# AmesStakingModels

Notebook saved.

#### Assistance

Routine Description

☐ importFiles Import file(s) into H<sub>2</sub>O☐ importSqlTable Import SQL table into H<sub>2</sub>O

 $\blacksquare$  getFrames Get a list of frames in  $H_2O$ 

**X** splitFrame Split a frame into two or more frames

 • mergeFrames Merge two frames into one

 • getModels Get a list of models in H₂O

getGrids Get a list of grid search results in H<sub>2</sub>O
 getPredictions Get a list of predictions in H<sub>2</sub>O

≡ getJobs Get a list of jobs running in H<sub>2</sub>O

runAutoML Automatically train and tune many models

buildModel Build a model

importModel Import a saved modelpredict Make a prediction

# ( Import Files

Search: c:\Users\vigarcia\Desktop\Ames\H2o\train.csv

Search Results: (All files added)

Selected Files: 1 file selected: Clear All

\*\* c:\Users\vigarcia\Desktop\Ames\H2o\train.csv

Actions: | Malphort | Malph

# 1 / 1 files imported.

Files C:\Users\vigarcia\Desktop\Ames\H2o\train.csv

Actions Parse these files...

# Setup Parse

PARSE CONFIGURATION

Sources ID train.hex

Parser CSV 
Separator : '044'

Column Headers Auto
First row contains column names
First row contains data

Options Enable single quotes as a field quotation character

✓ Delete on done

Search by column name...

EDIT COLUMN NAMES AND TYPES

1	LotFrontage	Numeric ▼	-0.0518096796424063	0.570487760272347	0.0832634294826105	-0.291196770182661	0
2	LotArea	Numeric ▼	-0.0778773291956158	0.190259941755559	0.523564507451361	0.17928624407868	1
3	YearBuilt	Numeric ▼	1.04607836558577	0.154737490652949	0.98005311559075	-1.85903263419529	0
4	MasVnrArea	Numeric ▼	1.22246022284227	-0.793325290666826	1.15017497762078	-0.793325290666826	1

```
BsmtFinSF2
               Numeric ▼ -0.362503310368561
                                             -0.362503310368561 -0.362503310368561 -0.362503310368561 -
               Numeric ▼
                         -0.3252127614345
                                             0.0145883378473003 0.240792854875087 saved 357449717806527 0
6
  BsmtUnfSF
                                                                 -0.555058934777369 -0.418197897919521 0
  X1stFlrSF
               Numeric ▼ -0.781374035846745
                                             0.43723254456605
  X2ndFlrSF
               Numeric ▼ 1.19667047343491
                                             -0.86346176331678 1.20092357336847
                                                                                    1.15952145066452
              Numeric ▼ | -0.116912359625568
                                            -0.116912359625568 -0.116912359625568 -0.116912359625568 -
   LowQualFinS
10 GrLivArea
               Numeric ▼ 0.577737288088703
                                             -0.368660529155416 0.713219115239534 0.590464992067988
11 BsmtFullBat | Numeric ▼ | 1.08714781758737
                                             -0.818788806759497 1.08714781758737 1.08714781758737
12 BsmtHalfBat
               Numeric ▼ -0.251501295144843 3.88819026609115
                                                                -0.251501295144843 -0.251501295144843 -
13 FullBath
               Numeric ▼ 0.781231961533791
                                             0.781231961533791  0.781231961533791  -1.02718682967726  0
14 HalfBath
               Numeric ▼ 1.23238772066244
                                             -0.756191470389888 1.23238772066244 -0.756191470389888 1
                         0.169897979034312
                                             0.169897979034312  0.169897979034312  0.169897979034312  1
15 BedroomAbvG Numeric ▼
← Previous page
              → Next page
```

**■** Parse

#### **≔** Job

Run Time 00:00:00.751
Remaining Time 00:00:00.0

Type Frame
Key Q train.hex
Description Parse
Status DONE
Progress 100%
Done.
Actions Q View

#### **H** train.hex

Rows			Col	lumns				Comp	ressed Siz	re		
1460			16	4				426K	В			
COLUMN SI	ЈММА	RIES										
label	type	Missing	Zeros	+Inf	-Inf	min	max	mean	sigma	cardinality	Actions	
LotFrontage	real	0	0	0	0	-3.3901	4.6877	0.0293	0.9840	•	•	
otArea	real	0	0	0	0	-4.0104	2.8877	0.0201	0.9809		•	
/earBuilt	real	0	0	0	0	-3.2786	1.2772	-0.0015	0.9971		•	
MasVnrArea	real	Θ	Θ	0	0	-0.7933	2.0219	0.0157	1.0030			
BsmtFinSF2	real	0	Θ	0	0	-0.3625	3.5216	-0.0136	0.9822		•	
BsmtUnfSF	real	0	0	0	0	-3.0092	1.1402	0.0124	0.9918		•	
(1stFlrSF	real	Θ	Θ	0	0	-3.7330	3.0481	0.0126	0.9827			
X2ndFlrSF	real	Θ	Θ	0	0	-0.8635	1.4659	0.0107	1.0050			
LowQualFinSF	real	0	0	0	0	-0.1169	9.5133	0.0344	1.1333		•	
GrLivArea	real	Θ	Θ	0	0	-4.5058	2.9208	0.0182	1.0203			
BsmtFullBath	real	0	Θ	0	0	-0.8188	4.8990	-0.0081	0.9890		•	
BsmtHalfBath	real	0	Θ	0	0	-0.2515	6.3098	-0.0157	0.9711		•	
FullBath	real	0	Θ	0	0	-2.8356	2.5897	-0.0053	0.9963		•	
HalfBath	real	0	Θ	0	0	-0.7562	3.2210	0.0052	1.0		•	
BedroomAbvGr	real	0	Θ	0	0	-3.4767	6.2475	0.0076	0.9916		•	
KitchenAbvGr	real	0	Θ	0	0	-8.0597	7.6587	0.0096	1.0122		•	
TotRmsAbvGrd	real	0	Θ	0	0	-2.8365	4.8098	0.0422	1.0357		•	
Fireplaces	real	0	0	0	0	-0.9242	3.7189	0.0246	0.9977		•	
GarageCars	real	0	0	0	0	-2.3170	2.9310	0.0015	0.9805		•	
VoodDeckSF	real	0	0	0	0	-0.9450	1.6606	0.0029	1.0016		•	

Notebook saved.

lambda max, given lambda is then interpreted as lambda min

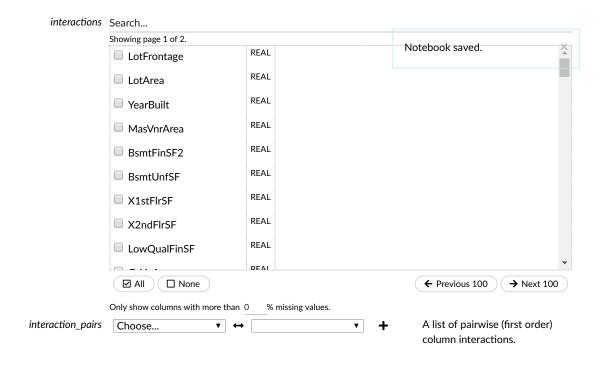
Select an algorithm: Generalized Linear Modeling PARAMETERS GR model\_id Liner\_Regression Destination id for this model; auto-generated if not specified. training\_frame train.hex ▼ Id of the training data frame. Id of the validation data frame. validation\_frame train.hex ▼ Number of folds for K-fold nfolds 10 cross-validation (0 to disable or >= 2). seed 12345 Seed for pseudo random number generator (if applicable) Response variable column. ignored\_columns Search... Showing page 1 of 2. REAL LotFrontage REAL LotArea REAL YearBuilt REAL MasVnrArea **REAL** ■ BsmtFinSF2 REAL BsmtUnfSF REAL X1stFlrSF REAL X2ndFlrSF REAL LowQualFinSF DΕΔΙ → Next 100 ✓ All ☐ None ← Previous 100 Only show columns with more than 0 ignore\_const\_cols 

✓ Ignore constant columns. family gaussian Family. Use binomial for classification with logistic regression, others are for regression problems. solver AUTO • AUTO will set the solver based on given data and the other parameters. IRLSM is fast on on problems with small number of predictors and for lambdasearch with L1 penalty, L\_BFGS scales better for datasets with many columns. alpha 1 Distribution of regularization between the L1 (Lasso) and L2 (Ridge) penalties. A value of 1 for alpha represents Lasso regression, a value of 0 produces Ridge regression, and anything in between specifies the amount of mixing between the two. Default value of alpha is 0 when SOLVER = 'L-BFGS'; 0.5 otherwise. lambda 0 Regularization strength lambda\_search Use lambda search starting at

standardize	<b>₹</b>	Standardize numeric columns to	
		have zero mean and unit Notebook saved.	
non_negative		Restrict coefficients (not	
		intercept) to be non-negative	
beta_constraints	(Choose) ▼	Beta constraints	
obj_reg	-1	Likelihood divider in objective	
		value computation, default is	
		1/nobs	
A D V A N C E D			G R
fold_assignment	AUTO •	Cross-validation fold	
Joiu_ussigiiment	A010 Y	assignment scheme, if	
		fold_column is not specified.	
		The 'Stratified' option will	
		stratify the folds based on the	
		response variable, for	
		classification problems.	
score_each_iteration		Whether to score during each iteration of model training.	
offset_column	(Choose) ▼	Offset column. This will be	
		added to the combination of	
		columns before applying the link function.	
weights_column	(Choose)	Column with observation	
weights_column	(CHOO3C)	weights. Giving some	
		observation a weight of zero is	
		equivalent to excluding it from	
		the dataset; giving an	
		observation a relative weight of	
		2 is equivalent to repeating that row twice. Negative weights are	
		not allowed. Note: Weights are	
		per-row observation weights	
		and do not increase the size of	
		the data frame. This is typically	
		the number of times a row is repeated, but non-integer	
		values are supported as well.	
		During training, rows with	
		higher weights matter more,	
		due to the larger loss function	
		pre-factor.	
compute_p_values		Request p-values computation, p-values work only with IRLSM	
		solver and no regularization	
remove_collinear_columns		In case of linearly dependent	
		columns, remove some of the	
		dependent columns	
max_iterations	-1	Maximum number of iterations	
link	family_default ▼		
export_checkpoints_dir		Automatically export generated	
		models to this directory.	_
max_runtime_secs	0	Maximum allowed runtime in	
		seconds for model training. Use 0 to disable.	
custom_metric_func		Reference to custom evaluation	
		function, format:	
		`language:keyName=funcName`	
EXPERT			G R

keep_cross_validation_predictions		Whether to keep the
reep_cross_validation_predictions	•	predictions of the cross- Noteblack fave models.
$keep\_cross\_validation\_fold\_assignment$		Whether to keep the cross-
		validation fold assignment.
missing_values_handling	MeanImputation ▼	Handling of missing values. Either MeanImputation or Skip.
intercept	•	Include constant term in the model
objective_epsilon	-1	Converge if objective value
		changes less than this. Default indicates: If lambda_search is set to True the value of objective_epsilon is set to .0001. If the lambda_search is set to False and lambda is equal to zero, the value of objective_epsilon is set to .000001, for any other value of lambda the default value of objective_epsilon is set to
beta_epsilon	0.0001	.0001.  Converge if beta changes less
seta_epsilo.ii	0.0001	(using L-infinity norm) than beta esilon, ONLY applies to IRLSM solver
gradient_epsilon		Converge if objective changes less (using L-infinity norm) than this, ONLY applies to L-BFGS solver. Default indicates: If lambda_search is set to False and lambda is equal to zero, the default value of gradient_epsilon is equal to .000001, otherwise the default value is .0001. If lambda_search is set to True, the conditional values above are 1E-8 and 1E-6 respectively.
prior	-1	Prior probability for y==1. To be used only for logistic regression iff the data has been sampled and the mean of response does not reflect reality.
max_active_predictors	_1	Maximum number of active predictors during computation. Use as a stopping criterion to prevent expensive model building with many predictors. Default indicates: If the IRLSM solver is used, the value of max_active_predictors is set to 5000 otherwise it is set to

100000000.



**⊞** Build Model

### **≆≡** Job

Run Time 00:00:01.641

Remaining Time 00:00:00.0

Type Model

Key Q Liner\_Regression

Description GLM

Status DONE

Progress 100%

Done.

Don

Actions Q View

# Build a Model

PARAME

Select an algorithm: Generalized Linear Modeling ▼

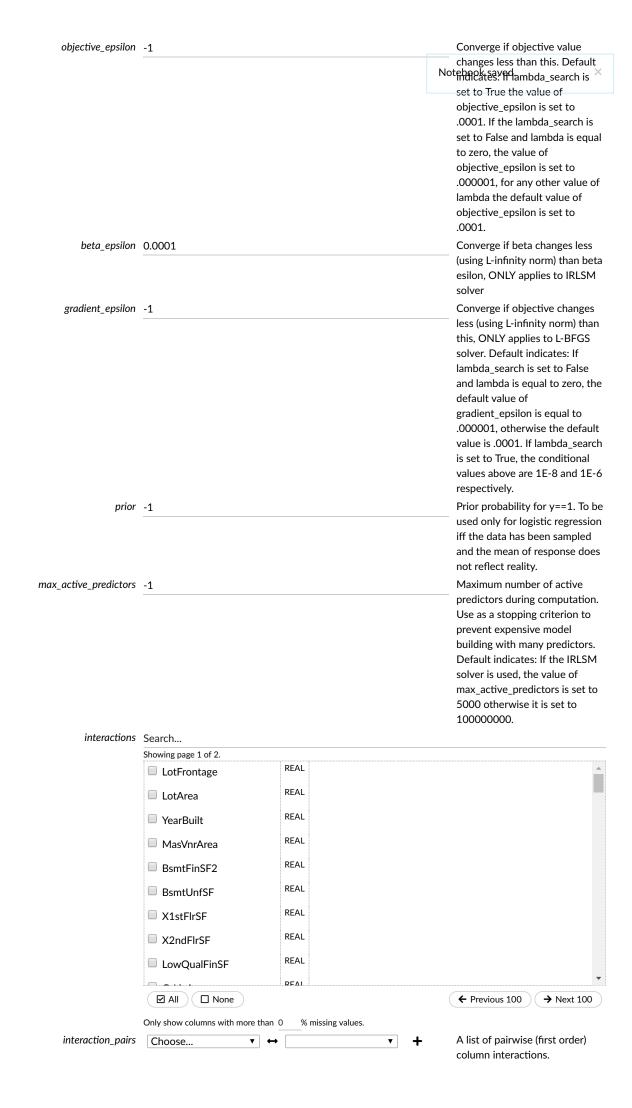
TERS				GR
	model_id	Lasso_Regression	Destination id for this model; auto-generated if not specified.	
t	training_frame	train.hex ▼	Id of the training data frame.	
val	lidation_frame	train.hex •	Id of the validation data frame.	
	nfolds	10	Number of folds for K-fold cross-validation (0 to disable or	
			>= 2).	
	seed	12345	Seed for pseudo random number generator (if applicable)	
resp	ponse_column	SalePrice ▼	Response variable column.	

	Showing page 1 of 2.		Nietokooleenkool	
	Showing page 1 of 2.  LotFrontage  LotArea  YearBuilt  MasVnrArea  BsmtFinSF2  BsmtUnfSF  X1stFlrSF  X2ndFlrSF  LowQualFinSF	REAL REAL REAL REAL REAL REAL REAL REAL	Notebook saved.	
	☑ All None		← Previous 100 → Next 100	
ignore_const_cols	Only show columns with more t	than 0 % missing values.	Ignore constant columns.	
_			Family. Use binomial for classification with logistic regression, others are for regression problems.	
solver	AUTO	•	AUTO will set the solver based on given data and the other parameters. IRLSM is fast on on problems with small number of predictors and for lambdasearch with L1 penalty, L_BFGS scales better for datasets with many columns.	
alpha	1		Distribution of regularization between the L1 (Lasso) and L2 (Ridge) penalties. A value of 1 for alpha represents Lasso regression, a value of 0 produces Ridge regression, and anything in between specifies the amount of mixing between the two. Default value of alpha is 0 when SOLVER = 'L-BFGS'; 0.5 otherwise.	
lambda	0.0001		Regularization strength	
lambda_search standardize			Use lambda search starting at lambda max, given lambda is then interpreted as lambda min Standardize numeric columns to have zero mean and unit	
			variance	
non_negative beta_constraints		▼	Restrict coefficients (not intercept) to be non-negative  Beta constraints	
obj_reg			Likelihood divider in objective value computation, default is 1/nobs	
DVANCED				GR
fold_assignment	AUTO v		Cross-validation fold assignment scheme, if fold_column is not specified. The 'Stratified' option will stratify the folds based on the response variable, for	

classification problems.

ignored\_columns Search...

score_each_iteration		Whether to score during each	
offset_column	(Choose)	iteration of model training. Notebook saved. Offset column. This will be	
		added to the combination of	
		columns before applying the link function.	
weights_column	(Choose) ▼	Column with observation weights. Giving some observation a weight of zero is equivalent to excluding it from the dataset; giving an observation a relative weight of 2 is equivalent to repeating that row twice. Negative weights are not allowed. Note: Weights are per-row observation weights and do not increase the size of the data frame. This is typically the number of times a row is repeated, but non-integer values are supported as well. During training, rows with higher weights matter more, due to the larger loss function pre-factor.	
compute_p_values		Request p-values computation, p-values work only with IRLSM solver and no regularization	
remove_collinear_columns		In case of linearly dependent columns, remove some of the dependent columns	
max_iterations	-1	Maximum number of iterations	
link	family_default ▼		
	ranniy_ucradit ·		
export_checkpoints_dir		Automatically export generated models to this directory.	
max_runtime_secs	0	Maximum allowed runtime in	
		seconds for model training. Use 0 to disable.	
custom_metric_func		Reference to custom evaluation	
		function, format:	
		`language:keyName=funcName`	
EXPERT			G R
keep_cross_validation_models		Whether to keep the cross-	
keep_cross_validation_models		validation models.	
keep_cross_validation_predictions		Whether to keep the predictions of the cross-validation models.	
keep_cross_validation_fold_assignment	€	Whether to keep the cross-validation fold assignment.	
missing_values_handling	MeanImputation ▼	Handling of missing values. Either MeanImputation or Skip.	
intercept	€	Include constant term in the model	



regression problems.

#### X

#### **≔** Job

Run Time 00:00:00.664

Remaining Time 00:00:00.0

Type Model

Key Q Lasso\_Regression

Description GLM

Status DONE

Progress 100%

Done.

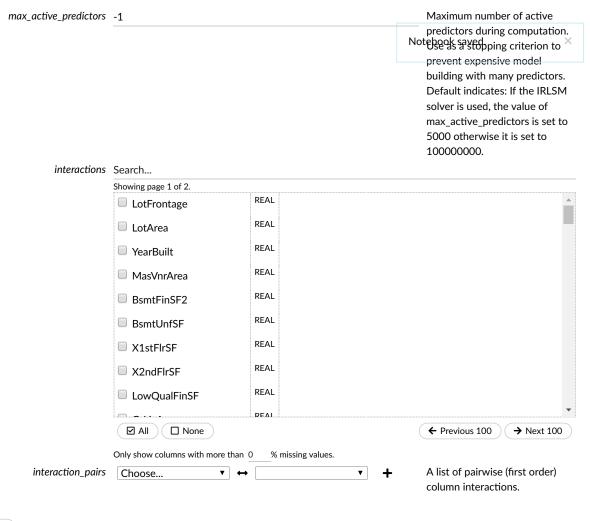
Actions Q View

#### Build a Model

Select an algorithm: Generalized Linear Modeling PARAMETERS GR model\_id Ridge\_Regression Destination id for this model; auto-generated if not specified. • training\_frame train.hex Id of the training data frame. • Id of the validation data frame. validation\_frame train.hex Number of folds for K-fold nfolds 10 cross-validation (0 to disable or >= 2). seed 12345 Seed for pseudo random number generator (if applicable) Response variable column. ignored\_columns Search... Showing page 1 of 2. REAL LotFrontage REAL LotArea YearBuilt REAL REAL MasVnrArea REAL ■ BsmtFinSF2 REAL BsmtUnfSF REAL X1stFlrSF REAL X2ndFlrSF REAL LowQualFinSF DΕΔΙ → Next 100 ← Previous 100 Only show columns with more than 0 % missing values. ignore\_const\_cols <a> ✓</a> Ignore constant columns. family gaussian Family. Use binomial for classification with logistic regression, others are for

	AUTO	·	AUTO will set the solver based	
		N	on given data and the other otebook sayed Parameters. IRLSM is fast on on	
			problems with small number of predictors and for lambda- search with L1 penalty, L_BFGS scales better for datasets with	
alpha			many columns.  Distribution of regularization between the L1 (Lasso) and L2 (Ridge) penalties. A value of 1 for alpha represents Lasso regression, a value of 0 produces Ridge regression, and anything in between specifies the amount of mixing between the two. Default value of alpha is 0 when SOLVER = 'L-BFGS'; 0.5 otherwise.	
lambda	0.0102		Regularization strength	
lambda_search standardize			Use lambda search starting at lambda max, given lambda is then interpreted as lambda min Standardize numeric columns to	
non_negative			have zero mean and unit variance Restrict coefficients (not	
beta_constraints	(Choose )	▼	intercept) to be non-negative  Beta constraints	
obj_reg			Likelihood divider in objective	
<i>2</i> 0			value computation, default is 1/nobs	
ADVANCED				G R
fold_assignment	AUTO ▼		Cross-validation fold	
, <u>-</u> <b>g</b>			assignment scheme, if fold_column is not specified. The 'Stratified' option will stratify the folds based on the response variable, for classification problems.	
score_each_iteration			fold_column is not specified. The 'Stratified' option will stratify the folds based on the response variable, for	
			fold_column is not specified. The 'Stratified' option will stratify the folds based on the response variable, for classification problems. Whether to score during each	

compute_p_values		Request p-values computation, p-values work only with IRLSM Notebook saved Notebook saved	
remove_collinear_columns max_iterations		in case of linearly dependent columns, remove some of the dependent columns Maximum number of iterations	
link	family_default ▼		
export_checkpoints_dir		Automatically export generated models to this directory.	
max_runtime_secs custom_metric_func	0	Maximum allowed runtime in seconds for model training. Use 0 to disable.  Reference to custom evaluation	
		function, format: `language:keyName=funcName`	
EXPERT			GR
keep_cross_validation_models		Whether to keep the cross- validation models.	
keep_cross_validation_predictions	€	Whether to keep the predictions of the cross-validation models.	
keep_cross_validation_fold_assignment	<b>€</b>	Whether to keep the cross- validation fold assignment.	
missing_values_handling	MeanImputation ▼	Handling of missing values. Either MeanImputation or Skip.	
intercept		Include constant term in the model	
objective_epsilon	-1	Converge if objective value changes less than this. Default indicates: If lambda_search is set to True the value of objective_epsilon is set to .0001. If the lambda_search is set to False and lambda is equal to zero, the value of objective_epsilon is set to .000001, for any other value of lambda the default value of objective_epsilon is set to .0001.	
beta_epsilon	0.0001	Converge if beta changes less (using L-infinity norm) than beta esilon, ONLY applies to IRLSM solver	
gradient_epsilon	-1	Converge if objective changes less (using L-infinity norm) than this, ONLY applies to L-BFGS solver. Default indicates: If lambda_search is set to False and lambda is equal to zero, the default value of gradient_epsilon is equal to .000001, otherwise the default value is .0001. If lambda_search is set to True, the conditional values above are 1E-8 and 1E-6 respectively.	
prior	-1	Prior probability for y==1. To be used only for logistic regression iff the data has been sampled and the mean of response does not reflect reality.	



Build Model

#### **₹** Job

Run Time 00:00:00.420

Remaining Time 00:00:00.0

Type Model

Key Q Ridge\_Regression

Description GLM

Status DONE

Progress 100%

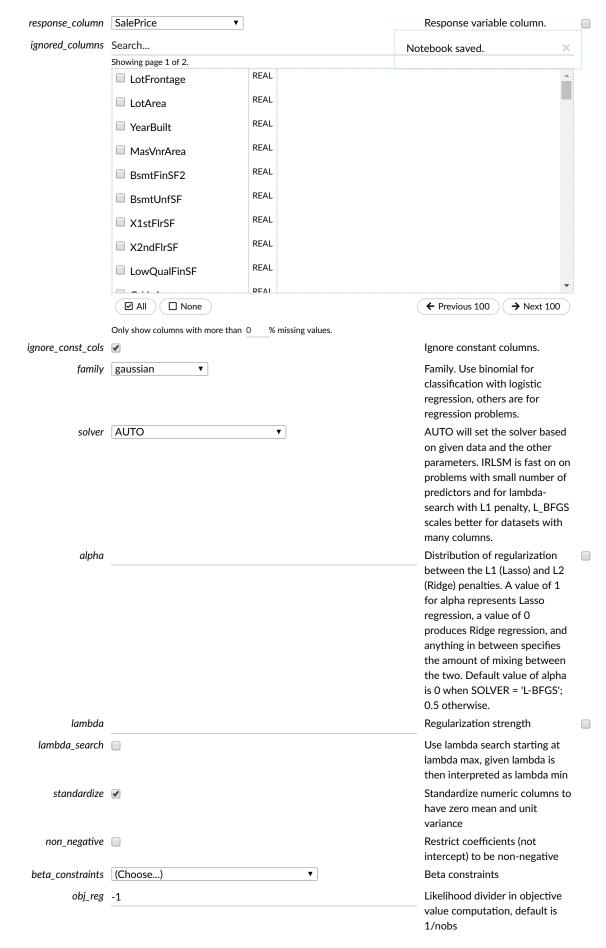
Done.

Actions Q View

# Build a Model

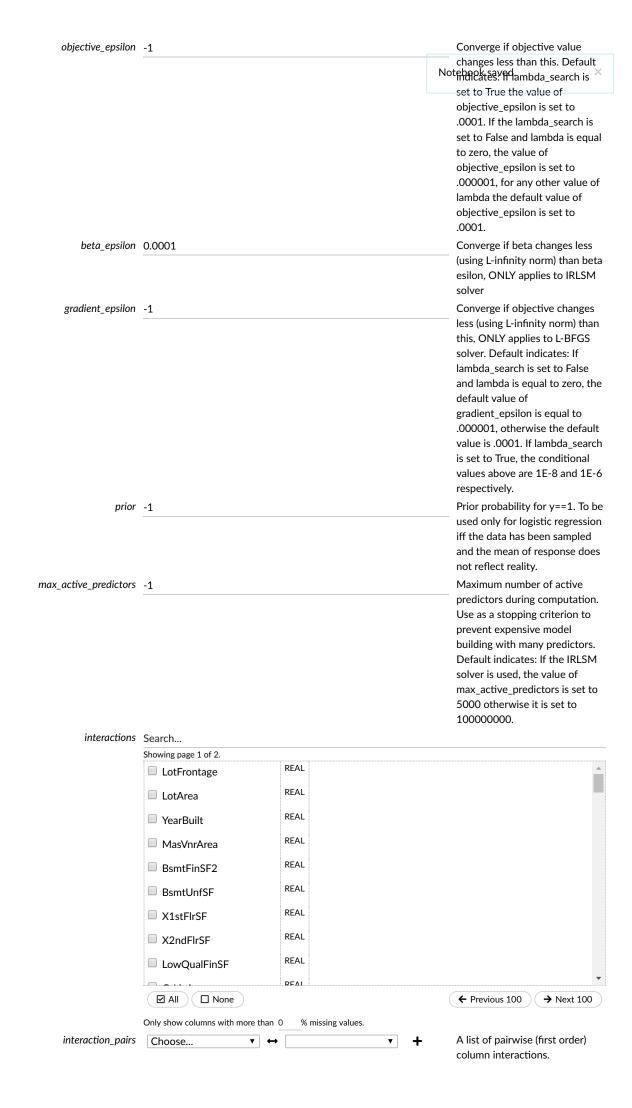
Select an algorithm: Generalized Linear Modeling

PARAMETERS GR model\_id ElasticNet Destination id for this model; auto-generated if not specified. training\_frame train.hex Id of the training data frame. validation\_frame train.hex Id of the validation data frame. Number of folds for K-fold nfolds 10 cross-validation (0 to disable or >= 2). Seed for pseudo random seed 12345 number generator (if applicable)



A D V A N C E D G R i

fold_assignment	AUTO ▼	Cross-validation fold	
		assignment scheme, if Notebook sayed is not specified.  The 'Stratified' option will	
		stratify the folds based on the	
		response variable, for	
		classification problems.	
score_each_iteration		Whether to score during each iteration of model training.	
offset_column	(Choose)	Offset column. This will be	
		added to the combination of	
		columns before applying the link function.	
weights_column	(Choose) ▼	Column with observation	
weights_column	(choose)	weights. Giving some	
		observation a weight of zero is	
		equivalent to excluding it from	
		the dataset; giving an	
		observation a relative weight of 2 is equivalent to repeating that	
		row twice. Negative weights are	
		not allowed. Note: Weights are	
		per-row observation weights	
		and do not increase the size of the data frame. This is typically	
		the number of times a row is	
		repeated, but non-integer	
		values are supported as well.	
		During training, rows with higher weights matter more,	
		due to the larger loss function	
		pre-factor.	
compute_p_values		Request p-values computation,	
		p-values work only with IRLSM solver and no regularization	
remove_collinear_columns		In case of linearly dependent	
remove_commear_commis		columns, remove some of the	
		dependent columns	
max_iterations	-1	Maximum number of iterations	
link	family_default ▼		
export_checkpoints_dir		Automatically export generated	
		models to this directory.	
max_runtime_secs	0	Maximum allowed runtime in seconds for model training. Use	
		0 to disable.	
custom_metric_func		Reference to custom evaluation	
		function, format:	
		`language:keyName=funcName`	
EXPERT			GR
keep_cross_validation_models		Whether to keep the cross-validation models.	
keep_cross_validation_predictions		Whether to keep the	
		predictions of the cross- validation models.	
keep_cross_validation_fold_assignment	<b>€</b>	Whether to keep the cross- validation fold assignment.	
missing_values_handling	MeanImputation ▼	Handling of missing values. Either MeanImputation or Skip.	
intercept	€	Include constant term in the model	



#### X

GRID?

#### **≔** Job

Run Time 00:00:00.344

Remaining Time 00:00:00.0

Type Model

Key Q ElasticNet

Description GLM

Status DONE

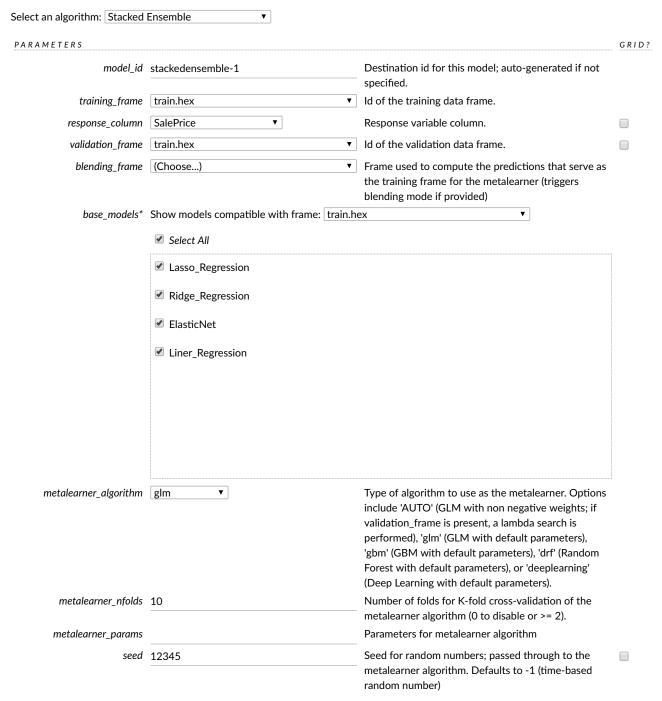
Progress 100%

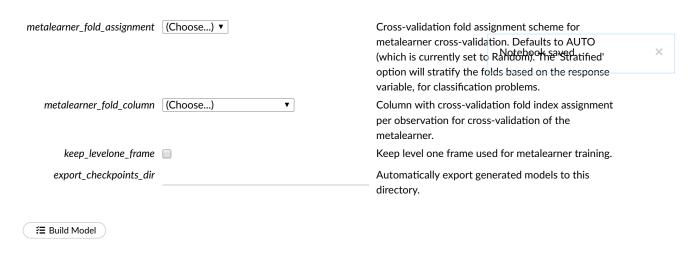
Done.

Actions Q View

#### Build a Model

ADVANCED





#### **₹** Job

Run Time 00:00:00.455

Remaining Time 00:00:00.0

Type Model

Key Q stackedensemble-1

Description StackedEnsemble

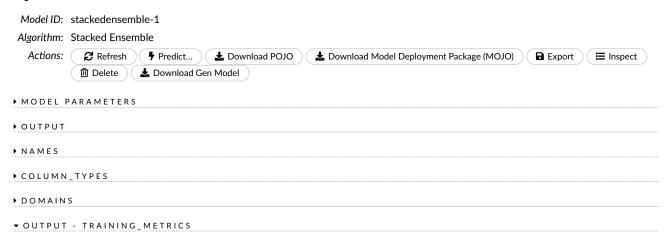
Status DONE

Progress 100%

Done.

Actions Q View

## **⊘** Model



```
model stackedensemble-1
             model_checksum 4766109664223675936
                      frame train.hex
             frame_checksum -4933500844521803980
                description .
             model_category Regression
               scoring_time 1568289978464
                predictions ·
                       MSE 0.011326
                       RMSE 0.106425
                       nobs 1460
         \textit{custom\_metric\_name} \quad \cdot
        custom_metric_value 0
                        r2 0.928968
    mean_residual_deviance 0.011326
                       mae 0.073881
                      rmsle 0.008231
          residual_deviance 16.536328
              null_deviance 232.800671
                       AIC -2386.422273
    null_degrees_of_freedom 1459
residual_degrees_of_freedom 1455
```

Notebook saved.

#### ▼ OUTPUT - VALIDATION\_METRICS

```
model metalearner_glm_stackedensemble-1
            model_checksum -661633969583491072
                     frame .
            frame_checksum 0
               description ·
            model_category Regression
              scoring_time 1568289978256
               predictions ·
                      MSE 0.011326
                     RMSE 0.106425
                     nobs 1460
        custom_metric_name .
       custom_metric_value 0
                       r2 0.928968
    mean_residual_deviance 0.011326
                      mae 0.073881
                     rmsle 0.008231
         residual_deviance 16.536328
             null_deviance 232.800671
                      AIC -2386.422273
   null_degrees_of_freedom 1459
residual_degrees_of_freedom 1455
```

## ▼ OUTPUT - CROSS\_VALIDATION\_METRICS

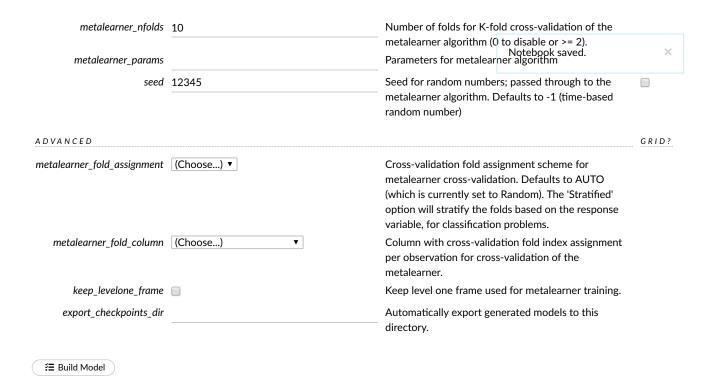
model stackedensemble-1 model\_checksum 4766109664223675936 Notebook saved. frame · frame checksum 0 description 10-fold cross-validation on training data (Metrics computed for combined holdout predictions) model\_category Regression scoring\_time 1568289978257 predictions · MSE 0.015060 RMSE 0.122719 nobs 1460  $custom\_metric\_name$  . custom\_metric\_value 0 r2 0.905552 mean\_residual\_deviance 0.015060 mae 0.082725 rmsle 0.009486 residual\_deviance 21.987593 null\_deviance 232.936688 AIC -1970.441037 null\_degrees\_of\_freedom 1459 residual\_degrees\_of\_freedom 1455

OUTPUT - HELP

### Build a Model

Select an algorithm: Stacked Ensemble PARAMETERS GRID? model\_id stackedensemble-2 Destination id for this model; auto-generated if not specified. training\_frame train.hex ▼ Id of the training data frame. Response variable column. validation\_frame train.hex ▼ Id of the validation data frame. blending\_frame (Choose...) ▼ Frame used to compute the predictions that serve as the training frame for the metalearner (triggers blending mode if provided) base\_models\* Show models compatible with frame: train.hex ▼ Select All ✓ Ridge\_Regression ElasticNet ✓ Liner\_Regression Lasso\_Regression stackedensemble-1 metalearner\_algorithm gbm Type of algorithm to use as the metalearner. Options

include 'AUTO' (GLM with non negative weights; if validation\_frame is present, a lambda search is performed), 'glm' (GLM with default parameters), 'gbm' (GBM with default parameters), 'drf' (Random Forest with default parameters), or 'deeplearning' (Deep Learning with default parameters).



**≆≣** Job

Run Time 00:00:03.141

Remaining Time 00:00:00.0

Type Model

Key **Q** stackedensemble-2

Description StackedEnsemble

Status DONE

Progress 100%

Done.

Actions Q View

# **₩** Model

Model ID:	stackedensemble-2
Algorithm:	Stacked Ensemble
Actions:	
	☐ Delete
▶ MODEL F	PARA M ETERS
• OUTPUT	
• NAMES	
• COLUMN	_TYPES
DOMAIN	S
▼ OUTPUT	- TRAINING_METRICS

```
model_checksum -684629815996900896
                                                                                       Notebook saved.
                 frame train.hex
        frame_checksum -4933500844521803980
           description ·
        model_category Regression
          scoring_time 1568290136086
           predictions ·
                   MSE 0.010847
                  RMSE 0.104150
                  nobs 1460
    \textit{custom\_metric\_name} \quad \cdot
   custom_metric_value 0
                   r2 0.931972
 mean_residual_deviance 0.010847
                  mae 0.071845
                 rmsle 0.008047
▼ OUTPUT - VALIDATION_METRICS
                 model metalearner_gbm_stackedensemble-2
        model_checksum 3814499315818121984
                 frame ·
        frame_checksum 0
           description ·
        model_category Regression
          scoring_time 1568290135993
           predictions ·
                   MSE 0.010847
                  RMSE 0.104150
                  nobs 1460
    custom_metric_name ·
   custom_metric_value 0
                   r2 0.931972
 mean_residual_deviance 0.010847
                  mae 0.071845
                 rmsle 0.008047
▼ OUTPUT - CROSS_VALIDATION_METRICS
                 model stackedensemble-2
        model_checksum -684629815996900896
         frame_checksum 0
           description 10-fold cross-validation on training data (Metrics computed for combined holdout predictions)
        model_category Regression
          scoring_time 1568290135998
           predictions ·
                   MSE 0.016408
                  RMSE 0.128094
                  nobs 1460
    custom_metric_name .
   custom_metric_value 0
                   r2 0.897097
 mean_residual_deviance 0.016408
                   mae 0.086703
                 rmsle 0.009909
▶ OUTPUT - HELP
Build a Model
Select an algorithm: Stacked Ensemble
PARAMETERS
                                                                                                               GRID?
```

model\_id stackedensemble-3

model stackedensemble-2

Destination id for this model; auto-generated if not specified.

training_frame	train.hex	Id of the training data frame.		
response_column	SalePrice •	Response variable column. Notebook saved.		×
validation_frame	train.hex	Id of the validation data frame.		
blending_frame	(Choose)	Frame used to compute the predictions that serve as the training frame for the metalearner (triggers blending mode if provided)		
base_models*	Show models compatible with frame: train.			
	Select All			
	stackedensemble-2			
	▼ Ridge_Regression			
	✓ Liner_Regression			
	✓ Lasso_Regression			
	stackedensemble-1			
metalearner_algorithm	drf ▼	Type of algorithm to use as the metalearner. Options include 'AUTO' (GLM with non negative weights; if validation_frame is present, a lambda search is performed), 'glm' (GLM with default parameters), 'gbm' (GBM with default parameters), 'drf' (Random Forest with default parameters), or 'deeplearning' (Deep Learning with default parameters).		
metalearner_nfolds	10	Number of folds for K-fold cross-validation of the		
metalearner_params		metalearner algorithm (0 to disable or >= 2).  Parameters for metalearner algorithm		
seed	12345	Seed for random numbers; passed through to the metalearner algorithm. Defaults to -1 (time-based random number)		
A D V A N C E D			GRI	D?
metalearner_fold_assignment	(Choose) ▼	Cross-validation fold assignment scheme for metalearner cross-validation. Defaults to AUTO (which is currently set to Random). The 'Stratified' option will stratify the folds based on the response variable, for classification problems.		
metalearner_fold_column	(Choose)	Column with cross-validation fold index assignment per observation for cross-validation of the metalearner.		
keep_levelone_frame		Keep level one frame used for metalearner training.		
export_checkpoints_dir		Automatically export generated models to this directory.		
≆ Puild Model				
Build Model     Build				

# **≨≣** Job

Run Time 00:00:04.128
Remaining Time 00:00:00.0
Type Model

Key **Q** stackedensemble-3

Description StackedEnsemble

Status DONE

Progress 100%

Done.

Actions Q View

Notebook saved.

 $\times$ 

**⊘** Model Model ID: stackedensemble-3 Algorithm: Stacked Ensemble Actions: Refresh Predict... Download POJO Lownload Model Deployment Package (MOJO) Export (\equiv Inspect ► MODEL PARAMETERS ▶ OUTPUT ▶ COLUMN\_TYPES ▶ DOMAINS ▼ OUTPUT - TRAINING\_METRICS model stackedensemble-3 model\_checksum -7085086949227900512 frame train.hex frame\_checksum -4933500844521803980 description ·  $model\_category$  Regression

scoring\_time 1568290243297

predictions ·

MSE 0.010732

RMSE 0.103594 nobs 1460

r2 0.932697

rmsle 0.008006

#### ▼ OUTPUT - VALIDATION\_METRICS

model metalearner\_drf\_stackedensemble-3

model\_checksum -4344697905864719784

frame ·

frame\_checksum 0

description .

model\_category Regression
scoring\_time 1568290243258

predictions .

MSE 0.010732

RMSE 0.103594

nobs 1460

 $\textit{custom\_metric\_name} \quad \cdot$ 

custom\_metric\_value 0

r2 0.932697

mean\_residual\_deviance 0.010732

mae 0.070900

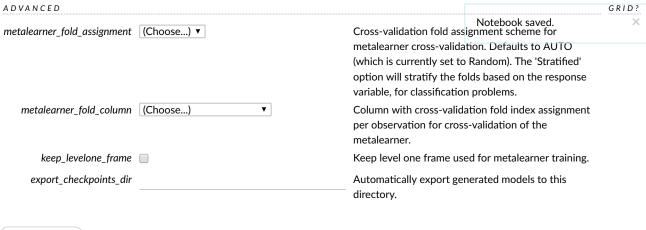
rmsle 0.008006

model stackedensemble-3 model\_checksum -7085086949227900512 Notebook saved. frame · frame checksum 0 description 10-fold cross-validation on training data (Metrics computed for combined holdout predictions) model\_category Regression scoring\_time 1568290243261 predictions · MSE 0.017421 RMSE 0.131990 nobs 1460  $\textit{custom\_metric\_name} \quad \cdot$ custom\_metric\_value 0 r2 0.890743 mean\_residual\_deviance 0.017421 mae 0.090034 rmsle 0.010199

▶ OUTPUT - HELP

## Build a Model

Select an algorithm: Stacked Ensemble PARAMETERS GRID? model\_id stackedensemble-4 Destination id for this model; auto-generated if not specified. training\_frame train.hex ▼ Id of the training data frame. response\_column | SalePrice • Response variable column. ▼ Id of the validation data frame. validation\_frame | train.hex Frame used to compute the predictions that serve as blending\_frame (Choose...) the training frame for the metalearner (triggers blending mode if provided) base\_models\* Show models compatible with frame: train.hex ▼ Select All stackedensemble-2 ✓ Ridge\_Regression ElasticNet ✓ Liner\_Regression ✓ Lasso\_Regression stackedensemble-3 stackedensemble-1 metalearner\_algorithm deeplearning ▼ Type of algorithm to use as the metalearner. Options include 'AUTO' (GLM with non negative weights; if validation\_frame is present, a lambda search is performed), 'glm' (GLM with default parameters), 'gbm' (GBM with default parameters), 'drf' (Random Forest with default parameters), or 'deeplearning' (Deep Learning with default parameters). metalearner\_nfolds 10 Number of folds for K-fold cross-validation of the metalearner algorithm (0 to disable or >= 2). Parameters for metalearner algorithm metalearner\_params Seed for random numbers; passed through to the seed 12345 metalearner algorithm. Defaults to -1 (time-based random number)



Build Model

#### **₹** Job

Run Time 00:00:26.534

Remaining Time 00:00:00.0

Type Model

Key Q stackedensemble-4

Description StackedEnsemble

Status DONE

Progress 100%

Done.

Actions Q View

## **⊘** Model

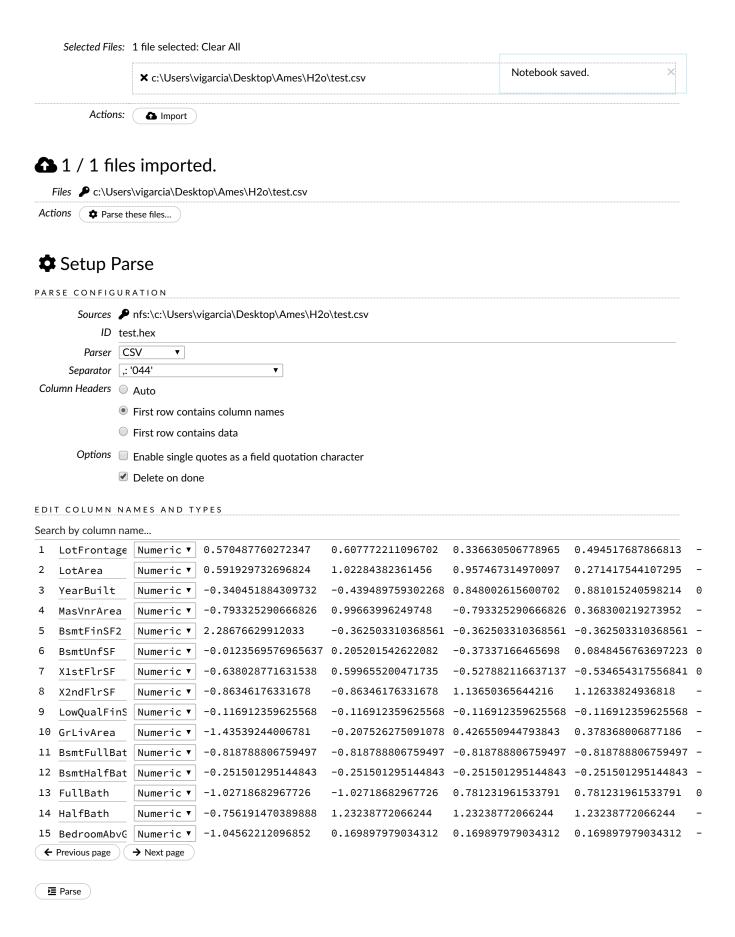
```
model stackedensemble-4
        model_checksum -1219408000543728064
                                                                                        Notebook saved.
                 frame train.hex
         frame_checksum -4933500844521803980
           description ·
        model_category Regression
          scoring_time 1568290344441
           predictions ·
                   MSE 0.011754
                  RMSE 0.108417
                  nobs 1460
    \textit{custom\_metric\_name} \quad \cdot
   custom_metric_value 0
                   r2 0.926284
 mean_residual_deviance 0.011754
                  mae 0.076771
                 rmsle 0.008391
▼ OUTPUT - VALIDATION_METRICS
                 model metalearner_deeplearning_stackedensemble-4
        model_checksum -1569179343419488056
                 frame ·
         frame_checksum 0
           description Metrics reported on full validation frame
        model_category Regression
          scoring_time 1568290343908
           predictions ·
                   MSE 0.011754
                  RMSE 0.108417
                  nobs 1460
    custom metric name ·
   custom_metric_value 0
                   r2 0.926284
 mean_residual_deviance 0.011754
                  mae 0.076771
                 rmsle 0.008391
▼ OUTPUT - CROSS_VALIDATION_METRICS
                 model stackedensemble-4
        model_checksum -1219408000543728064
         frame_checksum 0
           description 10-fold cross-validation on training data (Metrics computed for combined holdout predictions)
        model_category Regression
          scoring_time 1568290343915
           predictions ·
                   MSE 0.015660
                  RMSE 0.125139
                  nobs 1460
    custom_metric_name .
   custom_metric_value 0
                   r2 0.901791
 mean_residual_deviance 0.015660
                   mae 0.084490
                 rmsle 0.009684
▶ OUTPUT - HELP
```

# ( Import Files

Search: c:\Users\vigarcia\Desktop\Ames\H2o\test.csv

Search Results: (All files added)

(Q)



#### ₹**三** Job

Run Time 00:00:00.191
Remaining Time 00:00:00.0

Type Frame
Key **Q** test.hex
Description Parse
Status DONE

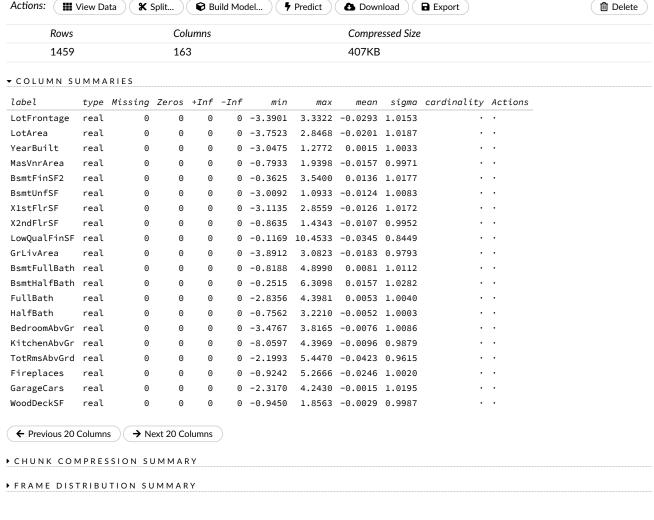
Progress 100%

Done.

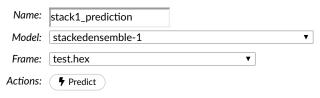
Actions Q View

Notebook saved.

#### **test.hex**







# **7** Prediction

# **⊞** stack1\_prediction

Rows	Columns	Compressed Size	
1459	1	12KB	



