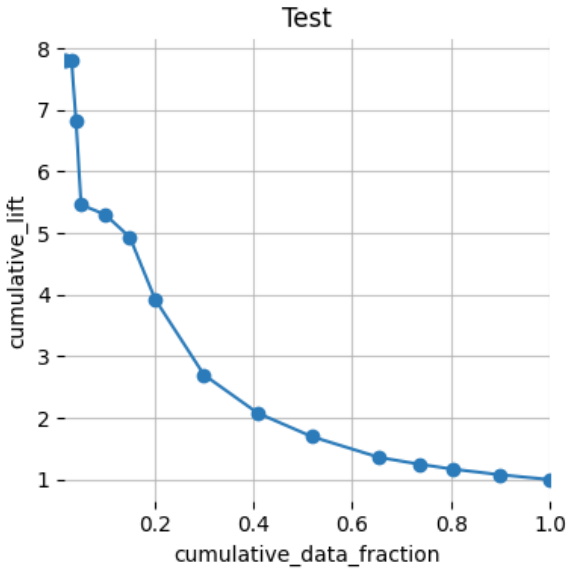


Cumulative Lift

This chart shows the cumulative lift. For example, "How many times more observations of the positive target class are in the top predicted 1%, 2%, 10%, etc. (cumulative) compared to selecting observations randomly?" By definition, the Lift at 100% is 1.0.







Population Stability Index (PSI)

Population Stability Index is a statistic used to describe a variable's distribution shift. It can measure the shift between the training dataset's model score distribution and any other given dataset (i.e. validation or test dataset).

A PSI value lower than 0.10 indicates a small shift in the model predictions, a value between 0.10 and 0.25 indicates a moderate shift, and a value greater than 0.25 indicates a strong shift. Strong shift values can indicate that the model trained on the training dataset might not be suitable for the provided validation or test datasets.

Summary PSI table

Dataset	PSI
Validation	0.2916
Test	0.2841

Details on the PSI calculations can be found in the Appendix.

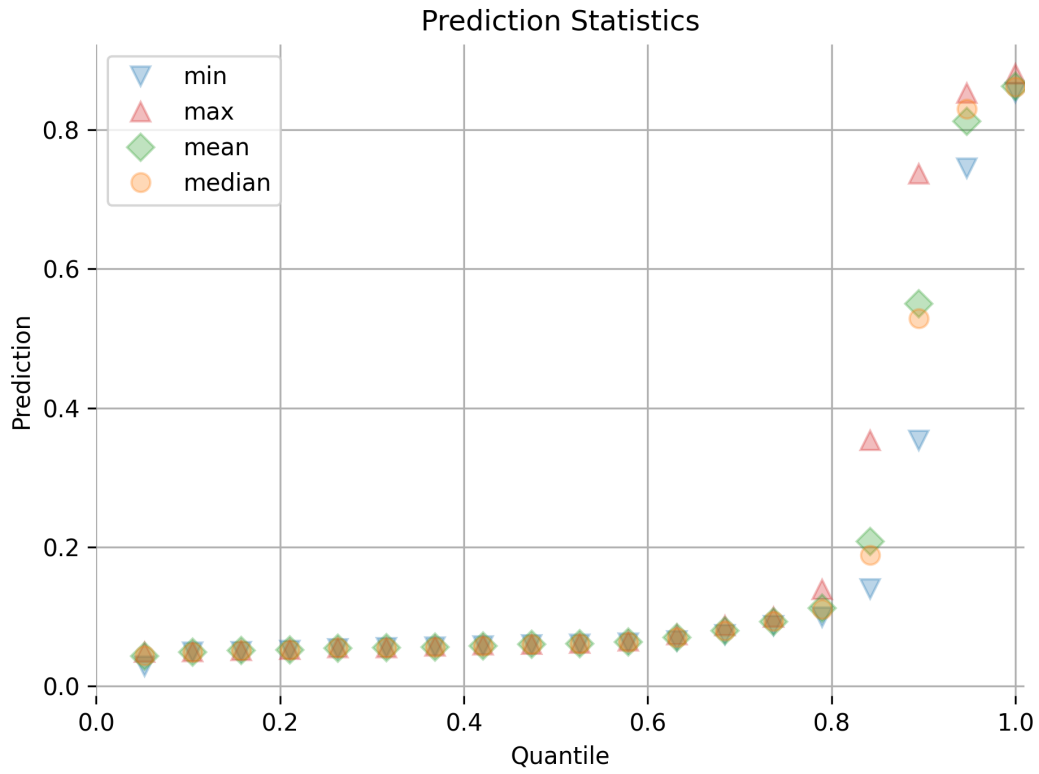
Prediction statistics

The following tables and plots show the min, max, mean, and median quantile prediction values for each dataset split. Note: values are rounded to the fourth decimal place. For example, .000025 and .000010 would both appear as 0.0.

Train

quantile	min	max	mean	median
0.0526	0.0288	0.0489	0.0435	0.044
0.1053	0.049	0.0495	0.0491	0.0492
0.1579	0.05	0.0513	0.0513	0.0513
0.2105	0.0515	0.0532	0.052	0.052
0.2632	0.0537	0.0552	0.0547	0.0552
0.3158	0.0552	0.0557	0.0555	0.0556
0.3684	0.0559	0.0575	0.0565	0.0559

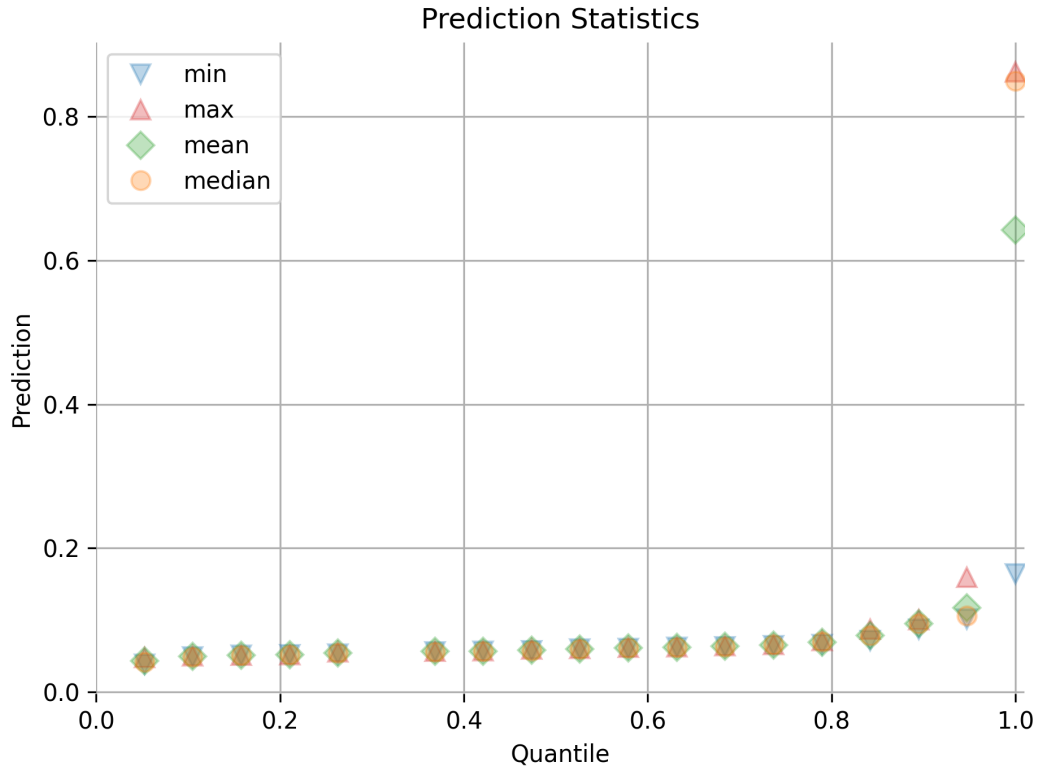
0.4211	0.0575	0.0594	0.0578	0.0575
0.4737	0.0594	0.0603	0.0599	0.0599
0.5263	0.0603	0.0622	0.0614	0.0614
0.5789	0.0622	0.065	0.0637	0.0635
0.6316	0.0651	0.0737	0.07	0.0704
0.6842	0.0737	0.0874	0.08	0.08
0.7368	0.0874	0.0995	0.0928	0.0938
0.7895	0.0995	0.1393	0.1122	0.1102
0.8421	0.1395	0.3529	0.2081	0.1889
0.8947	0.3532	0.7365	0.5501	0.5283
0.9474	0.745	0.8529	0.8124	0.8299
1.0	0.8535	0.8798	0.8622	0.8617



Validation

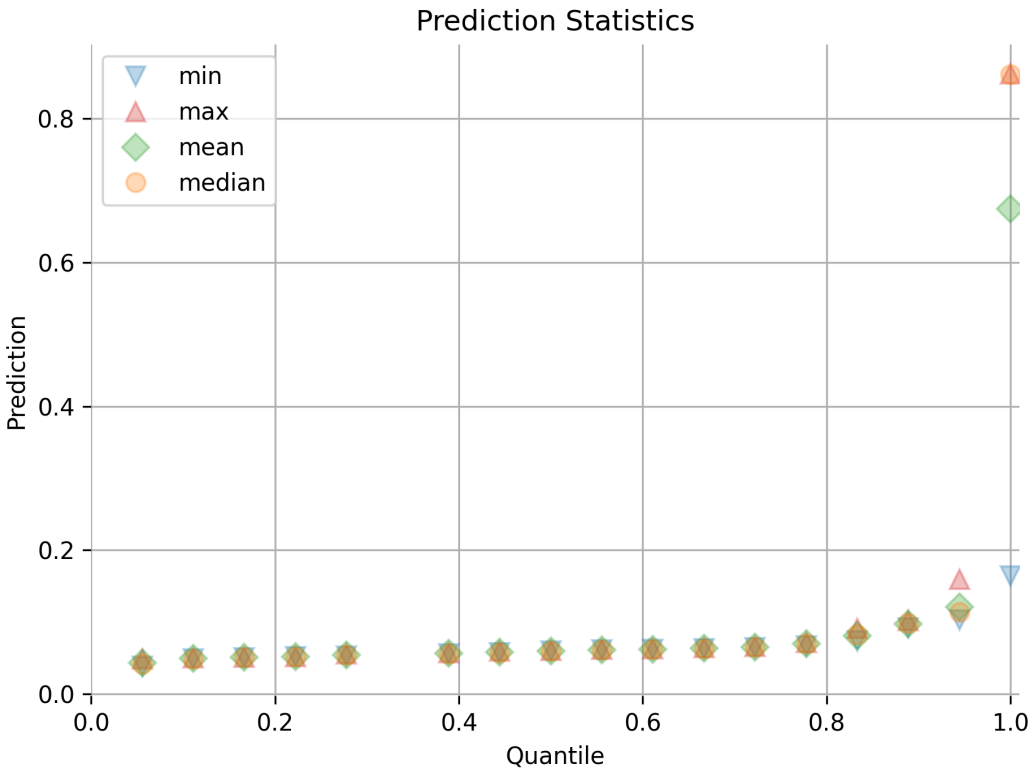
quantile	min	max	mean	median
0.0526	0.0386	0.0481	0.0432	0.0416
0.1053	0.0488	0.0505	0.0492	0.049
0.1579	0.0511	0.0513	0.0513	0.0513
0.2105	0.0515	0.0522	0.0521	0.0522
0.2632	0.0528	0.0556	0.0545	0.0543
0.3684	0.0559	0.0564	0.0561	0.0559
0.4211	0.0565	0.0571	0.0566	0.0565
0.4737	0.0575	0.0599	0.0578	0.0575

0.5263	0.0599	0.0601	0.0599	0.0599
0.5789	0.0604	0.0616	0.0612	0.0613
0.6316	0.0619	0.0625	0.0621	0.0619
0.6842	0.0632	0.0647	0.0634	0.0632
0.7368	0.0647	0.0659	0.0653	0.0658
0.7895	0.066	0.0713	0.069	0.0695
0.8421	0.0714	0.0868	0.0782	0.0782
0.8947	0.0868	0.1003	0.0948	0.0951
0.9474	0.1013	0.1593	0.1167	0.1062
1.0	0.1643	0.8618	0.6422	0.8495



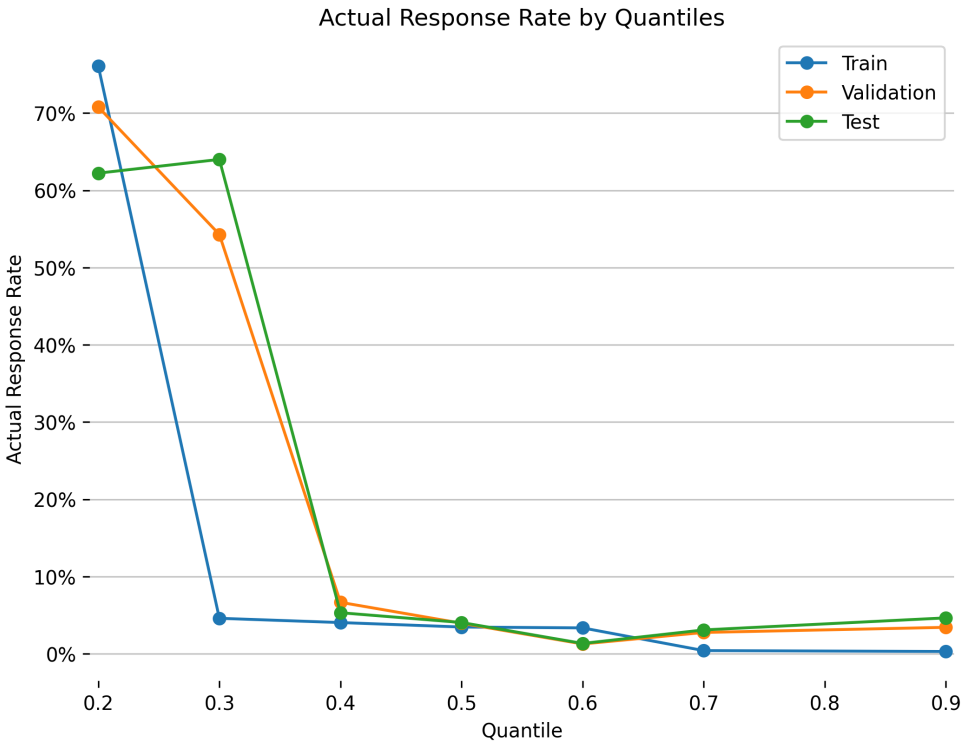
Test

quantile	min	max	mean	median
0.0556	0.0386	0.0488	0.0434	0.0416
0.1111	0.049	0.0503	0.0492	0.049
0.1667	0.0505	0.0513	0.0513	0.0513
0.2222	0.0515	0.0522	0.0521	0.0522
0.2778	0.0528	0.0556	0.0545	0.0543
0.3889	0.0559	0.0571	0.0561	0.0559
0.4444	0.0575	0.0599	0.0578	0.0575
0.5	0.0599	0.0601	0.0599	0.0599
0.5556	0.0609	0.0619	0.0612	0.0614
0.6111	0.0619	0.0625	0.062	0.0619
0.6667	0.0632	0.0641	0.0633	0.0632
0.7222	0.0647	0.0661	0.0653	0.0649
0.7778	0.0669	0.0715	0.0695	0.0696
0.8333	0.0736	0.091	0.0808	0.0821
0.8889	0.0917	0.1017	0.0972	0.098
0.9444	0.1025	0.1593	0.1213	0.1136
1.0	0.1643	0.8618	0.6748	0.8616



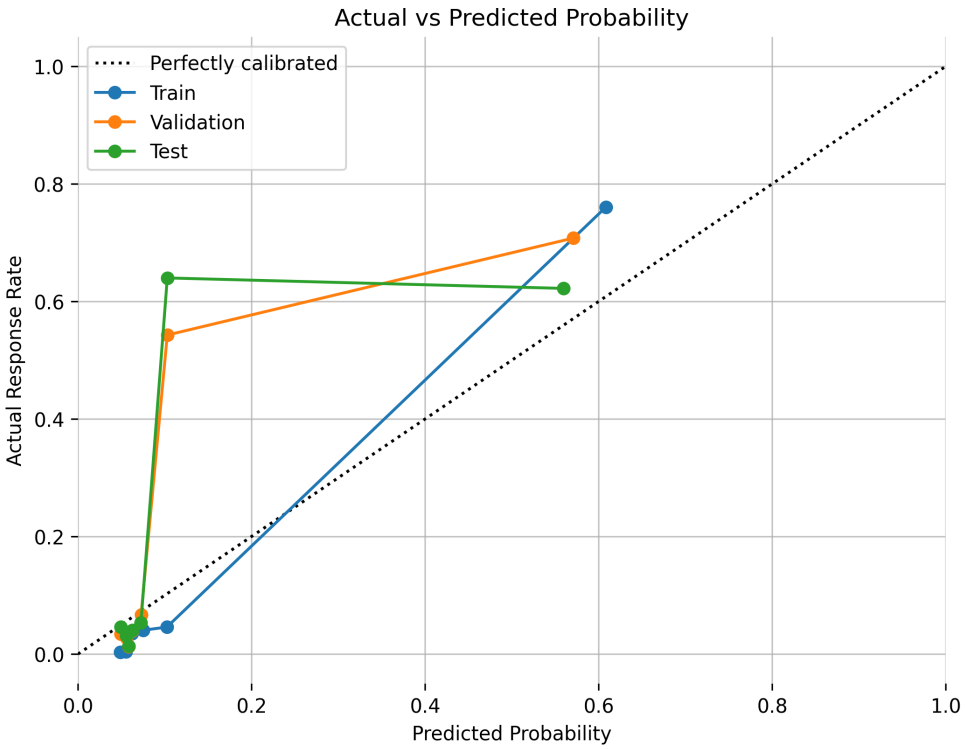
Quantile Response Rates

The response rate, for a given quantile, is equal to the number of positive-labeled data points divided by the total number of data points. Quantiles are sorted in decreasing order.



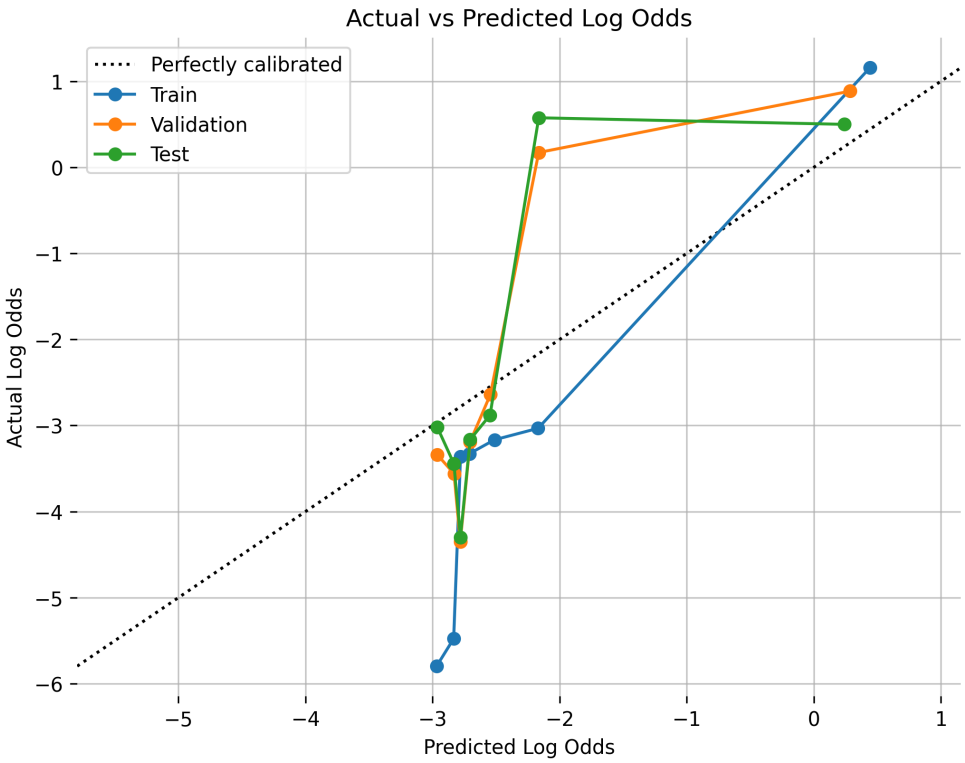
Actual vs Predicted Probabilities

This plot shows the alignment between the predicted and the actual probabilities. The predicted probabilities are binned into quantiles. For each, bin the average predicted value and the actual response rate (i.e., the number positive-labeled records divided by the total number of records within each bin) is calculated.



Actual vs Predicted Log Odds

This plot shows the alignment between the predicted and the actual probabilities within the log odds space. In this case, the log odds are the log transformation of the probability of a positive record divided by the probability of a negative record.



Details on the quantile-based plots' calculations can be found in the Appendix.

Alternative Models

The alternative model section consists of two part: alternative model performance and parameters. The performance section shows how each alternative model performed for a given dataset split (train, valid, test, etc.,). Performance tables are sorted by AUC. The parameter section provides the model arguments which a user has the control to grid over. Note: The parameter value "auto" corresponds to the default value for that model's H2O-3 version.

PERFORMANCE OVERVIEW TABLE

model_id	train_auc	valid_auc	test_auc
DRF_1_AutoML_20201110_095422	0.9263	0.8238	0.7987
XGBoost_3_AutoML_20201110_095422	0.9974	0.783	0.7668
XGBoost_1_AutoML_20201110_095422	0.9899	0.8068	0.7776

XGBoost_2_AutoML_20201110_095422	0.9566	0.6823	0.6609

PERFORMANCE TABLES

Train Data Split

model_id	F1	accuracy	logloss	mcc	auc
XGBoost_3_AutoML_20201110_095422	0.952	0.9838	0.0781	0.943	0.9974
XGBoost_1_AutoML_20201110_095422	0.9165	0.9722	0.1141	0.9012	0.9899
XGBoost_2_AutoML_20201110_095422	0.8429	0.9502	0.1811	0.819	0.9566
DRF_1_AutoML_20201110_095422	0.8141	0.9392	0.4646	0.779	0.9263

Valid Data Split

model_id	F1	accuracy	logloss	mcc	auc
DRF_1_AutoML_20201110_095422	0.5123	0.9174	0.3405	0.513	0.8238
XGBoost_1_AutoML_20201110_095422	0.4926	0.908	0.293	0.4386	0.8068
XGBoost_3_AutoML_20201110_095422	0.5342	0.9193	0.3016	0.5284	0.783
XGBoost_2_AutoML_20201110_095422	0.3946	0.8986	0.3409	0.3572	0.6823

Test Data Split

model_id	F1	accuracy	logloss	mcc	auc
DRF_1_AutoML_20201110_095422	0.4672	0.9051	0.3744	0.4842	0.7987

XGBoost_1_AutoML_20201110_095422	0.4639	0.8987	0.3327	0.4319	0.7776
XGBoost_3_AutoML_20201110_095422	0.4795	0.909	0.3433	0.51	0.7668
XGBoost_2_AutoML_20201110_095422	0.3766	0.891	0.376	0.3639	0.6609

PARAMETER TUNING TABLES

Distributed Random Forest Tuning

Distributed Random Forest (DRF) is a powerful classification and regression tool. When given a set of data, DRF generates a forest of classification or regression trees, rather than a single classification or regression tree. Each of these trees is a weak learner built on a subset of rows and columns. More trees will reduce the variance.

Parameters	**DISTRIBUTED RANDOM FOREST_1**
model_id	DRF_1_AutoML_20201110_095422
balance_classes	False
categorical_encoding	Enum
class_sampling_factors	None
col_sample_rate_change_per_level	1.0
col_sample_rate_per_tree	1.0
distribution	bernoulli
fold_assignment	None
fold_column	None
histogram_type	UniformAdaptive

max_after_balance_size	5.0
max_depth	20
max_runtime_secs	768614359311056.9
min_rows	1.0
min_split_improvement	1e-05
mtries	-1
nbins	20
nbins_cats	1024
nbins_top_level	1024
ntrees	50
offset_column	None
r2_stopping	1.7976931348623157e+308
response_column	Churn
sample_rate	0.632
sample_rate_per_class	None
seed	4
stopping_metric	logloss
stopping_rounds	3
stopping_tolerance	0.02400768368836718

validation_frame	py_11_sid_8916
weights_column	None

Xgboost Tuning

XGBoost is a supervised learning algorithm that implements a process called boosting to yield accurate models. Boosting refers to the ensemble learning technique of building many models sequentially, with each new model attempting to correct for the deficiencies in the previous model.

Parameters	**XGBOOST_1**	**XGBOOST_2**	**XGBOOST_3**
model_id	XGBoost_3_AutoML_20201110_095422	XGBoost_1_AutoML_20201110_095422	XGBoost_2_AutoML_20201110_095422
backend	cpu	cpu	cpu
booster	gbtree	gbtree	gbtree
categorical_encoding	OneHotInternal	OneHotInternal	OneHotInternal
col_sample_rate	0.8	0.8	0.8
col_sample_rate_per_tree	0.8	0.8	0.8
colsample_bylevel	0.8	0.8	0.8
colsample_bynode	1.0	1.0	1.0
colsample_bytree	0.8	0.8	0.8

distribution	bernoulli	bernoulli	bernoulli
dmatrix_type	dense	dense	dense
eta	0.3	0.3	0.3
fold_assignment	None	None	None
fold_column	None	None	None
gamma	0.0	0.0	0.0
grow_policy	depthwise	depthwise	depthwise
learn_rate	0.3	0.3	0.3
max_abs_leafnode_pred	0.0	0.0	0.0
max_bins	256	256	256
max_delta_step	0.0	0.0	0.0
max_depth	5	10	20
max_leaves	0	0	0
max_runtime_secs	709490183111704.6	614891501192740.9	658812317797974.0
min_child_weight	3.0	5.0	10.0
min_rows	3.0	5.0	10.0
min_split_improvement	0.0	0.0	0.0

normalize_type	tree	tree	tree
ntrees	10000	10000	10000
offset_column	None	None	None
one_drop	False	False	False
rate_drop	0.0	0.0	0.0
reg_alpha	0.0	0.0	0.0
reg_lambda	1.0	1.0	1.0
response_column	Churn	Churn	Churn
sample_rate	0.8	0.6	0.6
sample_type	uniform	uniform	uniform
seed	3	1	2
skip_drop	0.0	0.0	0.0
stopping_metric	logloss	logloss	logloss
stopping_rounds	3	3	3
stopping_tolerance	0.0240076836883671 8	0.0240076836883671 8	0.0240076836883671 8
subsample	0.8	0.6	0.6
tree_method	exact	exact	exact

tweedie_power	1.5	1.5	1.5
validation_frame	py_11_sid_8916	py_11_sid_8916	py_11_sid_8916
weights_column	None	None	None

Partial Dependence Plots

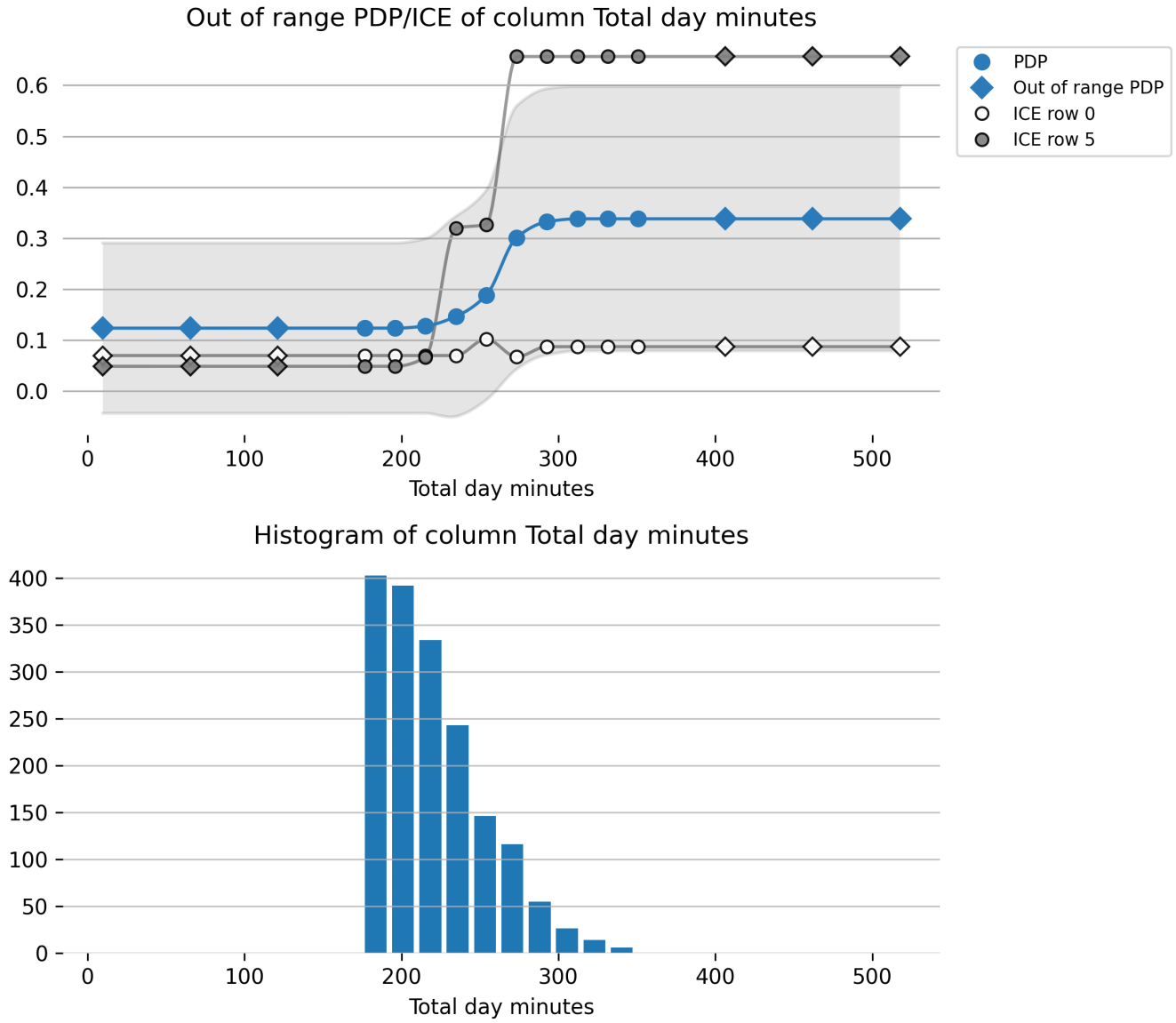
Partial dependence plots show the partial dependence as a function of specific values for a feature subset. The plots show how machine-learned response functions change based on the values of an input feature of interest, while taking nonlinearity into consideration and averaging out the effects of all 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 4 user-selected features.

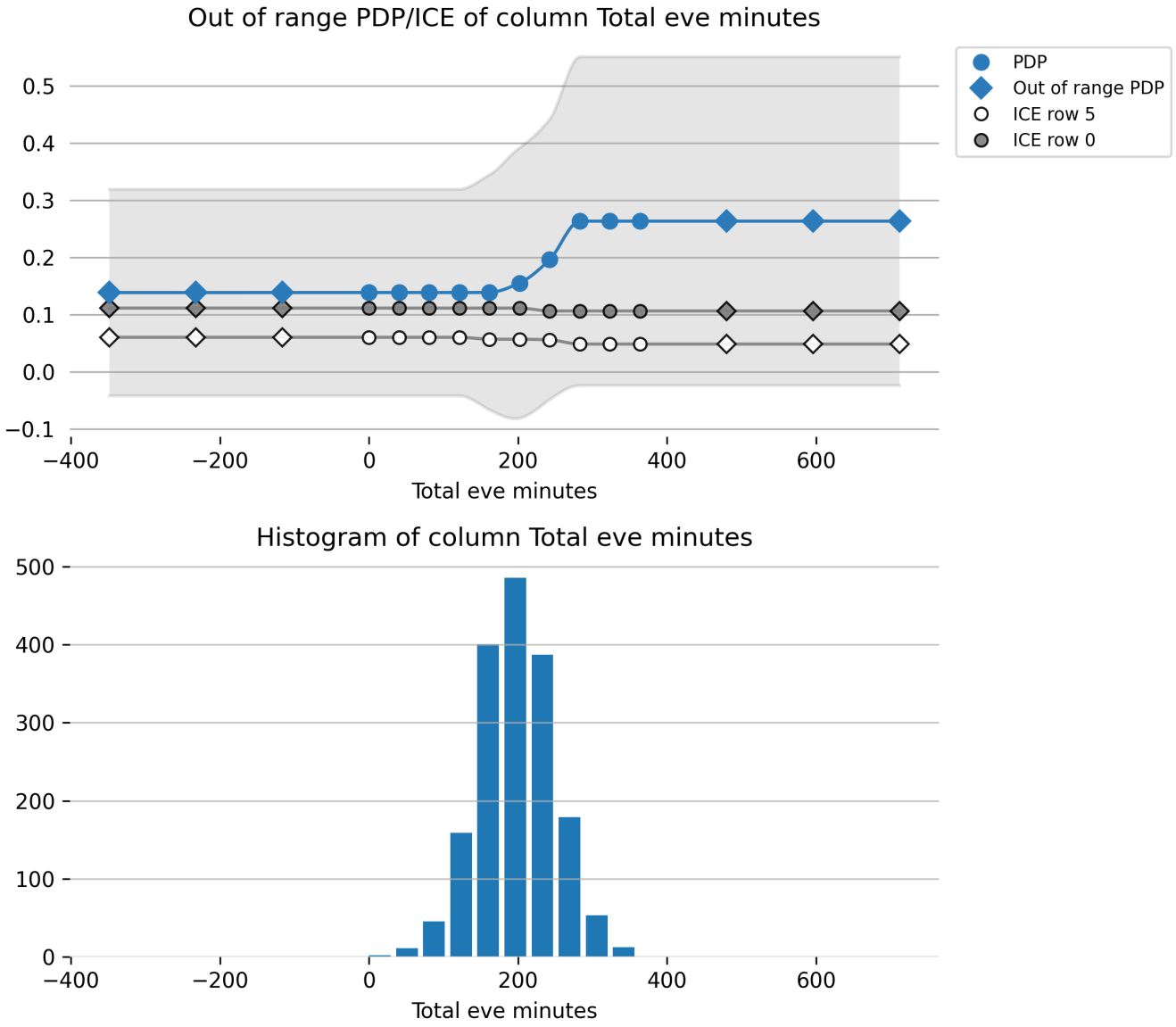
Plot Details

In the H2O-3 PDP, the y-axis represents the mean response, and a shaded region (for numeric features) or shaded bar (for categorical features) represents $\pm$ 1 standard deviation.

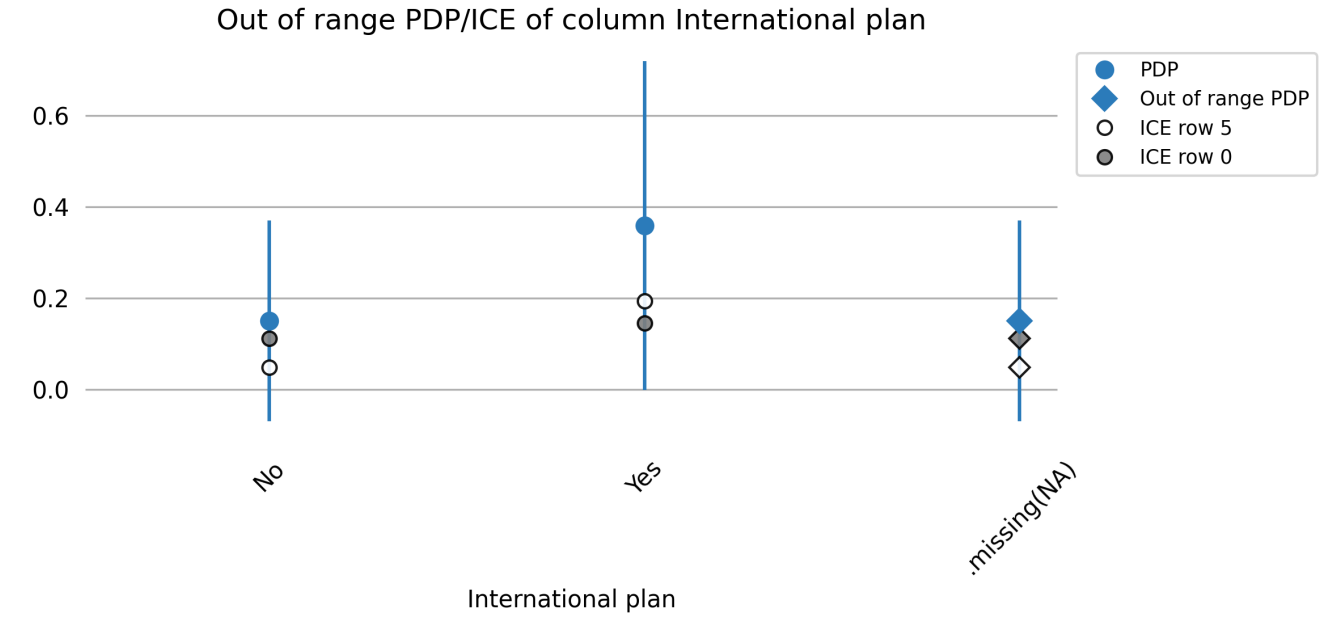
Feature **Total day minutes**

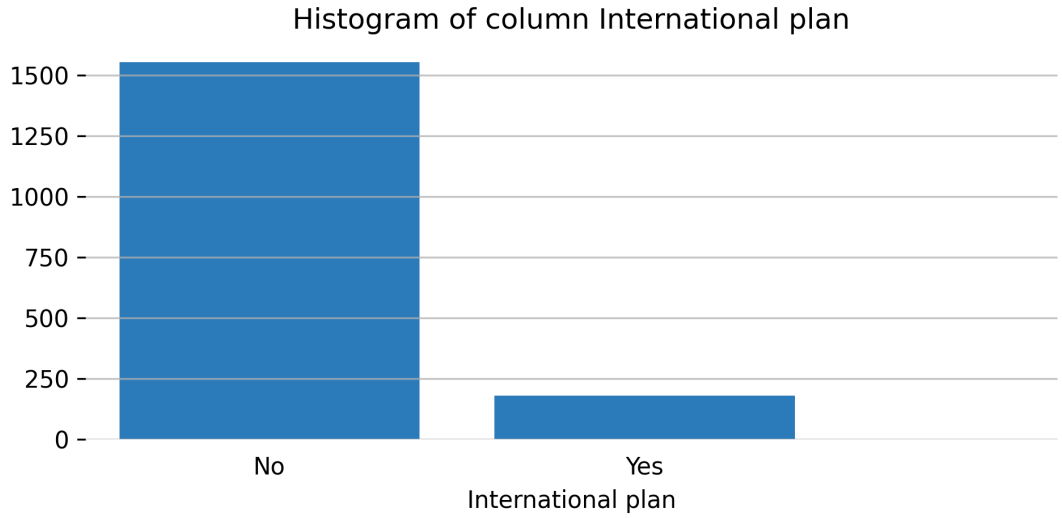


Feature **Total eve minutes**

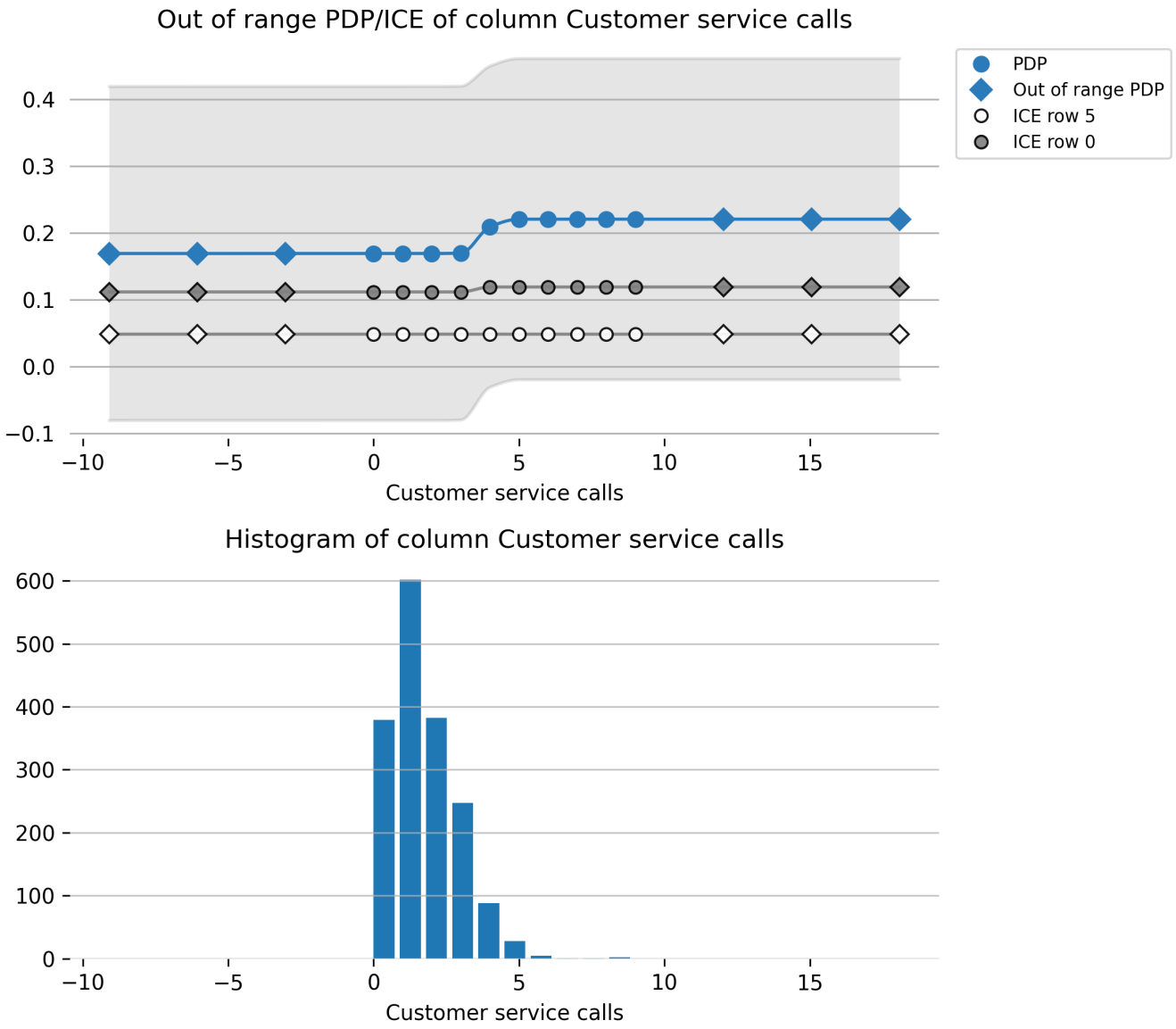


Feature **International plan**





Feature **Customer service calls**



Model Reproducibility

This model is not reproducible, because early stopping was enabled without setting the **score_tree_interval** parameter.

Early Stopping Parameters

Model Parameter	Value
score_tree_interval	0
stopping_rounds	3
stopping_metric	logloss
stopping_tolerance	0.01

Single-Node Cluster Reproducibility Requirements

The following criteria must be met to guarantee reproducibility in a single node cluster:

- Same training data. **Note**: If you have H2O-3 import a whole directory with multiple files instead of a single file, H2O-3 does not guarantee reproducibility because the data may be shuffled during import.
- Same parameters are used to train the model.
- Same seed is used if sampling is done. The following parameters perform sampling:
    - sample_rate
    - sample_rate_per_class
    - col_sample_rate
    - col_sample_rate_change_per_level
    - col_sample_rate_per_tree
- No early stopping performed or early stopping with score_tree_interval is explicitly specified and the same validation data.

Multi-Node Cluster Reproducibility Requirements

The following criteria must be met to guarantee reproducibility in a multi-node cluster:

- Reproducible requirements for single node cluster are met. (See above.)
- The cluster configuration is the same:
    - Clusters must have the same number of nodes.
    - Nodes must have the same number of CPU cores available (or same restriction on number of threads).

If you do not have a machine with the same number of CPU cores, see the question "How can I reproduce a model if machines have a different number of CPU cores?" below.

- The model training is triggered from the leader node of the cluster. (See note below.)

Note: When H2O is running on Hadoop, the leader node is automatically returned by the h2odriver as the node that the user should connect to. In multi-node deployments of Standalone H2O, the leader node must be manually identified by the user. Flow users can easily check whether they are connected to the leader node by opening Cluster Status (from the Admin menu) and checking that the first node has the same IP address as they see in their browser's address bar.

Appendix

Final Model Details

Model Parameters (Complete List)	**Values**
model_id	gbm
balance_classes	False
build_tree_one_node	False
calibrate_model	False
calibration_frame	None
categorical_encoding	Enum
check_constant_response	True
checkpoint	None
class_sampling_factors	None
col_sample_rate	1.0
col_sample_rate_change_per_level	1.0
col_sample_rate_per_tree	1.0
custom_distribution_func	None

custom_metric_func	None
distribution	bernoulli
export_checkpoints_dir	None
fold_assignment	None
fold_column	None
gainslift_bins	-1
histogram_type	UniformAdaptive
huber_alpha	0.9
ignore_const_cols	True
ignored_columns	None
keep_cross_validation_fold_assignment	False
keep_cross_validation_models	True
keep_cross_validation_predictions	False
learn_rate	0.1
learn_rate_annealing	1.0
max_abs_leafnode_pred	1.7976931348623157e+308
max_after_balance_size	5.0
max_confusion_matrix_size	20
max_depth	5

max_runtime_secs	0.0
min_rows	10.0
min_split_improvement	1e-05
monotone_constraints	None
nbins	20
nbins_cats	1024
nbins_top_level	1024
nfolds	0
ntrees	50
offset_column	None
pred_noise_bandwidth	0.0
quantile_alpha	0.5
r2_stopping	1.7976931348623157e+308
response_column	Churn
sample_rate	1.0
sample_rate_per_class	None
score_each_iteration	False
score_tree_interval	0
seed	1234

stopping_metric	logloss
stopping_rounds	3
stopping_tolerance	0.01
training_frame	py_8_sid_8916
tweedie_power	1.5
validation_frame	py_11_sid_8916
weights_column	None

Population Stability Index (PSI) Final Model Details

The PSI and calculation table is provided for each dataset below. The corresponding table columns are defined as follows:

- *Quantile: the bin to which the ordered predicted probabilities belong.*
- *Upper Bound: the upper bound of the corresponding bin.*
- *Test Count: the total number of Test records within the corresponding bin.*
- *Test Fraction (Tst): Test Count divided by the total number of Test records.*
- *Train Count: the total number of Train records within the corresponding bin.*
- *Train Fraction (Trn): Train Count divided by the total number of Train records.*
- *Tst - Trn: the difference between the Test Fraction and the Train Fraction.*
- *ln(Tst / Trn): the natural logarithm of the Test Fraction divided by the Train Fraction.*
- *PSI: the Population Stability Index for each bin - the dataset PSI is the total sum of these PSI values.*

Validation

The Population Stability Index is 0.2916.

Quantile	Upper Bound	Test Count	Test Fraction	Train Count	Train Fraction	Tst - Trn	ln(Tst / Trn)	PSI

			(Tst)		(Trn)			
0.1	0.0497	149	0.0932	174	0.1003	-0.007	-0.0729	0.0005
0.2	0.0537	201	0.1258	156	0.0899	0.0359	0.3357	0.012
0.3	0.0559	67	0.0419	136	0.0784	-0.0365	-0.6257	0.0228
0.4	0.0575	150	0.0939	104	0.0599	0.0339	0.4485	0.0152
0.5	0.0603	314	0.1965	298	0.1718	0.0247	0.1346	0.0033
0.6	0.065	278	0.174	173	0.0997	0.0743	0.5566	0.0413
0.7	0.0874	210	0.1314	173	0.0997	0.0317	0.2761	0.0088
0.8	0.1393	140	0.0876	174	0.1003	-0.0127	-0.1352	0.0017
0.9	0.7365	40	0.025	174	0.1003	-0.0753	-1.3879	0.1045
1.0	inf	49	0.0307	173	0.0997	-0.069	-1.1792	0.0814

Test

The Population Stability Index is 0.2841.

Quantile	Upper Bound	Test Count	Test Fraction (Tst)	Train Count	Train Fraction (Trn)	Tst - Trn	ln(Tst / Trn)	PSI
0.1	0.0497	74	0.0949	174	0.1003	-0.0054	-0.0555	0.0003

0.2	0.0537	98	0.1256	156	0.0899	0.0357	0.3346	0.012
0.3	0.0559	33	0.0423	136	0.0784	-0.0361	-0.6167	0.0222
0.4	0.0575	64	0.0821	104	0.0599	0.0221	0.314	0.0069
0.5	0.0603	149	0.191	298	0.1718	0.0193	0.1063	0.002
0.6	0.065	148	0.1897	173	0.0997	0.09	0.6434	0.0579
0.7	0.0874	94	0.1205	173	0.0997	0.0208	0.1895	0.0039
0.8	0.1393	75	0.0962	174	0.1003	-0.0041	-0.0421	0.0002
0.9	0.7365	20	0.0256	174	0.1003	-0.0746	-1.3639	0.1018
1.0	inf	25	0.0321	173	0.0997	-0.0677	-1.1349	0.0768

Quantile Plots Calculation Table

The following table is used to calculate the Quantile Response Rates, Actual vs Predicted Probabilities, and Actual vs Predicted Log Odds plot.

table columns are defined as follows:

- *Quantile: the bin to which the ordered predicted probabilities belong.*
- *bound: the upper bound of the corresponding bin.*
- *{dataset name} cnt: the number of records within the corresponding bin.*
- *{dataset name} sum: the number of positive-labeled records within the corresponding bin.*
- *{dataset name} act: the fraction of positive-labeled records within the corresponding bin.*
- *{dataset name} pred: the mean of the predicted values that fall within the corresponding bin.*

Quantile	bound	Train cnt	Train sum	Train act	Train pred	Validation cnt	Validation sum	Validation act	Validation pred	Test cnt	Test sum	Test act	Test pred
0.2	inf	347	264	0.7608	0.6086	89	63	0.7079	0.571	45	28	0.6222	0.5593
0.3	0.1393	174	8	0.0046	0.1025	140	76	0.5429	0.103	75	48	0.64	0.103
0.4	0.0874	173	7	0.0405	0.075	210	14	0.0667	0.0728	94	5	0.0532	0.0726
0.5	0.065	173	6	0.0347	0.0625	278	11	0.0396	0.0627	148	6	0.0405	0.0627
0.6	0.0603	298	10	0.0336	0.0584	314	4	0.0127	0.0583	149	2	0.0134	0.0584
0.7	0.0575	240	1	0.0042	0.0555	217	6	0.0276	0.0557	97	3	0.0309	0.0556
0.9	0.0537	330	1	0.003	0.0489	350	12	0.0343	0.0491	172	8	0.0465	0.0492