Homework_Regression1_UsedCarSales

❷ Assistance

	Routine	Description
⎘	importFiles	Import file(s) into H_2O
▦	importSqlTable	Import SQL table into H_2O
▦	getFrames	Get a list of frames in H_2O
✂	splitFrame	Split a frame into two or more frames
🔗	mergeFrames	Merge two frames into one
⛁	getModels	Get a list of models in H_2O
⠿	getGrids	Get a list of grid search results in H_2O
⚡	getPredictions	Get a list of predictions in H_2O
☷	getJobs	Get a list of jobs running in H_2O
⬡	runAutoML	Automatically train and tune many models
◈	buildModel	Build a model
◈	importModel	Import a saved model
⚡	predict	Make a prediction

⎙ Import Files

Search:	/Users/apple/Downloads/dataset_regression_car_prices.csv 🔍
Search Results:	(All files added)
Selected Files:	1 file selected: Clear All

✖ /Users/apple/Downloads/dataset_regression_car_prices.csv

Actions: ☁ Import

☁ 1 / 1 files imported.

Files 🔑 /Users/apple/Downloads/dataset_regression_car_prices.csv

Actions ⚙ Parse these files...

⚙ Setup Parse

PARSE CONFIGURATION

Sources 🔑 nfs://Users/apple/Downloads/dataset_regression_car_prices.csv

ID dataset_regression_car_prices.hex

Parser [CSV ⇅]

Separator [,: '044' ⇅]

Escape Character 0

Column Headers ○ Auto

• First row contains column names

○ First row contains data

Options ☐ Enable single quotes as a field quotation character

✓ Delete on done

EDIT COLUMN NAMES AND TYPES

Search by column name...

1	car_Name	Enum	ritz	sx4	ciaz	wagon r	swift	vitara brezza	ciaz	s cross	ciaz
2	year	Numeric	2014	2013	2017	2011	2014	2018	2015	2015	2016
3	selling_price	Numeric	3.35	4.75	7.25	2.85	4.6	9.25	6.75	6.5	8.75
4	present_price	Numeric	5.59	9.54	9.85	4.15	6.87	9.83	8.12	8.61	8.89
5	kms_driven	Numeric	27000	43000	6900	5200	42450	2071	18796	33429	20273
6	fuel_type	Enum	Petrol	Diesel	Petrol	Petrol	Diesel	Diesel	Petrol	Diesel	Diesel
7	seller_type	Enum	Dealer	Dealer	Dealer	Dealer	Dealer	Dealer	Dealer	Dealer	Dealer
8	transmission	Enum	Manual	Manual	Manual	Manual	Manual	Manual	Manual	Manual	Manual
9	owner	Numeric	0	0	0	0	0	0	0	0	0

← Previous page   → Next page

≔ Parse

≔ Job

Run Time	00:00:00.208
Remaining Time	00:00:00.0
Type	Frame
Key	🔍 dataset_regression_car_prices.hex
Description	Parse
Status	DONE
Progress	100%

Done.

Actions   🔍 View

⊞ dataset_regression_car_prices.hex

Actions: ▦ View Data ✂ Split ⬡ Build Model ⛛ Run AutoML ⚡ Predict ☁ Download ▤ Export   🗑 Delete

Rows	*Columns*	*Compressed Size*
301	9	16KB

▾ COLUMN SUMMARIES

label	*type*	*Missing*	*Zeros*	*+Inf*	*−Inf*	*min*	*max*	*mean*	*sigma*	*cardinality*	*Actions*
car_Name	enum	0	1	0	0	0	97.0	·	·	98	Convert to numeric
year	int	0	0	0	0	2003.0	2018.0	2013.6279	2.8916	·	Convert to enum
selling_price	real	0	0	0	0	0.1000	35.0	4.6613	5.0828	·	·
present_price	real	0	0	0	0	0.3200	92.6000	7.6285	8.6441	·	·
kms_driven	int	0	0	0	0	500.0	500000.0	36947.2060	38886.8839	·	Convert to enum
fuel_type	enum	0	2	0	0	0	2.0	·	·	3	Convert to numeric
seller_type	enum	0	195	0	0	0	1.0	0.3522	0.4784	2	Convert to numeric
transmission	enum	0	40	0	0	0	1.0	0.8671	0.3400	2	Convert to numeric
owner	int	0	290	0	0	0	3.0	0.0432	0.2479	·	Convert to enum

← Previous 20 Columns   → Next 20 Columns

▸ CHUNK COMPRESSION SUMMARY

▸ FRAME DISTRIBUTION SUMMARY

⊞ dataset_regression_car_prices.hex

▾ D A T A

← Previous 20 Columns   → Next 20 Columns

75	etios cross	2014.0	4.9000	8.9300	83000.0	Diesel	Dealer	Manual	0
76	etios g	2015.0	3.9500	6.8000	36000.0	Petrol	Dealer	Manual	0
77	corolla altis	2013.0	5.5000	14.6800	72000.0	Petrol	Dealer	Manual	0
78	corolla	2004.0	1.5000	12.3500	135154.0	Petrol	Dealer	Automatic	0
79	corolla altis	2010.0	5.2500	22.8300	80000.0	Petrol	Dealer	Automatic	0
80	fortuner	2012.0	14.5000	30.6100	89000.0	Diesel	Dealer	Automatic	0
81	corolla altis	2016.0	14.7300	14.8900	23000.0	Diesel	Dealer	Manual	0
82	etios gd	2015.0	4.7500	7.8500	40000.0	Diesel	Dealer	Manual	0
83	innova	2017.0	23.0	25.3900	15000.0	Diesel	Dealer	Automatic	0
84	innova	2015.0	12.5000	13.4600	38000.0	Diesel	Dealer	Manual	0
85	innova	2005.0	3.4900	13.4600	197176.0	Diesel	Dealer	Manual	0
86	camry	2006.0	2.5000	23.7300	142000.0	Petrol	Individual	Automatic	3.0
87	land cruiser	2010.0	35.0	92.6000	78000.0	Diesel	Dealer	Manual	0
88	corolla altis	2012.0	5.9000	13.7400	56000.0	Petrol	Dealer	Manual	0
89	etios liva	2013.0	3.4500	6.0500	47000.0	Petrol	Dealer	Manual	0
90	etios g	2014.0	4.7500	6.7600	40000.0	Petrol	Dealer	Manual	0
91	corolla altis	2009.0	3.8000	18.6100	62000.0	Petrol	Dealer	Manual	0
92	innova	2014.0	11.2500	16.0900	58242.0	Diesel	Dealer	Manual	0
93	innova	2005.0	3.5100	13.7000	75000.0	Petrol	Dealer	Manual	0
94	fortuner	2015.0	23.0	30.6100	40000.0	Diesel	Dealer	Automatic	0
95	corolla altis	2008.0	4.0	22.7800	89000.0	Petrol	Dealer	Automatic	0
96	corolla altis	2012.0	5.8500	18.6100	72000.0	Petrol	Dealer	Manual	0
97	innova	2016.0	20.7500	25.3900	29000.0	Diesel	Dealer	Automatic	0
98	corolla altis	2017.0	17.0	18.6400	8700.0	Petrol	Dealer	Manual	0
99	corolla altis	2013.0	7.0500	18.6100	45000.0	Petrol	Dealer	Manual	0

← Previous 20 Columns   → Next 20 Columns

✂ Split Frame

Frame: dataset_regression_car_prices.hex

Splits:

Ratio	Key	
0.80	dataset_regression_car_prices.hex_0.80	✖
0.20	dataset_regression_car_prices.hex_0.20	
Add a new split		

Seed: 43

✂ Create

⊞ Split Frames

Type	Key	Ratio
⊞	dataset_regression_car_prices.hex_0.80	0.8
⊞	dataset_regression_car_prices.hex_0.20	0.1999999999999996

⛭ Run AutoML

PARAMETERS

project_name usedcar_sales_model_1

distribution [AUTO   ⇕]

training_frame* [dataset_regression_car_prices.hex_0.80   ]

response_column* [selling_price   ⇕]

validation_frame [(Choose...)   ]

blending_frame [(Choose...)   ]

leaderboard_frame [(Choose...)   ]

A D V A N C E D

nfolds -1

balance_classes ⃝

custom_metric_func

fold_column [(Choose...) ⇅]

weights_column [(Choose...) ⇅]

ignored_columns

⃝ car_Name	ENUM(98)
⃝ year	INT
⃝ selling_price	REAL
⃝ present_price	REAL
⃝ kms_driven	INT
⃝ fuel_type	ENUM(3)
⃝ seller_type	ENUM(2)
⃝ transmission	ENUM(2)

☐ owner   INT

☑ All ☐ None

Only show columns with more than ___ % missing values.

sort_metric AUTO ⇕

exclude_algos

☐ GLM

☐ DRF

☐ GBM

☐ DeepLearning

☐ StackedEnsemble

☐ XGBoost

☑ All ☐ None

exploitation_ratio -1

monotone_constraints Choose... ⇕ ↔ ⇕ **+**

seed -1

max_models   0

max_runtime_secs   0

max_runtime_secs_per_model   0

stopping_rounds   3

stopping_metric AUTO

stopping_tolerance -1

EXPERT

keep_cross_validation_predictions ✓

keep_cross_validation_models ✓

keep_cross_validation_fold_assignment ☐

export_checkpoints_dir

⊞ Build Models

⿴ Job

Run Time	01:00:03.47
Remaining Time	00:00:00.0
Type	AutoML
Key	🔍 usedcar_sales_model_1@@selling_price
Description	AutoML build
Status	DONE
Progress	100%
	Done.
Actions	🔍 View

CS

⿴ Leaderboard

🔁 Monitor Live

▼ M O D E L S

models sorted in order of rmse, best first

	model_id	rmse	mse	mae
0	StackedEnsemble_BestOfFamily_4_AutoML_2_20260206_233833	0.879336059505872	0.7732319055473146	0.525674471419556:
1	DeepLearning_grid_1_AutoML_2_20260206_233833_model_7	0.8996110984562009	0.8093001284655723	0.591205027221134:
2	DeepLearning_grid_1_AutoML_2_20260206_233833_model_3	0.959819730670448	0.9212539153842914	0.639831752578921:

3	DeepLearning_grid_1_AutoML_2_20260206_233833_model_1	0.9702731051611252	0.9414298985990118	0.656890701688316!
4	DeepLearning_grid_1_AutoML_2_20260206_233833_model_4	0.9955894110724416	0.9911982754395711	0.621217558275087
5	DeepLearning_grid_1_AutoML_2_20260206_233833_model_11	1.0682801326000126	1.1412224417079007	0.718315607970073!
6	DeepLearning_grid_1_AutoML_2_20260206_233833_model_5	1.0920035477912355	1.1924717483886453	0.744952619523549:
7	DeepLearning_grid_1_AutoML_2_20260206_233833_model_10	1.1267243656460322	1.2695077961404535	0.748781503980318!
8	DeepLearning_grid_1_AutoML_2_20260206_233833_model_8	1.1488316401334366	1.319814137371682	0.733426296087539!
9	XGBoost_grid_1_AutoML_2_20260206_233833_model_327	1.1580299718145886	1.3410334156208967	0.577534168771620:
10	DeepLearning_grid_2_AutoML_2_20260206_233833_model_3	1.1970850273846558	1.4330125627885222	0.798692873823616
11	XGBoost_grid_1_AutoML_2_20260206_233833_model_558	1.1986733464037653	1.436817791378801	0.607782124381933:
12	XGBoost_grid_1_AutoML_2_20260206_233833_model_91	1.2245369931910943	1.499490847693486	0.621484672268813.
13	XGBoost_grid_1_AutoML_2_20260206_233833_model_723	1.2421312725059408	1.542890098137228	0.598774659077181
14	DeepLearning_grid_1_AutoML_2_20260206_233833_model_6	1.2464165152294677	1.5535541294367698	0.884762701307761!
15	XGBoost_grid_1_AutoML_2_20260206_233833_model_773	1.2569601890096829	1.5799489167552576	0.650071840624180!
16	DeepLearning_grid_1_AutoML_2_20260206_233833_model_2	1.2957636961606256	1.679003556287846	0.845292897531145!
17	XGBoost_grid_1_AutoML_2_20260206_233833_model_490	1.3094351814892955	1.7146204945219041	0.628996177479063'
18	XGBoost_grid_1_AutoML_2_20260206_233833_model_171	1.324130975686328	1.7533228407720267	0.634917025629696
19	XGBoost_grid_1_AutoML_2_20260206_233833_model_9	1.3271004206154102	1.7611955263975985	0.641329410622309.
20	XGBoost_grid_1_AutoML_2_20260206_233833_model_160	1.3301033435362588	1.7691749044863347	0.635160270376933'
21	XGBoost_grid_1_AutoML_2_20260206_233833_model_356	1.3329060803050887	1.7766386189142758	0.653259536773589!
22	StackedEnsemble_AllModels_3_AutoML_2_20260206_233833	1.3331019067295768	1.7771606937260336	0.777145140699266'

▼ E V E N T   L O G

Actions taken and discoveries made by AutoML

	timestamp	level	stage	message	name
0	23:38:33.767	INFO	Workflow	Project: usedcar_sales_model_1	
1	23:38:33.770	INFO	Validation	5-fold cross-validation will be used.	
2	23:38:33.770	INFO	Validation	User didn't set any runtime constraints (max runtime or max models), using default 1h time limit	
3	23:38:33.771	INFO	Validation	Setting stopping tolerance adaptively based on the training frame: 0.05	
4	23:38:33.771	INFO	Validation	Build control seed: -1 (random)	
				training frame: Frame key:	

5	23:38:33.779	INFO	DataImport	AutoML_2_20260206_233833_training_dataset_regression_car_prices.hex_0.80 cols: 9 rows: 239 chunks: 1 size: 15303 checksum: 3625281056691907533
6	23:38:33.779	INFO	DataImport	validation frame: NULL
7	23:38:33.780	INFO	DataImport	leaderboard frame: NULL
8	23:38:33.780	INFO	DataImport	blending frame: NULL
9	23:38:33.780	INFO	DataImport	response column: selling_price
10	23:38:33.781	INFO	DataImport	fold column: null
11	23:38:33.782	INFO	DataImport	weights column: null
12	23:38:33.785	INFO	Workflow	Loading execution steps: [{XGBoost : [def_2 (1g, 10w), def_1 (2g, 10w), def_3 (3g, 10w), grid_1 (4g, 90w), lr_search (6g, 30w)]}, {GLM : [def_1 (1g, 10w)]}, {DRF : [def_1 (2g, 10w), XRT (3g, 10w)]}, {GBM : [def_5 (1g, 10w), def_2 (2g, 10w), def_3 (2g, 10w), def_4 (2g, 10w), def_1 (3g, 10w), grid_1 (4g, 60w), lr_annealing (6g, 10w)]}, {DeepLearning : [def_1 (3g, 10w), grid_1 (4g, 30w), grid_2 (5g, 30w), grid_3 (5g, 30w)]}, {completion : [resume_best_grids (10g, 60w)]}, {StackedEnsemble : [best_of_family_1 (1g, 5w), best_of_family_2 (2g, 5w), best_of_family_3 (3g, 5w), best_of_family_4 (4g, 5w), best_of_family_5 (5g, 5w), all_2 (2g, 10w), all_3 (3g, 10w), all_4 (4g, 10w), all_5 (5g, 10w), monotonic (6g, 10w), best_of_family_gbm (6g, 10w), all_gbm (7g, 10w),

☰ Leaderboard

⟳ Monitor Live

▾ M O D E L S

models sorted in order of rmse, best first

	model_id	rmse	mse	mae
0	StackedEnsemble_BestOfFamily_4_AutoML_2_20260206_233833	0.879336059505872	0.7732319055473146	0.525674471419556:
1	DeepLearning_grid_1_AutoML_2_20260206_233833_model_7	0.8996110984562009	0.8093001284655723	0.591205027221134
2	DeepLearning_grid_1_AutoML_2_20260206_233833_model_3	0.959819730670448	0.9212539153842914	0.639831752578921:
3	DeepLearning_grid_1_AutoML_2_20260206_233833_model_1	0.9702731051611252	0.9414298985990118	0.656890701688316!

4	DeepLearning_grid_1_AutoML_2_20260206_233833_model_4	0.9955894110724416	0.9911982754395711	0.621217558275087
5	DeepLearning_grid_1_AutoML_2_20260206_233833_model_11	1.0682801326000126	1.1412224417079007	0.718315607970073
6	DeepLearning_grid_1_AutoML_2_20260206_233833_model_5	1.0920035477912355	1.1924717483886453	0.744952619523549
7	DeepLearning_grid_1_AutoML_2_20260206_233833_model_10	1.1267243656460322	1.2695077961404535	0.748781503980318
8	DeepLearning_grid_1_AutoML_2_20260206_233833_model_8	1.1488316401334366	1.319814137371682	0.733426296087539
9	XGBoost_grid_1_AutoML_2_20260206_233833_model_327	1.1580299718145886	1.3410334156208967	0.577534168771620
10	DeepLearning_grid_2_AutoML_2_20260206_233833_model_3	1.1970850273846558	1.4330125627885222	0.798692873823616
11	XGBoost_grid_1_AutoML_2_20260206_233833_model_558	1.1986733464037653	1.436817791378801	0.607782124381933
12	XGBoost_grid_1_AutoML_2_20260206_233833_model_91	1.2245369931910943	1.499490847693486	0.621484672268813
13	XGBoost_grid_1_AutoML_2_20260206_233833_model_723	1.2421312725059408	1.542890098137228	0.598774659077181
14	DeepLearning_grid_1_AutoML_2_20260206_233833_model_6	1.2464165152294677	1.5535541294367698	0.884762701307761
15	XGBoost_grid_1_AutoML_2_20260206_233833_model_773	1.2569601890096829	1.5799489167552576	0.650071840624180
16	DeepLearning_grid_1_AutoML_2_20260206_233833_model_2	1.2957636961606256	1.679003556287846	0.845292897531145
17	XGBoost_grid_1_AutoML_2_20260206_233833_model_490	1.3094351814892955	1.7146204945219041	0.628996177479063
18	XGBoost_grid_1_AutoML_2_20260206_233833_model_171	1.324130975686328	1.7533228407720267	0.634917025629696
19	XGBoost_grid_1_AutoML_2_20260206_233833_model_9	1.3271004206154102	1.7611955263975985	0.641329410622309
20	XGBoost_grid_1_AutoML_2_20260206_233833_model_160	1.3301033435362588	1.7691749044863347	0.635160270376933
21	XGBoost_grid_1_AutoML_2_20260206_233833_model_356	1.3329060803050887	1.7766386189142758	0.653259536773589
22	StackedEnsemble_AllModels_3_AutoML_2_20260206_233833	1.3331019067295768	1.7771606937260336	0.777145140699266
23	XGBoost_grid_1_AutoML_2_20260206_233833_model_309	1.3335511806590659	1.7783587514371884	0.607906561578666

▼ E V E N T   L O G

Actions taken and discoveries made by AutoML

	timestamp	level	stage	message	name
0	23:38:33.767	INFO	Workflow	Project: usedcar_sales_model_1	
1	23:38:33.770	INFO	Validation	5-fold cross-validation will be used.	
2	23:38:33.770	INFO	Validation	User didn't set any runtime constraints (max runtime or max models), using default 1h time limit	
3	23:38:33.771	INFO	Validation	Setting stopping tolerance adaptively based on the training frame: 0.05	
4	23:38:33.771	INFO	Validation	Build control seed: -1 (random)	
5	23:38:33.779	INFO	DataImport	training frame: Frame key: AutoML_2_20260206_233833_training_dataset_regression_car_prices.hex_0.80	

				cols: 9 rows: 239 chunks: 1 size: 15303 checksum: 3625281056691907533
6	23:38:33.779	INFO	DataImport	validation frame: NULL
7	23:38:33.780	INFO	DataImport	leaderboard frame: NULL
8	23:38:33.780	INFO	DataImport	blending frame: NULL
9	23:38:33.780	INFO	DataImport	response column: selling_price
10	23:38:33.781	INFO	DataImport	fold column: null
11	23:38:33.782	INFO	DataImport	weights column: null
12	23:38:33.785	INFO	Workflow	Loading execution steps: [{XGBoost : [def_2 (1g, 10w), def_1 (2g, 10w), def_3 (3g, 10w), grid_1 (4g, 90w), lr_search (6g, 30w)]}, {GLM : [def_1 (1g, 10w)]}, {DRF : [def_1 (2g, 10w), XRT (3g, 10w)]}, {GBM : [def_5 (1g, 10w), def_2 (2g, 10w), def_3 (2g, 10w), def_4 (2g, 10w), def_1 (3g, 10w), grid_1 (4g, 60w), lr_annealing (6g, 10w)]}, {DeepLearning : [def_1 (3g, 10w), grid_1 (4g, 30w), grid_2 (5g, 30w), grid_3 (5g, 30w)]}, {completion : [resume_best_grids (10g, 60w)]}, {StackedEnsemble : [best_of_family_1 (1g, 5w), best_of_family_2 (2g, 5w), best_of_family_3 (3g, 5w), best_of_family_4 (4g, 5w), best_of_family_5 (5g, 5w), all_2 (2g, 10w), all_3 (3g, 10w), all_4 (4g, 10w), all_5 (5g, 10w), monotonic (6g, 10w), best_of_family_gbm (6g, 10w), all_gbm (7g, 10w), best_of_family_xglm (8g, 10w), all_xglm (8g, 10w), best_of_family (10g,