

Employee Attrition

?

 Assistance

Routine	Description
importFiles	Import file(s) into H ₂ O
importSqlTable	Import SQL table into H ₂ O
getFrames	Get a list of frames in H ₂ O
splitFrame	Split a frame into two or more frames
mergeFrames	Merge two frames into one
getModels	Get a list of models in H ₂ O
getGrids	Get a list of grid search results in H ₂ O
getPredictions	Get a list of predictions in H ₂ O
getJobs	Get a list of jobs running in H ₂ O
runAutoML	Automatically train and tune many models
buildModel	Build a model
importModel	Import a saved model
predict	Make a prediction

?

 Import Files

Search: C:\h2o



Search Results: Found 3 files: Add all

- + C:\h2o\allyears2k.csv
- + C:\h2o\WA_Fn-UseC_-HR-Employee-Attrition.xlsx

Selected Files: 1 file selected: Clear All

- x C:\h2o\WA_Fn-UseC_-HR-Employee-Attrition.csv

Actions: Import

Cloud 1 / 1 files imported.

Files C:\h2o\WA_Fn-UseC_-HR-Employee-Attrition.csv

Actions Parse these files...

⚙ Setup Parse

PARSE CONFIGURATION

Sources nfs:\C:\h2o\WA_Fn-UseC_-HR-Employee-Attrition.csv

ID WA_Fn_UseC_HR_Employee_Attrition.hex

Parser CSV

Separator ,: '044'

Column Headers Auto First row contains column names First row contains dataOptions Enable single quotes as a field quotation character Delete on done

EDIT COLUMN NAMES AND TYPES

Search by column name...

1	Age	Numeric ▾	41	49	37	33	27	32
2	Attrition	Enum ▾	Yes	No	Yes	No	No	No
3	BusinessTravel	Enum ▾	Travel_Rarely	Travel_Frequently	Travel_Rarely	Travel_Frequently	Travel_Rarely	Travel_Fre
4	DailyRate	Numeric ▾	1102	279	1373	1392	591	1005
5	Department	Enum ▾	Sales	Research & Development				

6	DistanceFromHome	Enum ▼	1	8	2	3	2	2
7	Education	Enum ▼	2	1	2	4	1	2
8	EducationField	Enum ▼	Life Sciences	Life Sciences	Other	Life Sciences	Medical	Life Sciences
9	EmployeeCount	Numeric ▼	1	1	1	1	1	1
10	EmployeeNumber	Numeric ▼	1	2	4	5	7	8
11	Environment	Enum ▼	2	3	4	4	1	4
12	Gender	Enum ▼	Female	Male	Male	Female	Male	Male
13	HourlyRate	Numeric ▼	94	61	92	56	40	79
14	JobInvolvement	Enum ▼	3	2	2	3	3	3
15	JobLevel	Enum ▼	2	2	1	1	1	1

[◀ Previous page](#) [▶ Next page](#)

[Parse](#)

Job

Run Time 00:00:00.261

Remaining Time 00:00:00.0

Type Frame

Key [WA_Fn_UseC_HR_Employee_Attrition.hex](#)

Description Parse

Status DONE

Progress 100%

Done.

Actions [View](#)

WA_Fn_UseC_HR_Employee_Attrition.hex

Actions: [View Data](#) [Split...](#) [Build Model...](#) [Predict](#) [Download](#) [Export](#) [Delete](#)

Rows	Columns	Compressed Size
1470	35	218KB

COLUMN SUMMARIES

label	type	Missing	Zeros	+Inf	-Inf	min	max	mean	sigma	cardinality	Actions
NumCompaniesWorked	int	0	197	0	0	0	9.0	2.6932	2.4980	· Convert to enum	
Over18	enum	0	1470	0	0	0	0	0	0	1 Convert to numeric	
Overtime	enum	0	1054	0	0	0	1.0	0.2830	0.4506	2 Convert to numeric	
PercentSalaryHike	int	0	0	0	0	11.0	25.0	15.2095	3.6599	· Convert to enum	
PerformanceRating	enum	0	1244	0	0	0	1.0	0.1537	0.3608	2 Convert to numeric	
RelationshipSatisfaction	enum	0	276	0	0	0	3.0	·	·	4 Convert to numeric	
StandardHours	int	0	0	0	0	80.0	80.0	80.0	0	· Convert to enum	
StockOptionLevel	enum	0	631	0	0	0	3.0	·	·	4 Convert to numeric	
TotalWorkingYears	int	0	11	0	0	0	40.0	11.2796	7.7808	· Convert to enum	
TrainingTimesLastYear	int	0	54	0	0	0	6.0	2.7993	1.2893	· Convert to enum	
WorkLifeBalance	enum	0	80	0	0	0	3.0	·	·	4 Convert to numeric	
YearsAtCompany	int	0	44	0	0	0	40.0	7.0082	6.1265	· Convert to enum	
YearsInCurrentRole	int	0	244	0	0	0	18.0	4.2293	3.6231	· Convert to enum	
YearsSinceLastPromotion	int	0	581	0	0	0	15.0	2.1878	3.2224	· Convert to enum	
YearsWithCurrManager	int	0	263	0	0	0	17.0	4.1231	3.5681	· Convert to enum	

[◀ Previous 20 Columns](#)

[▶ Next 20 Columns](#)

CHUNK COMPRESSION SUMMARY

FRAME DISTRIBUTION SUMMARY

SplitOptions

Frame: [WA_Fn_UseC_HR_Employee_Attrition.hex](#)

Splits:	Ratio	Key
0.70		WA_Fn_UseC_HR_Employee_Attrition.hex_0.70_ <input checked="" type="checkbox"/>
0.15		WA_Fn_UseC_HR_Employee_Attrition.hex_0.15_ <input checked="" type="checkbox"/>
0.150		WA_Fn_UseC_HR_Employee_Attrition.hex_0.15_
Add a new split		
Seed:	1234	

[Create](#)

Split Frames

Type	Key	Ratio
#	WA_Fn_UseC_HR_Employee_Attrition.hex_0.70_train	0.7
#	WA_Fn_UseC_HR_Employee_Attrition.hex_0.15_validation	0.15
#	WA_Fn_UseC_HR_Employee_Attrition.hex_0.15_test	0.15000000000000002

Build a Model

Select an algorithm: Gradient Boosting Machine ▾

PARAMETERS

model_id gbm-508038d8-418e-4c20-b39b-ecbafde01b37

training_frame WA_Fn_UseC_HR_Employee_Attrition.hex_0.70_train

validation_frame WA_Fn_UseC_HR_Employee_Attrition.hex_0.15_validation

nfold 5

response_column Attrition

ignored_columns Search...

Showing page 1 of 1. 5 ignored.

<input type="checkbox"/> Age	INT
<input type="checkbox"/> Attrition	ENUM(2)
<input type="checkbox"/> BusinessTravel	ENUM(3)
<input type="checkbox"/> DailyRate	INT
<input type="checkbox"/> Department	ENUM(3)
<input type="checkbox"/> DistanceFromHome	INT
<input type="checkbox"/> Education	ENUM(5)
<input type="checkbox"/> EducationField	ENUM(6)
<input checked="" type="checkbox"/> EmployeeCount	INT
<input checked="" type="checkbox"/> EmployeeNumber	INT

All None

Only show columns with more than 0 % missing values.

ignore_const_cols

ntrees 50

max_depth 5

min_rows 10

nbins 20

seed -1

learn_rate 0.1

sample_rate 1

col_sample_rate 1

ADVANCED

score_each_iteration

score_tree_interval 0

fold_assignment AUTO ▾

offset_column (Choose...) ▾

weights_column (Choose...) ▾

balance_classes

nbins_top_level 1024

nbins_cats 1024

r2_stopping 1.7976931348623157e+308

stopping_rounds 0

stopping_metric AUTO ▾

stopping_tolerance 0.001

max_runtime_secs 0

learn_rate_annealing 1

distribution AUTO ▾

huber_alpha 0.9

checkpoint

col_sample_rate_per_tree 1

min_split_improvement 0.00001

histogram_type AUTO ▾

categorical_encoding AUTO ▾

custom_metric_func

EXPERT

keep_cross_validation_predictions

keep_cross_validation_fold_assignment

build_tree_one_node

sample_rate_per_class

col_sample_rate_change_per_level 1

max_abs_leafnode_pred 1.7976931348623157e+308

pred_noise_bandwidth 0

calibrate_model

calibration_frame

 Build Model

Job

Run Time 00:00:08.569

Remaining Time 00:00:00.0

Type Model

Key  gbm-508038d8-418e-4c20-b39b-ecbafde01b37

Description GBM

Status DONE

Progress 100%

Done.

Actions View

Model

Model ID: gbm-508038d8-418e-4c20-b39b-ecbafde01b37

Algorithm: Gradient Boosting Machine

Actions: Refresh Predict... Download POJO Download Model Deployment Package (MOJO) Export Inspect Delete Download Gen Model

MODEL PARAMETERS

▼ Show all parameters

Parameter	Value	Description
model_id	gbm-508038d8-418e-4c20-b39b-ecbafde01b37	Destination id for this model; auto-generated if not specified.
training_frame	WA_Fn_UseC_HR_Employee_Attrition.hex_0.70_train	Id of the training data frame.
validation_frame	WA_Fn_UseC_HR_Employee_Attrition.hex_0.15_validation	Id of the validation data frame.
nfolds	5	Number of folds for K-fold cross-validation (0 to disable or >= 2).
response_column	Attrition	Response variable column.
ignored_columns	EmployeeCount, EmployeeNumber, Over18, StandardHours, StockOptionLevel	Names of columns to ignore for training.
seed	2143993807418951000	Seed for pseudo random number generator (if applicable)
distribution	bernoulli	Distribution function
custom_metric_func		Reference to custom evaluation function, format: `language:keyName=funcName`

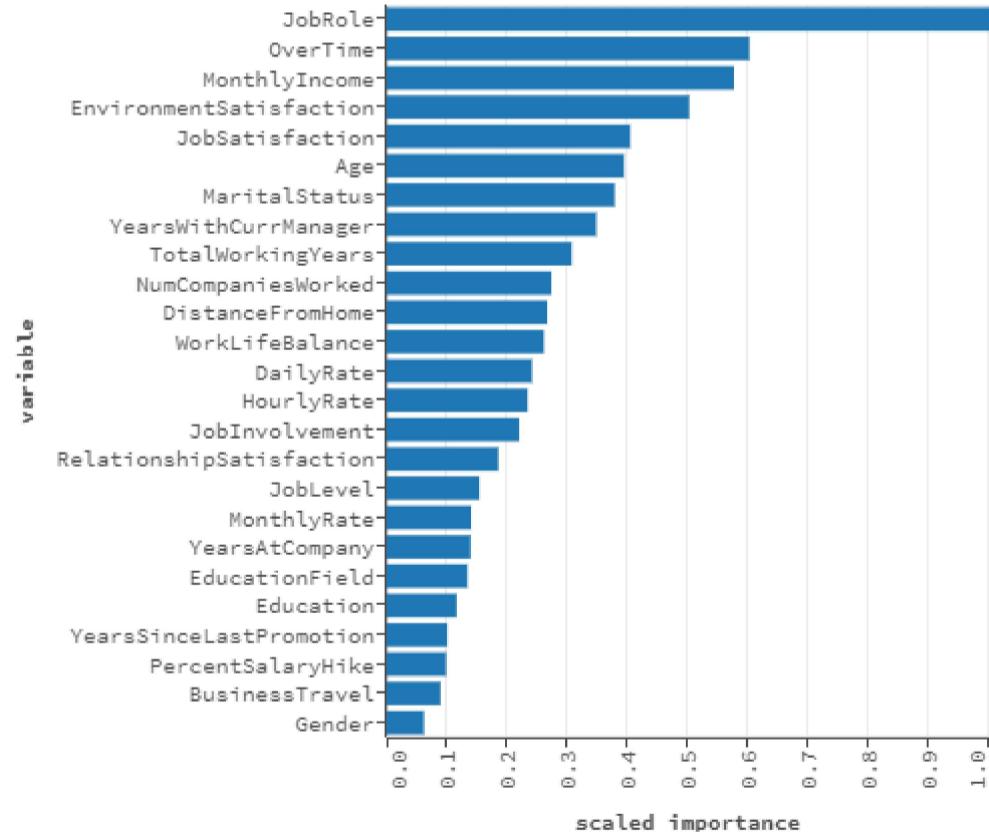
SCORING HISTORY - LOGLOSS

ROC CURVE - TRAINING METRICS , AUC = 0.999699

ROC CURVE - VALIDATION METRICS , AUC = 0.759523

ROC CURVE - CROSS VALIDATION METRICS , AUC = 0.777390

VARIABLE IMPORTANCES



TRAINING METRICS - CONFUSION MATRIX ROW LABELS: ACTUAL CLASS; COLUMN LABELS: PREDICTED CLASS

	No	Yes	Error	Rate	Recall
No	869	1	0.0011	1 / 870	1.0
Yes	4	166	0.0235	4 / 170	0.99
Total	873	167	0.0048	5 / 1,040	
Precision	1.0	0.98			

VALIDATION METRICS - CONFUSION MATRIX ROW LABELS: ACTUAL CLASS; COLUMN LABELS: PREDICTED CLASS

	No	Yes	Error	Rate	Recall
No	156	25	0.1381	25 / 181	0.92
Yes	14	24	0.3684	14 / 38	0.49
Total	170	49	0.1781	39 / 219	
Precision	0.86	0.63			

► CROSS VALIDATION METRICS - CONFUSION MATRIX ROW LABELS: ACTUAL CLASS; COLUMN LABELS: PREDICTED CLASS

	No	Yes	Error	Rate	Recall
No	750	120	0.1379	120 / 870	0.92
Yes	68	102	0.4000	68 / 170	0.46
Total	818	222	0.1808	188 / 1,040	
Precision	0.86	0.60			

► TRAINING METRICS - GAINS/LIFT TABLE

► VALIDATION METRICS - GAINS/LIFT TABLE

► CROSS VALIDATION METRICS - GAINS/LIFT TABLE

► OUTPUT

► CROSS_VALIDATION_MODELS

► OUTPUT - MODEL SUMMARY

► OUTPUT - SCORING HISTORY

► OUTPUT - TRAINING_METRICS

► DOMAIN

► OUTPUT - TRAINING_METRICS - METRICS FOR THRESHOLDS (BINOMIAL METRICS AS A FUNCTION OF CLASSIFICATION THRESHOLDS)

► OUTPUT - TRAINING_METRICS - MAXIMUM METRICS (MAXIMUM METRICS AT THEIR RESPECTIVE THRESHOLDS)

► OUTPUT - TRAINING_METRICS - GAINS/LIFT TABLE (AVG RESPONSE RATE: 16.35 %, AVG SCORE: 16.40 %)

► OUTPUT - VALIDATION_METRICS

► OUTPUT - VALIDATION_METRICS - METRICS FOR THRESHOLDS (BINOMIAL METRICS AS A FUNCTION OF CLASSIFICATION THRESHOLDS)

► OUTPUT - VALIDATION_METRICS - MAXIMUM METRICS (MAXIMUM METRICS AT THEIR RESPECTIVE THRESHOLDS)

► OUTPUT - VALIDATION_METRICS - GAINS/LIFT TABLE (AVG RESPONSE RATE: 17.35 %, AVG SCORE: 14.07 %)

► OUTPUT - CROSS_VALIDATION_METRICS

► OUTPUT - CROSS_VALIDATION_METRICS - METRICS FOR THRESHOLDS (BINOMIAL METRICS AS A FUNCTION OF CLASSIFICATION THRESHOLDS)

► OUTPUT - CROSS_VALIDATION_METRICS - MAXIMUM METRICS (MAXIMUM METRICS AT THEIR RESPECTIVE THRESHOLDS)

► OUTPUT - CROSS_VALIDATION_METRICS - GAINS/LIFT TABLE (AVG RESPONSE RATE: 16.35 %, AVG SCORE: 12.86 %)

► OUTPUT - CROSS-VALIDATION METRICS SUMMARY

► OUTPUT - VARIABLE IMPORTANCES

variable	relative_importance	scaled_importance	percentage
JobRole	72.6629	1.0	0.1366
Overtime	43.7560	0.6022	0.0822
MonthlyIncome	41.8926	0.5765	0.0787
EnvironmentSatisfaction	36.5161	0.5025	0.0686
JobSatisfaction	29.3752	0.4043	0.0552
Age	28.6346	0.3941	0.0538
MaritalStatus	27.5133	0.3786	0.0517
YearsWithCurrManager	25.3036	0.3482	0.0476
TotalWorkingYears	22.3306	0.3073	0.0420
NumCompaniesWorked	19.8586	0.2733	0.0373
DistanceFromHome	19.3862	0.2668	0.0364
WorkLifeBalance	18.9674	0.2610	0.0356
DailyRate	17.5538	0.2416	0.0330
HourlyRate	16.9960	0.2339	0.0319
JobInvolvement	15.9783	0.2199	0.0300
RelationshipSatisfaction	13.4867	0.1856	0.0253
JobLevel	11.1743	0.1538	0.0210
MonthlyRate	10.1858	0.1402	0.0191
YearsAtCompany	10.1216	0.1393	0.0190
EducationField	9.7584	0.1343	0.0183
Education	8.4440	0.1162	0.0159
YearsSinceLastPromotion	7.3011	0.1005	0.0137
PercentSalaryHike	7.1824	0.0988	0.0135
BusinessTravel	6.5001	0.0895	0.0122

▼ PREVIEW POJO

⚡ Predict

Name:

Model: gbm-508038d8-418e-4c20-b39b-ecbafde01b37

Frame: ▾Compute

Leaf Node

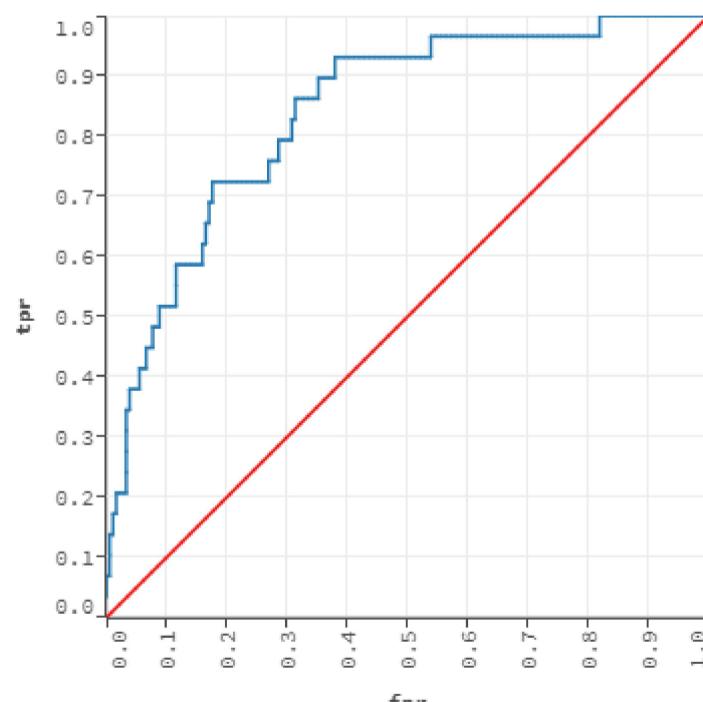
Assignment:

Actions:

⚡ Prediction

Actions:

▼ ROC CURVE



▼ PREDICTION

```

model gbm-508038d8-418e-4c20-b39b-ecbafde01b37
model_checksum 534588866288345280
frame WA_Fn_UseC_HR_Employee_Attrition.hex_0.15_test
frame_checksum -4046159514200313344
description .
model_category Binomial

```

```

scoring_time 1539545509168
predictions prediction-7e873f97-9f05-45bf-948f-6499d8603bc8
    MSE 0.092225
    RMSE 0.303686
    nobs 211
    custom_metric_name .
    custom_metric_value 0
        r2 0.222064
    logloss 0.303185
        AUC 0.839333
    pr_auc 0.456760
        Gini 0.678666
mean_per_class_error 0.225843

```

Combine predictions with frame

► DOMAIN

▼ PREDICTION - CONFUSION MATRIX ROW LABELS: ACTUAL CLASS; COLUMN LABELS: PREDICTED CLASS

	No	Yes	Error	Rate	Recall
No	150	32	0.1758	32 / 182	0.95
Yes	8	21	0.2759	8 / 29	0.40
Total	158	53	0.1896	40 / 211	
Precision	0.82	0.72			

► PREDICTION - METRICS FOR THRESHOLDS

► PREDICTION - MAXIMUM METRICS

► PREDICTION - GAINS/LIFT TABLE

Run AutoML

Project Name:

Training Frame: ▾

Response Column: ▾

Fold Column: ▾

Weights Column: ▾

Ignored Columns:

Showing page 1 of 1.3 ignored.

<input type="checkbox"/> Age	INT
<input type="checkbox"/> Attrition	ENUM(2)
<input type="checkbox"/> BusinessTravel	ENUM(3)
<input type="checkbox"/> DailyRate	INT
<input type="checkbox"/> Department	ENUM(3)
<input type="checkbox"/> DistanceFromHome	INT
<input type="checkbox"/> Education	ENUM(5)
<input type="checkbox"/> EducationField	ENUM(6)
<input checked="" type="checkbox"/> EmployeeCount	INT
<input type="checkbox"/> EmployeeNumber	INT

All None [← Previous 100](#) [Next 100 →](#)

Only show columns with more than 0 % missing values.

Validation Frame: ▾

Leaderboard Frame: ▾

Balance classes:

Exclude these algorithms:

- GLM
- DRF
- GBM
- DeepLearning
- StackedEnsemble

Max models to build:

Max Run Time (sec):

Early stopping metric:

Leaderboard sort metric:

Early stopping rounds:

Early stopping tolerance:

nfold:

Keep cross-validation predictions

Keep cross-validation models

Seed:

Build Model

Job

Run Time 01:00:33.120

Remaining Time 00:00:00.0

Type Auto Model

Key Employee Attrition AutoML

Description AutoML build

Status DONE

Progress 100%

Done.

Actions View

Leaderboard

Monitor Live

MODELS

models sorted in order of auc, best first

	model_id	auc	logloss	mean_per_class_error	rmse
0	GLM_grid_0_AutoML_20181014_195919_model_0	0.8523022312373225	0.31320911173457835	0.23083164300202838	0.30098973
1	DeepLearning_grid_0_AutoML_20181014_195919_model_20	0.8484448951994591	0.35243609766471506	0.21152805949966194	0.29915604
2	DeepLearning_grid_0_AutoML_20181014_195919_model_14	0.8465753887762001	0.3343732774241212	0.2202163624070318	0.30005533
3	DeepLearning_grid_0_AutoML_20181014_195919_model_10	0.846132521974307	0.34574441317995197	0.24601081812035158	0.29829891
4	DeepLearning_grid_0_AutoML_20181014_195919_model_21	0.845841784989858	0.3608989743932463	0.24668695064232588	0.31009247
5	StackedEnsemble_BestOfFamily_0_AutoML_20181014_195919	0.8451487491548343	0.29833680472654667	0.20307640297498308	0.29367698
6	DeepLearning_grid_0_AutoML_20181014_195919_model_11	0.8416632860040568	0.4601483362779503	0.23772819472616633	0.32434052
7	DeepLearning_grid_0_AutoML_20181014_195919_model_6	0.8402129817444219	0.3628892965345627	0.2205206220419202	0.30633490
8	DeepLearning_grid_0_AutoML_20181014_195919_model_3	0.8398208248816768	0.3861212037379842	0.22494929006085193	0.30824523
9	StackedEnsemble_AllModels_0_AutoML_20181014_195919	0.8374645030425963	0.30245807043933787	0.20446247464503042	0.29565432
10	DeepLearning_grid_0_AutoML_20181014_195919_model_27	0.8366599053414469	0.38055076614171174	0.2525693035835024	0.30820424
11	DeepLearning_grid_0_AutoML_20181014_195919_model_38	0.8362237998647735	0.3642995002043426	0.23593644354293442	0.30683304
12	DeepLearning_grid_0_AutoML_20181014_195919_model_31	0.8360547667342799	0.37440792583342525	0.2405341446923597	0.30794024
13	DeepLearning_grid_0_AutoML_20181014_195919_model_32	0.8350202839756593	0.37010937364076324	0.2663962136578769	0.31030354
14	DeepLearning_grid_0_AutoML_20181014_195919_model_4	0.8313353617308993	0.3708341550799633	0.2389452332657201	0.30901371
15	DeepLearning_grid_0_AutoML_20181014_195919_model_37	0.8279006085192697	0.39280179280156596	0.26331981068289384	0.31550807
16	DeepLearning_grid_0_AutoML_20181014_195919_model_13	0.8267241379310345	0.4328318079174838	0.23979039891818799	0.32694061
17	DeepLearning_grid_0_AutoML_20181014_195919_model_0	0.8242562542258283	0.37266340756783906	0.24354293441514538	0.30940426
18	DeepLearning_grid_0_AutoML_20181014_195919_model_23	0.824050033806626	0.39125138429802925	0.2436105476673428	0.32260153
19	DeepLearning_grid_0_AutoML_20181014_195919_model_29	0.8225185936443543	0.4037104740883276	0.24577417173766059	0.31893224
20	DeepLearning_grid_0_AutoML_20181014_195919_model_1	0.8222853279242731	0.40835065986113067	0.26118999323867476	0.32350909
21	DeepLearning_grid_0_AutoML_20181014_195919_model_22	0.8220453008789722	0.4009014645356866	0.23995943204868153	0.31825043
22	DeepLearning_grid_0_AutoML_20181014_195919_model_33	0.8214739688979039	0.37365621934899185	0.2661933739012846	0.31353433

USER FEEDBACK

Predict

Name: Model:

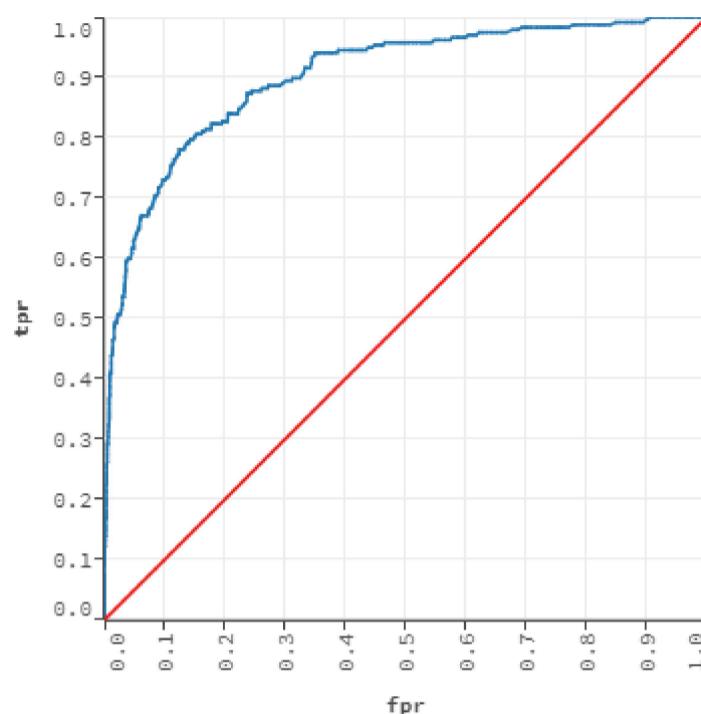
Frame: WA_Fn_UseC_HR_Employee_Attrition.hex

Actions:

⚡ Prediction

Actions:

ROC CURVE



PREDICTION

```

model GBM_grid_0_AutoML_20181014_195919_model_4
model_checksum 2632176253141243904
frame WA_Fn_UseC_HR_Employee_Attrition.hex
frame_checksum -4046162205839642624
description .
model_category Binomial
scoring_time 1539545628481
predictions prediction-87b3698f-2497-46e8-b7f2-09671b1eb4c2
    MSE 0.083582
    RMSE 0.289106
    nobs 1470
custom_metric_name .
custom_metric_value 0
    r2 0.381931
    logloss 0.283994
    AUC 0.903061
    pr_auc 0.728373
    Gini 0.806123
mean_per_class_error 0.196675

```

DOMAIN

PREDICTION - CONFUSION MATRIX ROW LABELS: ACTUAL CLASS; COLUMN LABELS: PREDICTED CLASS

	No	Yes	Error	Rate	Recall
No	1159	74	0.0600	74 / 1,233	0.94
Yes	79	158	0.3333	79 / 237	0.68
Total	1238	232	0.1041	153 / 1,470	
Precision	0.94	0.67			

PREDICTION - METRICS FOR THRESHOLDS

PREDICTION - MAXIMUM METRICS

PREDICTION - GAINS/LIFT TABLE

CS

Export Model

Model: GBM_grid_0_AutoML_20181014_191849_model_4

Path: C:\h2o\model

Overwrite: Actions:

Model Exported

The specified model was exported successfully.

