Chapter 1: Understanding H2O AutoML Basics

```
Checking whether there is an H2O instance running at http://localhost:54321 ..... not found.
Attempting to start a local H2O server...
; Java HotSpot(TM) 64-Bit Server VM (build 17.0.2+8-LTS-86, mixed mode, sharing)
  Starting server from C:\Users\ajgao\AppData\Local\Programs\Python\Python310\Lib\site-packages\h2o\backend\bin\h2o.jar
  Ice root: C:\Users\ajgao\AppData\Local\Temp\tmpmswmzso9
  JVM stdout: C:\Users\ajgao\AppData\Local\Temp\tmpmswmzso9\h2o_ajgao_started_from_python.out
  JVM stderr: C:\Users\ajgao\AppData\Local\Temp\tmpmswmzso9\h2o_ajgao_started_from_python.err
  Server is running at http://127.0.0.1:54321
Connecting to H2O server at http://127.0.0.1:54321 ... successful.
--------------------------  ------------------------------
H2O_cluster_uptime:         01 secs
H2O_cluster_timezone:       Europe/Dublin
H2O_data_parsing_timezone:  UTC
H2O_cluster_version:        3.36.1.2
H2O_cluster_version_age:    6 days
H2O_cluster_name:           H2O_from_python_ajgao_etoqk4
H2O_cluster_total_nodes:    1
H2O_cluster_free_memory:    3.963 Gb
H2O_cluster_total_cores:    12
H2O_cluster_allowed_cores:  12
H2O_cluster_status:         locked, healthy
H2O_connection_url:         http://127.0.0.1:54321
H2O_connection_proxy:       {"http": null, "https": null}
H2O_internal_security:      False
Python_version:             3.10.2 final
--------------------------  ------------------------------
>>>
```

```
H2O is not running yet, starting it now...

Note:  In case of errors look at the following log files:
    C:\Users\ajgao\AppData\Local\Temp\RtmpGOgnlL\filee50164e6416/h2o_ajgao_started_from_r.out
    C:\Users\ajgao\AppData\Local\Temp\RtmpGOgnlL\filee50756a39d3/h2o_ajgao_started_from_r.err

java version "17.0.2" 2022-01-18 LTS
Java(TM) SE Runtime Environment (build 17.0.2+8-LTS-86)
Java HotSpot(TM) 64-Bit Server VM (build 17.0.2+8-LTS-86, mixed mode, sharing)

Starting H2O JVM and connecting:  Connection successful!

R is connected to the H2O cluster:
    H2O cluster uptime:         1 seconds 849 milliseconds
    H2O cluster timezone:       Europe/Dublin
    H2O data parsing timezone:  UTC
    H2O cluster version:        3.36.1.2
    H2O cluster version age:    6 days
    H2O cluster name:           H2O_started_from_R_ajgao_kcp486
    H2O cluster total nodes:    1
    H2O cluster total memory:   3.96 GB
    H2O cluster total cores:    12
    H2O cluster allowed cores:  12
    H2O cluster healthy:        TRUE
    H2O Connection ip:          localhost
    H2O Connection port:        54321
    H2O Connection proxy:       NA
    H2O Internal Security:      FALSE
    R Version:                  R version 3.6.1 (2019-07-05)

>
```

C1	C2	C3	C4	C5
5.1	3.5	1.4	0.2	Iris-setosa
4.9	3	1.4	0.2	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa
4.6	3.1	1.5	0.2	Iris-setosa
5	3.6	1.4	0.2	Iris-setosa
5.4	3.9	1.7	0.4	Iris-setosa
4.6	3.4	1.4	0.3	Iris-setosa
5	3.4	1.5	0.2	Iris-setosa
4.4	2.9	1.4	0.2	Iris-setosa
4.9	3.1	1.5	0.1	Iris-setosa

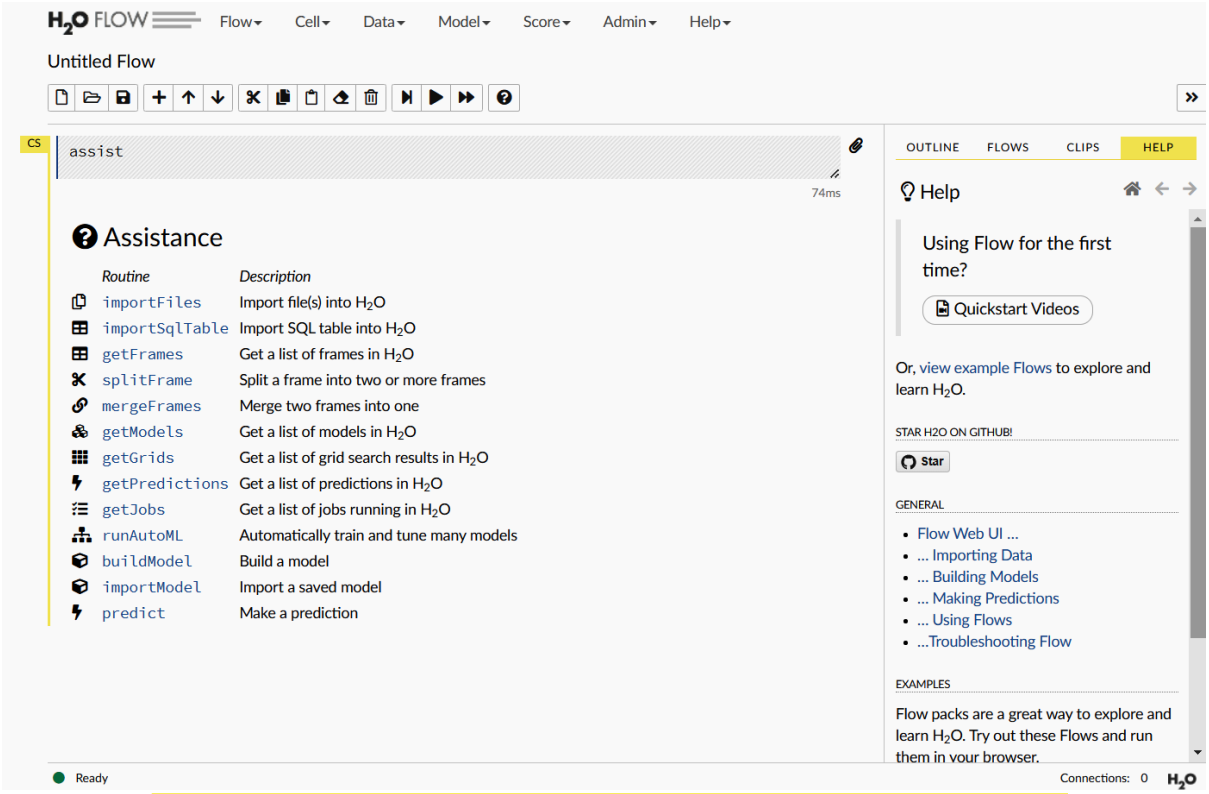
model_id	mean_per_class_error	logloss	rmse	mse
GLM_1_AutoML_1_20211221_224844	0.0254274	0.0730056	0.148617	0.0220871
StackedEnsemble_BestOfFamily_1_AutoML_1_20211221_224844	0.0254274	0.0889165	0.155702	0.0242432
StackedEnsemble_BestOfFamily_3_AutoML_1_20211221_224844	0.034188	0.234761	0.244628	0.0598429
StackedEnsemble_AllModels_4_AutoML_1_20211221_224844	0.0418803	0.212898	0.207104	0.042892
GBM_5_AutoML_1_20211221_224844	0.0423077	0.153568	0.196082	0.038448
XGBoost_3_AutoML_1_20211221_224844	0.0423077	0.171005	0.205224	0.0421171
StackedEnsemble_BestOfFamily_5_AutoML_1_20211221_224844	0.0423077	0.272447	0.208633	0.0435279
StackedEnsemble_AllModels_1_AutoML_1_20211221_224844	0.0425214	0.239748	0.247898	0.0614534
StackedEnsemble_BestOfFamily_2_AutoML_1_20211221_224844	0.0425214	0.242794	0.250567	0.0627836
GBM_2_AutoML_1_20211221_224844	0.0508547	0.164525	0.20544	0.0422055
XRT_1_AutoML_1_20211221_224844	0.0508547	0.154602	0.196747	0.0387094
DRF_1_AutoML_1_20211221_224844	0.0508547	0.155568	0.201598	0.0406418
XGBoost_2_AutoML_1_20211221_224844	0.0508547	0.230482	0.231617	0.0536463
GBM_4_AutoML_1_20211221_224844	0.0508547	0.159493	0.201367	0.0405486
StackedEnsemble_BestOfFamily_6_AutoML_1_20211221_224844	0.0508547	0.140577	0.198154	0.0392648
GBM_3_AutoML_1_20211221_224844	0.0508547	0.158446	0.203925	0.0415853
StackedEnsemble_AllModels_5_AutoML_1_20211221_224844	0.0508547	0.13417	0.197749	0.0391048
StackedEnsemble_AllModels_2_AutoML_1_20211221_224844	0.0515304	0.244011	0.252697	0.0638558
StackedEnsemble_AllModels_3_AutoML_1_20211221_224844	0.059188	0.183036	0.221313	0.0489796
StackedEnsemble_BestOfFamily_4_AutoML_1_20211221_224844	0.059188	0.246594	0.239717	0.0574643
XGBoost_1_AutoML_1_20211221_224844	0.0925214	0.478273	0.385523	0.148628

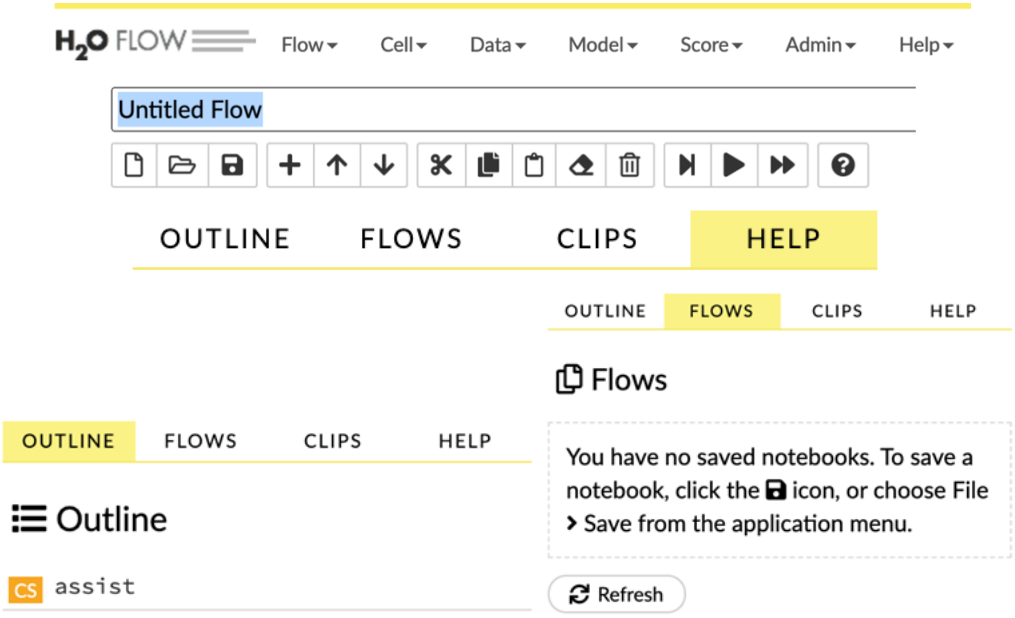
predict	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	0.996763	0.0029518	0.000284888
Iris-setosa	0.999722	0.000171882	0.000106306
Iris-setosa	0.99952	0.000345017	0.0001354
Iris-setosa	0.999739	0.000157519	0.000103133
Iris-setosa	0.999975	2.05119e-06	2.28322e-05
Iris-setosa	0.999801	0.000108085	9.04886e-05
Iris-setosa	0.999452	0.000405189	0.00014317
Iris-setosa	0.999515	0.000349113	0.000135955
Iris-setosa	0.999335	0.000509734	0.000155041
Iris-setosa	0.999627	0.000251481	0.000121322

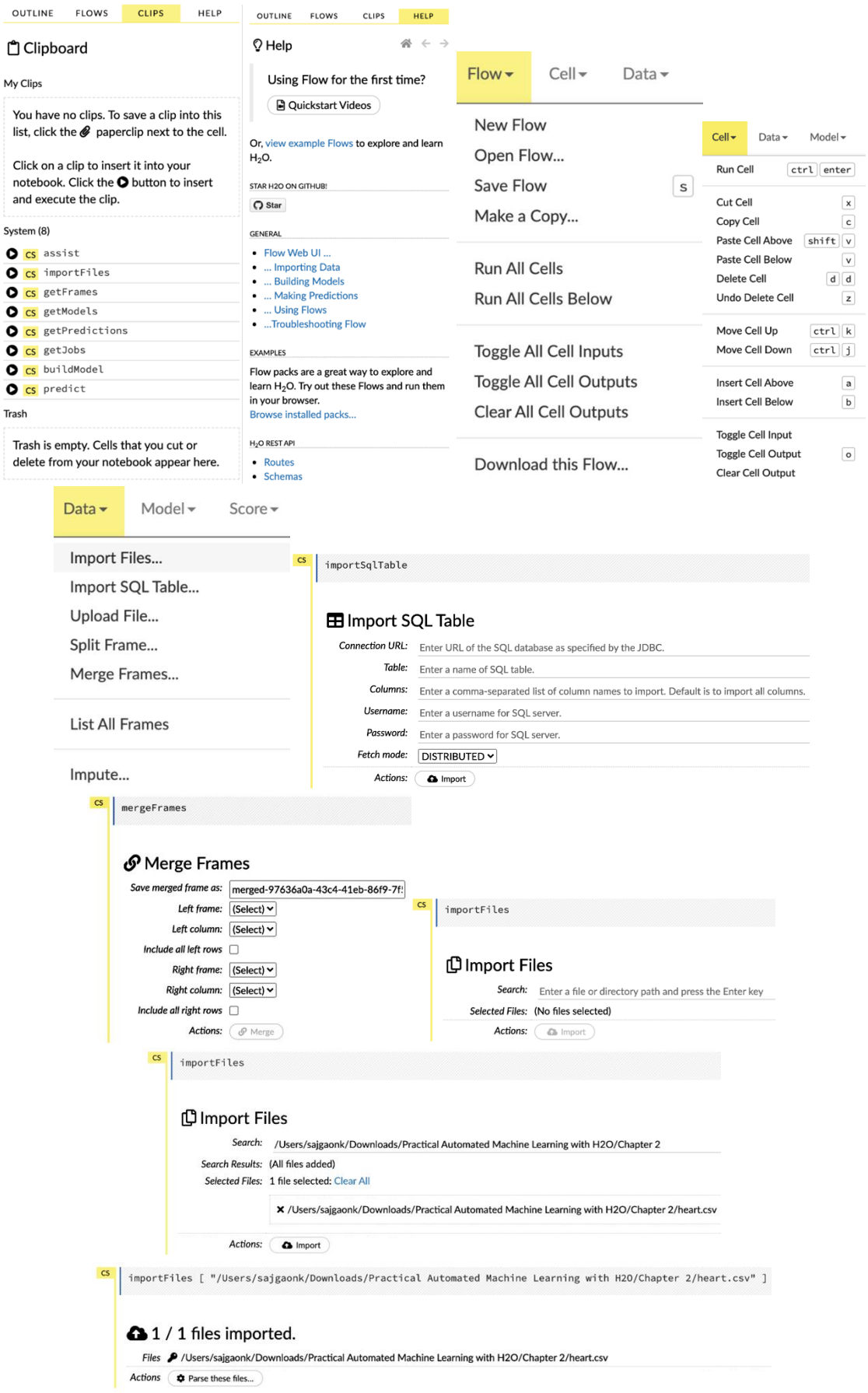
model_id	mean_per_class_error	logloss	rmse	mse
GBM_3_AutoML_8_20211222_02555	0.02503053	0.1441918	0.1790187	0.03204771
GLM_1_AutoML_8_20211222_02555	0.02503053	0.06051945	0.1367546	0.01870183
StackedEnsemble_AllModels_5_AutoML_8_20211222_02555	0.02503053	0.08676523	0.1584221	0.02509755
GBM_4_AutoML_8_20211222_02555	0.03296703	0.15864028	0.1871248	0.03501568
StackedEnsemble_BestOfFamily_2_AutoML_8_20211222_02555	0.03455433	0.22094574	0.2311208	0.05341682
StackedEnsemble_BestOfFamily_3_AutoML_8_20211222_02555	0.04090354	0.21032039	0.2249462	0.05060078
StackedEnsemble_BestOfFamily_4_AutoML_8_20211222_02555	0.04151404	0.15865856	0.1955133	0.03822546
StackedEnsemble_AllModels_4_AutoML_8_20211222_02555	0.04151404	0.18315141	0.1858142	0.03452691
StackedEnsemble_AllModels_3_AutoML_8_20211222_02555	0.04151404	0.17914264	0.2003599	0.0401441
StackedEnsemble_AllModels_1_AutoML_8_20211222_02555	0.04310134	0.2171082	0.2293443	0.05259883
StackedEnsemble_BestOfFamily_6_AutoML_8_20211222_02555	0.04310134	0.09428996	0.1672787	0.02798217
XRT_1_AutoML_8_20211222_02555	0.04884005	0.12547862	0.1850836	0.03425593
GBM_2_AutoML_8_20211222_02555	0.04884005	0.16726969	0.1934784	0.03743388
XGBoost_3_AutoML_8_20211222_02555	0.04884005	0.16601239	0.1983851	0.03935665
XGBoost_2_AutoML_8_20211222_02555	0.04884005	0.22718554	0.2309562	0.05334075
StackedEnsemble_BestOfFamily_5_AutoML_8_20211222_02555	0.04945055	0.26499193	0.221083	0.04887771
StackedEnsemble_BestOfFamily_1_AutoML_8_20211222_02555	0.05103785	0.10792744	0.187905	0.03530827
StackedEnsemble_AllModels_2_AutoML_8_20211222_02555	0.05103785	0.21228471	0.2263655	0.05124134
GBM_5_AutoML_8_20211222_02555	0.05738706	0.13549309	0.1926783	0.03712494
DRF_1_AutoML_8_20211222_02555	0.05738706	0.1279311	0.1848138	0.03415613
XGBoost_1_AutoML_8_20211222_02555	0.10989011	0.46102681	0.3760385	0.14140496

predict	Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	0.9876256	0.005647775	0.006726586
Iris-setosa	0.9869614	0.00706752	0.00597108
Iris-setosa	0.9888558	0.004439304	0.006704931
Iris-setosa	0.9874139	0.005875627	0.006710506
Iris-setosa	0.9873642	0.005833785	0.006802058
Iris-setosa	0.9876381	0.005642109	0.006719837

Chapter 2: Working with H2O Flow (H2O's Web UI)






OUTLINE FLOWS CLIPS HELP
Clipboard
My Clips
You have no clips. To save a clip into this list, click the paperclip next to the cell.
Click on a clip to insert it into your notebook. Click the button to insert and execute the clip.
System (8)
assist
importFiles
getFrames
getModels
getPredictions
getJobs
buildModel
predict
Trash
Trash is empty. Cells that you cut or delete from your notebook appear here.
OUTLINE FLOWS CLIPS HELP
Help
Using Flow for the first time?
Quickstart Videos
Or, view example Flows to explore and learn H2O.
STAR H2O ON GITHUB!
Star
GENERAL
Flow Web UI ...
... Importing Data
... Building Models
... Making Predictions
... Using Flows
...Troubleshooting Flow
EXAMPLES
Flow packs are a great way to explore and learn H2O. Try out these Flows and run them in your browser.
Browse installed packs...
H2O REST API
Routes
Schemas
Flow Cell Data
New Flow
Open Flow...
Save Flow s
Make a Copy...
Run All Cells
Run All Cells Below
Toggle All Cell Inputs
Toggle All Cell Outputs
Clear All Cell Outputs
Download this Flow...
Cell Data Model
Run Cell ctrl enter
Cut Cell x
Copy Cell c
Paste Cell Above shift v
Paste Cell Below v
Delete Cell d d
Undo Delete Cell z
Move Cell Up ctrl k
Move Cell Down ctrl j
Insert Cell Above a
Insert Cell Below b
Toggle Cell Input
Toggle Cell Output o
Clear Cell Output
Data Model Score
Import Files...
Import SQL Table...
Upload File...
Split Frame...
Merge Frames...
List All Frames
Impute...
importSqlTable
Import SQL Table
Connection URL: Enter URL of the SQL database as specified by the JDBC.
Table: Enter a name of SQL table.
Columns: Enter a comma-separated list of column names to import. Default is to import all columns.
Username: Enter a username for SQL server.
Password: Enter a password for SQL server.
Fetch mode: DISTRIBUTED
Actions: Import
mergeFrames
Merge Frames
Save merged frame as: merged-97636a0a-43c4-41eb-86f9-7f
Left frame: (Select)
Left column: (Select)
Include all left rows
Right frame: (Select)
Right column: (Select)
Include all right rows
Actions: Merge
importFiles
Import Files
Search: Enter a file or directory path and press the Enter key
Selected Files: (No files selected)
Actions: Import
importFiles
Import Files
Search: /Users/sajgaonk/Downloads/Practical Automated Machine Learning with H2O/Chapter 2
Search Results: (All files added)
Selected Files: 1 file selected: Clear All
/Users/sajgaonk/Downloads/Practical Automated Machine Learning with H2O/Chapter 2/heart.csv
Actions: Import
importFiles ["/Users/sajgaonk/Downloads/Practical Automated Machine Learning with H2O/Chapter 2/heart.csv"]
1 / 1 files imported.
Files /Users/sajgaonk/Downloads/Practical Automated Machine Learning with H2O/Chapter 2/heart.csv
Actions Parse these files...

CS

```
setupParse source_frames: [ "nfs://Users/sajgaonk/Downloads/Practical Automated Machine Learning with H2O/Chapter 2/heart.csv" ]
```

293ms

Setup Parse

PARSE CONFIGURATION

Sources: nfs://Users/sajgaonk/Downloads/Practical Automated Machine Learning with H2O/Chapter 2/heart.csv
ID: heart.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	Age	Numeric	40	49	37	48	54	39	45	54	37
2	Sex	Enum	M	F	M	F	M	M	F	M	M
3	ChestPainType	Enum	ATA	NAP	ATA	ASY	NAP	NAP	ATA	ATA	ASY
4	RestingBP	Numeric	140	160	130	138	150	120	130	110	140
5	Cholesterol	Numeric	289	180	283	214	195	339	237	208	207
6	FastingBS	Numeric	0	0	0	0	0	0	0	0	0
7	RestingECG	Enum	Normal	Normal	ST	Normal	Normal	Normal	Normal	Normal	Normal
8	MaxHR	Numeric	172	156	98	108	122	170	170	142	130
9	ExerciseAngina	Enum	N	N	N	Y	N	N	N	N	Y
10	Oldpeak	Numeric	0	1	0	1.5	0	0	0	0	1.5
11	ST_Slope	Enum	Up	Flat	Up	Flat	Up	Up	Up	Up	Flat
12	HeartDisease	Numeric	0	1	0	1	0	0	0	0	1

← Previous page → Next page

Parse

CS

```
parseFiles
  source_frames: ["nfs://Users/sajgaonk/Downloads/Practical Automated Machine Learning with H2O/Chapter 2/heart.csv"]
  destination_frame: "heart.hex"
  parse_type: "CSV"
  separator: 44
  number_columns: 12
  single_quotes: false
  column_names:
["Age","Sex","ChestPainType","RestingBP","Cholesterol","FastingBS","RestingECG","MaxHR","ExerciseAngina","Oldpeak","ST_Slope","HeartDisease"]
  column_types: ["Numeric","Enum","Enum","Numeric","Numeric","Enum","Enum","Numeric","Enum","Numeric","Enum","Enum"]
  delete_on_done: true
  check_header: 1
  chunk_size: 4194304
  escapechar: undefined
```

1.1s

Job

Run Time: 00:00:00.183
Remaining Time: 00:00:00.0
Type: Frame
Key: heart.hex
Description: Parse
Status: DONE
Progress: 100%

Done.

Actions: View

```
getFrameSummary "heart.hex"
```

134ms

heart.hex

Actions: View Data | Split | Build Model | Run AutoML | Predict | Download | Export | Delete

Rows	Columns	Compressed Size
918	12	15KB

▾ COLUMN SUMMARIES

label	type	Missing	Zeros	+Inf	-Inf	min	max	mean	sigma	cardinality	Actions
Age	int	0	0	0	0	28.0	77.0	53.5109	9.4326	·	Convert to enum
Sex	enum	0	193	0	0	0	1.0	0.7898	0.4077	2	Convert to numeric
ChestPainType	enum	0	496	0	0	0	3.0	·	·	4	Convert to numeric
RestingBP	int	0	1	0	0	0	200.0	132.3965	18.5142	·	Convert to enum
Cholesterol	int	0	172	0	0	0	603.0	198.7996	109.3841	·	Convert to enum
FastingBS	enum	0	704	0	0	0	1.0	0.2331	0.4230	2	Convert to numeric
RestingECG	enum	0	188	0	0	0	2.0	·	·	3	Convert to numeric
MaxHR	int	0	0	0	0	60.0	202.0	136.8094	25.4603	·	Convert to enum
ExerciseAngina	enum	0	547	0	0	0	1.0	0.4041	0.4910	2	Convert to numeric
Oldpeak	real	0	368	0	0	-2.6000	6.2000	0.8874	1.0666	·	·
ST_Slope	enum	0	63	0	0	0	2.0	·	·	3	Convert to numeric
HeartDisease	enum	0	410	0	0	0	1.0	0.5534	0.4974	2	Convert to numeric

← Previous 20 Columns | → Next 20 Columns

▸ CHUNK COMPRESSION SUMMARY

▸ FRAME DISTRIBUTION SUMMARY

▾ CHUNK COMPRESSION SUMMARY

chunk_type	chunk_name	count	count_percentage	size	size_percentage
CBS	Binary	4	33.3333	740 B	7.7390
C1N	1-Byte Integers (w/o NAs)	6	50.0	5.8 KB	61.8699
C1S	1-Byte Fractions	1	8.3333	1002 B	10.4790
C2	2-Byte Integers	1	8.3333	1.9 KB	19.9122

▾ FRAME DISTRIBUTION SUMMARY

	size	number_of_rows	number_of_chunks_per_column	number_of_chunks
192.168.0.157:54321	9.3 KB	918.0	1.0	12.0
mean	9.3 KB	918.0	1.0	12.0
min	9.3 KB	918.0	1.0	12.0
max	9.3 KB	918.0	1.0	12.0
stddev	0 B	0	0	0
total	9.3 KB	918.0	1.0	12.0

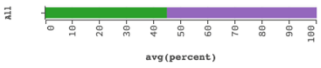
```
getColumnSummary "heart.hex", "HeartDisease"
```

131ms

Summary: HeartDisease

Actions: Impute | Inspect

CHARACTERISTICS



DOMAIN (MAX 1000 LEVELS)

```
getFrameData "heart.hex"
```

81ms

heart.hex

DATA

← Previous 20 Columns → Next 20 Columns

Row	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
1	40.0	M	ATA	140.0	289.0	0	Normal	172.0	N	0	Up	0
2	49.0	F	NAP	160.0	180.0	0	Normal	156.0	N	1.0	Flat	1
3	37.0	M	ATA	130.0	283.0	0	ST	98.0	N	0	Up	0
4	48.0	F	ASY	138.0	214.0	0	Normal	108.0	Y	1.5000	Flat	1
5	54.0	M	NAP	150.0	195.0	0	Normal	122.0	N	0	Up	0
6	39.0	M	NAP	120.0	339.0	0	Normal	170.0	N	0	Up	0
7	45.0	F	ATA	130.0	237.0	0	Normal	170.0	N	0	Up	0
8	54.0	M	ATA	110.0	208.0	0	Normal	142.0	N	0	Up	0
9	37.0	M	ASY	140.0	207.0	0	Normal	130.0	Y	1.5000	Flat	1
10	48.0	F	ATA	120.0	284.0	0	Normal	120.0	N	0	Up	0
11	37.0	F	NAP	130.0	211.0	0	Normal	142.0	N	0	Up	0
12	58.0	M	ATA	136.0	164.0	0	ST	99.0	Y	2.0	Flat	1
13	39.0	M	ATA	120.0	204.0	0	Normal	145.0	N	0	Up	0
14	49.0	M	ASY	140.0	234.0	0	Normal	140.0	Y	1.0	Flat	1
15	42.0	F	NAP	115.0	211.0	0	ST	137.0	N	0	Up	0
16	54.0	F	ATA	120.0	273.0	0	Normal	150.0	N	1.5000	Flat	0
17	38.0	M	ASY	110.0	196.0	0	Normal	166.0	N	0	Flat	1
18	43.0	F	ATA	120.0	201.0	0	Normal	165.0	N	0	Up	0
19	60.0	M	ASY	100.0	248.0	0	Normal	125.0	N	1.0	Flat	1
20	36.0	M	ATA	120.0	267.0	0	Normal	160.0	N	3.0	Flat	1
21	43.0	F	TA	100.0	223.0	0	Normal	142.0	N	0	Up	0
22	44.0	M	ATA	120.0	184.0	0	Normal	142.0	N	1.0	Flat	0
23	49.0	F	ATA	124.0	201.0	0	Normal	164.0	N	0	Up	0

← Previous 20 Columns → Next 20 Columns

```
assist splitFrame, "heart.hex"
```

21ms

Split Frame

Frame: heart.hex

Splits:	*Ratio*	*Key*
	0.70	training_dataframe
	0.20	testing_dataframe
	0.10	prediction_dataframe

Add a new split

Seed: 5

Create

Model ▾ Score ▾ Admin ▾ Help ▾

- Run AutoML...
- Aggregator...
- ANOVA for Generalized Linear Model...
- Cox Proportional Hazards...
- Deep Learning...
- Distributed Random Forest...
- Extended Isolation Forest...
- Gradient Boosting Machine...
- Generalized Linear Modeling...
- Generalized Low Rank Modeling...
- Isolation Forest...
- K-means...
- Maximum R Square Improvement (MAXR) to GLM...
- Naive Bayes...
- Principal Components Analysis...
- RuleFit...
- Stacked Ensemble...
- TargetEncoder...
- Word2Vec...
- XGBoost...
- Import MOJO Model
- List All Models
- List Grid Search Results
- Import Model...
- Export Model...

```
splitFrame "heart.hex", [0.7,0.2], ["training_dataframe","testing_dataframe","prediction_dataframe"], 5
```

94ms

Split Frames

Type	*Key*	*Ratio*
	training_dataframe	0.7
	testing_dataframe	0.2
	prediction_dataframe	0.10000000000000009

CS

```
assist runAutoML, training_frame: "heart.hex"
```

18ms

Run AutoML

PARAMETERS

*training_frame**	training_dataset	ID of the training data frame.
*response_column**	HeartDisease	Response column
validation_frame	testing_dataset	ID of the validation data frame (used for early stopping in grid searches and for early stopping of the AutoML process itself).
blending_frame	(Choose...)	ID of the H2OFrame used to train the the metalearning algorithm in Stacked Ensembles (instead of relying on cross-validated predicted values). When provided, it is also recommended to disable cross validation by setting `nfolds=0` and to provide a leaderboard frame for scoring purposes.
leaderboard_frame	(Choose...)	ID of the leaderboard data frame (used to score models and rank them on the AutoML Leaderboard).
project_name		Optional project name used to group models from multiple AutoML runs into a single Leaderboard; derived from the training data name if not specified.

ADVANCED

fold_column	(Choose...)	Fold column (contains fold IDs) in the training frame. These assignments are used to create the folds for cross-validation of the models.
weights_column	(Choose...)	Weights column in the training frame, which specifies the row weights used in model training.
ignored_columns	Search...	Names of columns to ignore in the training frame when building models.

Showing page 1 of 2. -24 ignored.

☐ Age	INT
☐ Sex	ENUM(2)
☐ ChestPainType	ENUM(4)
☐ RestingBP	INT
☐ Cholesterol	INT
☐ FastingBS	ENUM(2)
☐ RestingECG	ENUM(3)
☐ MaxHR	INT
☐ ExerciseAngina	ENUM(2)
☐ Oldpeak	REAL

☑ All ☐ None ← Previous 10 → Next 10

Only show columns with more than 0 % missing values.

sort_metric	AUTO	Metric used to sort leaderboard
nfolds	0	Number of folds for k-fold cross-validation (defaults to -1 (AUTO), otherwise it must be >=2 or use 0 to disable). Disabling prevents Stacked Ensembles from being built.
balance_classes	☐	Balance training data class counts via over/under-sampling (for imbalanced data).
exclude_algos	Search...	A list of algorithms to skip during the model-building phase.

- ☐ GLM
- ☐ DRF
- ☐ GBM
- ☐ DeepLearning
- ☐ StackedEnsemble
- ☐ XGBoost

☑ All ☐ None

exploitation_ratio	-1	The budget ratio (between 0 and 1) dedicated to the exploitation (vs exploration) phase.
monotone_constraints	Choose... ↔ +	A mapping representing monotonic constraints. Use +1 to enforce an increasing constraint and -1 to specify a decreasing constraint.
seed	5	Seed for random number generator; set to a value other than -1 for reproducibility.
max_models	10	Maximum number of models to build (optional).
max_runtime_secs	0	This argument specifies the maximum time that the AutoML process will run for. If neither max_runtime_secs nor max_models are specified by the user, then max_runtime_secs defaults to 3600 seconds (1 hour).
max_runtime_secs_per_model	0	Maximum time to spend on each individual model (optional).
stopping_rounds	3	Early stopping based on convergence of stopping_metric. Stop if simple moving average of length k of the stopping_metric does not improve for k:=stopping_rounds scoring events (0 to disable)
stopping_metric	AUTO	Metric to use for early stopping (AUTO: logloss for classification, deviance for regression)
stopping_tolerance	-1	Relative tolerance for metric-based stopping criterion (stop if relative improvement is not at least this much)

EXPERT

keep_cross_validation_predictions	☐	Whether to keep the predictions of the cross-validation predictions. This needs to be set to TRUE if running the same AutoML object for repeated runs because CV predictions are required to build additional Stacked Ensemble models in AutoML.
keep_cross_validation_models	☐	Whether to keep the cross-validated models. Keeping cross-validation models may consume significantly more memory in the H2O cluster.
keep_cross_validation_fold_assignment	☐	Whether to keep cross-validation assignments.
export_checkpoints_dir		Path to a directory where every generated model will be stored.

Build Models

```
runAutoML {"input_spec":
{"training_frame":"training_dataset","response_column":"HeartDisease","validation_frame":"testing_dataset","ignored_columns":
[],"sort_metric":"AUTO"},"build_control":{"nfolds":0,"balance_classes":false,"stopping_criteria":
{"seed":5,"max_models":10,"max_runtime_secs":0,"max_runtime_secs_per_model":0,"stopping_rounds":3,"stopping_metric":"AUTO","stopping_tolerance":-1},
"keep_cross_validation_predictions":false,"keep_cross_validation_models":false,"keep_cross_validation_fold_assignment":false},"build_models":
{"exclude_algos":[],"exploitation_ratio":-1,"monotone_constraints":[]}}, 'exec'
```

11.6s

Job

Run Time 00:00:10.158
Remaining Time 00:00:00.0
Type Auto Model
Key AutoML_1_20220104_223006@@HeartDisease
Description AutoML build
Status DONE
Progress 100%
Done.
Actions View

CS

```
getLeaderboard "AutoML_1_20220104_223006@@HeartDisease"
```

96ms

Leaderboard

Monitor Live

▾ MODELS

models sorted in order of auc, best first

	model_id	auc	logloss	aucpr	mean_per_class_error	rmse	mse
0	GBM_1_AutoML_1_20220104_223006	0.9369965477560414	0.3150888280397835	0.9526004839213847	0.10241657077100115	0.3052195935953662	0.09315900031452051
1	XGBoost_1_AutoML_1_20220104_223006	0.9365937859608745	0.31864288641084404	0.9510206216535636	0.0996547756041427	0.30389543850542716	0.09235243754440586
2	GLM_1_AutoML_1_20220104_223006	0.9350978135788263	0.321121239651523	0.9506935050645732	0.11962025316455696	0.31032593617677556	0.09630218666399219
3	GBM_4_AutoML_1_20220104_223006	0.933831990794016	0.33075852888909035	0.9531086456216924	0.13130034522439588	0.3203588891066754	0.10262981782966317
4	GBM_2_AutoML_1_20220104_223006	0.9315304948216341	0.32554761066460214	0.9417502787295773	0.1267548906789413	0.3115577763007514	0.09706824797346902
5	XRT_1_AutoML_1_20220104_223006	0.9283659378596087	0.34626540184045396	0.9410194226539295	0.10420023014959724	0.3205028494065838 6	0.10272207647773937
6	GBM_3_AutoML_1_20220104_223006	0.9253164556962025	0.3441538204067555	0.9366759429162057	0.13941311852704258	0.32372789720827333	0.10479975143089038
7	DRF_1_AutoML_1_20220104_223006	0.9222669735327963	0.341285000994188	0.937780404872303	0.1277330264672037	0.32437991383954023	0.10522232850254756
8	XGBoost_3_AutoML_1_20220104_223006	0.9181818181818181	0.3643481155776031	0.9336605482778707	0.14395857307249713	0.33199390950590124	0.11021995594901256
9	XGBoost_2_AutoML_1_20220104_223006	0.9127733026467204	0.3719410194756317 4	0.9214390408636409	0.14039125431530497	0.3322859827538344	0.11041397433468156

▸ EVENT LOG

CS

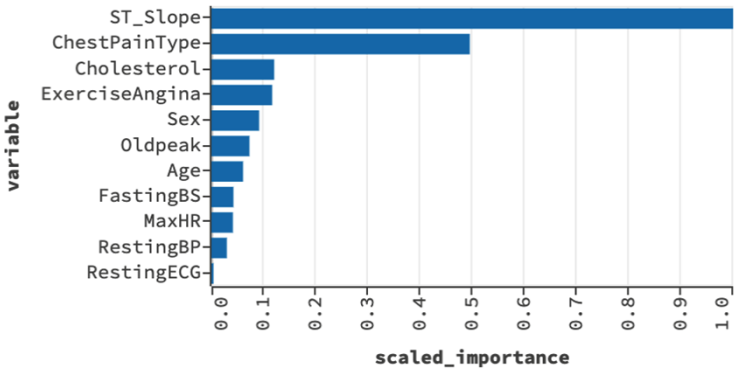
```
getModel "GBM_1_AutoML_1_20220104_223006"
```

575ms

Model

Model ID: GBM_1_AutoML_1_20220104_223006
Algorithm: Gradient Boosting Machine
Actions: Refresh | Predict... | Download POJO | Download Model Deployment Package (MOJO) | Export | Inspect | Delete | Download Gen Model

▾ VARIABLE IMPORTANCES



▾ OUTPUT

original_names	·
cross_validation_models	·
cross_validation_predictions	·
cross_validation_holdout_predictions_frame_id	·
cross_validation_fold_assignment_frame_id	·
model_category	Binomial
cv_scoring_history	·
cross_validation_metrics	·
cross_validation_metrics_summary	·
status	·
start_time	1641335408430
end_time	1641335408849
run_time	419
default_threshold	0.538636
init_f	0.184334

▾ COLUMN_TYPES

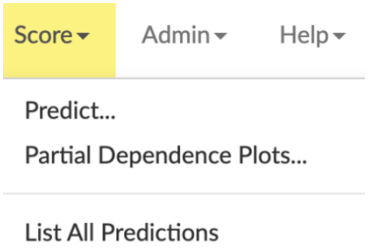
column_types
Numeric
Enum
Enum
Numeric
Numeric
Enum
Enum
Numeric
Enum
Numeric
Enum
Enum

▾ OUTPUT - TRAINING_METRICS

model	GBM_1_AutoML_1_20220104_223006
model_checksum	1667669939565867160
frame	AutoML_1_20220104_223006_training_training_dataset
frame_checksum	480731267713440174
description	·
model_category	Binomial
scoring_time	1641335408839
predictions	·
MSE	0.086861
RMSE	0.294722
nobs	729
custom_metric_name	·
custom_metric_value	0
r2	0.649596
logloss	0.294230
AUC	0.947350
pr_auc	0.949081
Gini	0.894700
mean_per_class_error	0.108940

▾ OUTPUT - VALIDATION_METRICS

model	GBM_1_AutoML_1_20220104_223006
model_checksum	1667669939565867160
frame	testing_dataset
frame_checksum	5628356784682958122
description	·
model_category	Binomial
scoring_time	1641335408842
predictions	·
MSE	0.093159
RMSE	0.305220
nobs	189
custom_metric_name	·
custom_metric_value	0
r2	0.617062
logloss	0.315089
AUC	0.936997
pr_auc	0.952600
Gini	0.873993
mean_per_class_error	0.102417



```
predict model: "GBM_1_AutoML_1_20220104_223006"
```

Predict

Name: prediction-8c7f82d6-f7

Model: GBM_1_AutoML_1_20220104_223006

Frame: testing_dataset

Compute Leaf Node Assignment:

Actions: Predict

```
predict model: "GBM_1_AutoML_1_20220104_223006", frame: "testing_dataset", predictions_frame: "prediction-8c7f82d6-f774-4c9a-9e25-37d408edd167"
```

456ms

Prediction

Actions: Inspect

- ROC CURVE
- PREDICTION

Combine predictions with frame

- DOMAIN
- PREDICTION - CONFUSION MATRIX ROW LABELS: ACTUAL CLASS; COLUMN LABELS: PREDICTED CLASS
- PREDICTION - METRICS FOR THRESHOLDS
- PREDICTION - MAXIMUM METRICS
- PREDICTION - GAINS/LIFT TABLE

PREDICTION

model	GBM_1_AutoML_1_20220104_223006
model_checksum	1667669939565867160
frame	testing_dataset
frame_checksum	5628356784682958122
description	·
model_category	Binomial
scoring_time	1641336595124
predictions	prediction-8c7f82d6-f774-4c9a-9e25-37d408edd167
MSE	0.093159
RMSE	0.305220
nobs	189
custom_metric_name	·
custom_metric_value	0
r2	0.617062
logloss	0.315089
AUC	0.936997
pr_auc	0.952600
Gini	0.873993
mean_per_class_error	0.102417

Combine predictions with frame

CS

```
getFrameSummary "prediction-8c7f82d6-f774-4c9a-9e25-37d408edd167"
```

114ms

prediction-8c7f82d6-f774-4c9a-9e25-37d408edd167

Actions: View Data | Split | Build Model | Run AutoML | Predict | Download | Export | Delete

Rows	Columns	Compressed Size
189	3	4KB

COLUMN SUMMARIES

label	type	Missing	Zeros	+Inf	-Inf	min	max	mean	sigma	cardinality	Actions
predict	enum	0	80	0	0	0	1.0	0.5767	0.4954	2	Convert to numeric
p0	real	0	0	0	0	0.8135	0.9779	0.4298	0.3598	·	·
p1	real	0	0	0	0	0.0221	0.9865	0.5710	0.3598	·	·

← Previous 20 Columns | → Next 20 Columns

- CHUNK COMPRESSION SUMMARY
- FRAME DISTRIBUTION SUMMARY

```
getFrameData "prediction-8c7f82d6-f774-4c9a-9e25-37d408edd167"
```

prediction-8c7f82d6-f774-4c9a-9e25-37d408edd167

▾ DATA

← Previous 20 Columns → Next 20 Columns

Row	predict	p0	p1
1	0	0.9138	0.0862
2	1	0.3062	0.6938
3	1	0.1320	0.8680
4	0	0.9404	0.0596
5	0	0.9589	0.0411
6	1	0.2106	0.7894
7	1	0.1018	0.8982
8	0	0.9369	0.0631
9	1	0.0344	0.9656
10	0	0.7541	0.2459
11	1	0.2028	0.7972
12	0	0.9469	0.0531
13	1	0.4461	0.5539
14	0	0.8049	0.1951
15	1	0.2738	0.7262
16	0	0.7119	0.2881
17	1	0.0266	0.9734
18	1	0.3805	0.6195
19	1	0.1011	0.8989
20	0	0.9710	0.0290
21	0	0.7275	0.2725
22	1	0.0254	0.9746
23	0	0.7764	0.2236

← Previous 20 Columns → Next 20 Columns

```
bindFrames "combined-prediction-8c7f82d6-f774-4c9a-9e25-37d408edd167", [ "prediction-8c7f82d6-f774-4c9a-9e25-37d408edd167", "testing_dataset" ]
```

Frames Combined

The specified frames were combined successfully.

View Frame

```
getFrameSummary "combined-prediction-8c7f82d6-f774-4c9a-9e25-37d408edd167"
```

168ms

combined-prediction-8c7f82d6-f774-4c9a-9e25-37d408edd167

Actions: View Data | Split | Build Model | Run AutoML | Predict | Download | Export | Delete

Rows	*Columns*	*Compressed Size*
189	15	12KB

▸ COLUMN SUMMARIES

▸ CHUNK COMPRESSION SUMMARY

▸ FRAME DISTRIBUTION SUMMARY

```
getFrameData "combined-prediction-8c7f82d6-f774-4c9a-9e25-37d408edd167"
```

combined-prediction-8c7f82d6-f774-4c9a-9e25-37d408edd167

▾ DATA

← Previous 20 Columns → Next 20 Columns

Row	predict	p0	p1	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
1	0	0.9138	0.0862	39.0	M	NAP	120.0	339.0	0	Normal	170.0	N	0	Up	0
2	1	0.3062	0.6938	38.0	M	ASY	110.0	196.0	0	Normal	166.0	N	0	Flat	1
3	1	0.1320	0.8680	41.0	M	ASY	130.0	172.0	0	ST	130.0	N	2.0	Flat	1
4	0	0.9404	0.0596	32.0	M	ATA	125.0	254.0	0	Normal	155.0	N	0	Up	0
5	0	0.9589	0.0411	50.0	M	ATA	140.0	216.0	0	Normal	170.0	N	0	Up	0
6	1	0.2106	0.7894	47.0	F	ASY	120.0	205.0	0	Normal	98.0	Y	2.0	Flat	1
7	1	0.1018	0.8982	52.0	M	ASY	112.0	342.0	0	ST	96.0	Y	1.0	Flat	1
8	0	0.9369	0.0631	49.0	M	ATA	100.0	253.0	0	Normal	174.0	N	0	Up	0
9	1	0.0344	0.9656	52.0	M	ASY	160.0	246.0	0	ST	82.0	Y	4.0	Flat	1
10	0	0.7541	0.2459	44.0	M	ASY	150.0	412.0	0	Normal	170.0	N	0	Up	0
11	1	0.2028	0.7972	63.0	M	ASY	150.0	223.0	0	Normal	115.0	N	0	Flat	1
12	0	0.9469	0.0531	52.0	M	ATA	160.0	196.0	0	Normal	165.0	N	0	Up	0
13	1	0.4461	0.5539	43.0	M	TA	120.0	291.0	0	ST	155.0	N	0	Flat	1
14	0	0.8049	0.1951	39.0	M	ASY	130.0	307.0	0	Normal	140.0	N	0	Up	0
15	1	0.2738	0.7262	46.0	M	ASY	118.0	186.0	0	Normal	124.0	N	0	Flat	1
16	0	0.7119	0.2881	50.0	M	ASY	140.0	129.0	0	Normal	135.0	N	0	Up	0
17	1	0.0266	0.9734	57.0	M	ASY	150.0	255.0	0	Normal	92.0	Y	3.0	Flat	1
18	1	0.3805	0.6195	33.0	F	ASY	100.0	246.0	0	Normal	150.0	Y	1.0	Flat	1
19	1	0.1011	0.8989	59.0	F	ASY	130.0	338.0	1	ST	130.0	Y	1.5000	Flat	1
20	0	0.9710	0.0290	34.0	F	ATA	130.0	161.0	0	Normal	190.0	N	0	Up	0
21	0	0.7275	0.2725	48.0	F	ASY	108.0	163.0	0	Normal	175.0	N	2.0	Up	0
22	1	0.0254	0.9746	56.0	M	ASY	170.0	388.0	0	ST	122.0	Y	2.0	Flat	1
23	0	0.7764	0.2236	39.0	M	ASY	110.0	273.0	0	Normal	132.0	N	0	Up	0

← Previous 20 Columns → Next 20 Columns

Chapter 3: Understanding Data Processing

important_column_1	important_column_2	important_column_3	important_column_4	important_column_5
-1.08563	0.997345	0.282978	-1.50629	-0.5786
1.65144	-2.42668	-0.428913	1.26594	-0.86674
-0.678886	-0.094709	1.49139	-0.638902	-0.443982
-0.434351	2.20593	2.18679	1.00405	0.386186
0.737369	1.49073	-0.935834	1.17583	-1.25388
-0.637752	0.907105	-1.42868	-0.140069	-0.861755
-0.255619	-2.79859	-1.77153	-0.699877	0.927462
-0.173636	0.00284592	0.688223	-0.879536	0.283627
-0.805367	-1.72767	-0.3909	0.573806	0.338589
-0.0118305	2.39237	0.412912	0.978736	2.23814

important_column_6	important_column_7
1.03973	-0.403366
-0.12603	-0.837517
-1.60596	1.25524
-0.688869	1.66095
0.807308	-0.314758
-1.0859	-0.732462
-1.21252	2.08711
0.164441	1.15021
-1.26735	0.181035
1.17786	-0.335011

important_column_1	important_column_2	important_column_3	important_column_4	important_column_5	important_column_6	important_column_7
-1.08563	0.997345	0.282978	-1.50629	-0.5786	1.03973	-0.403366
1.65144	-2.42668	-0.428913	1.26594	-0.86674	-0.12603	-0.837517
-0.678886	-0.094709	1.49139	-0.638902	-0.443982	-1.60596	1.25524
-0.434351	2.20593	2.18679	1.00405	0.386186	-0.688869	1.66095
0.737369	1.49073	-0.935834	1.17583	-1.25388	0.807308	-0.314758
-0.637752	0.907105	-1.42868	-0.140069	-0.861755	-1.0859	-0.732462
-0.255619	-2.79859	-1.77153	-0.699877	0.927462	-1.21252	2.08711
-0.173636	0.00284592	0.688223	-0.879536	0.283627	0.164441	1.15021
-0.805367	-1.72767	-0.3909	0.573806	0.338589	-1.26735	0.181035
-0.0118305	2.39237	0.412912	0.978736	2.23814	1.17786	-0.335011

important_column_1	important_column_2	important_column_3	important_column_4	important_column_5
-1.08563	0.997345	0.282978	-1.50629	-0.5786
1.65144	-2.42668	-0.428913	1.26594	-0.86674
-0.678886	-0.094709	1.49139	-0.638902	-0.443982
-0.434351	2.20593	2.18679	1.00405	0.386186
0.737369	1.49073	-0.935834	1.17583	-1.25388
-0.637752	0.907105	-1.42868	-0.140069	-0.861755
-0.255619	-2.79859	-1.77153	-0.699877	0.927462
-0.173636	0.00284592	0.688223	-0.879536	0.283627
-0.805367	-1.72767	-0.3909	0.573806	0.338589
-0.0118305	2.39237	0.412912	0.978736	2.23814

words	numerical_representation	letters
Hello	0	a
World	1	b
Welcome	2	c
To	3	d
Machine	4	
Learning	5	

other_words	numerical_representation	letters
How	0	a
Are	1	b
You	2	c
Doing	3	d
Today	4	e
My	5	
Friend	6	
Learning	7	
H2O	8	
Artificial	9	
Intelligence	nan	

numerical_representation	letters	other_words	words
0	a	How	Hello
1	b	Are	World
2	c	You	Welcome
3	d	Doing	To
5		My	Learning

numerical_representation	letters	other_words	words
0	a	How	Hello
1	b	Are	World
2	c	You	Welcome
3	d	Doing	To
4	e	Today	
5		My	Learning
6		Friend	
7		Learning	
8		Artificial	
9		Intelligence	

C1	C2	C3
nan	nan	77
94	14	-58
94	-26	-39
96	93	-56
-85	44	-53
58	-28	27
-84	63	92
78	-65	94
nan	81	nan
-27	-61	13

C1	C2	C3
nan	nan	77
94	14	-58
94	-26	-39
96	93	-56
-85	44	-53
58	-28	27
-84	63	92
78	-65	94
78	81	94
-27	-61	13

C1	C2	C3
-99	-18	75
-73	-10	58
-33	nan	70
21	nan	-61
-52	20	79
-22	-98	19
nan	-58	52
-58	-31	-9
79	-26	nan
12	-81	66

C1	C2	C3
-99	-18	75
-73	-10	58
-33	nan	70
21	99	-61
-52	20	79
-22	-98	19
nan	-58	52
-58	-31	-9
79	-26	nan
12	-81	66

C1	C2	C3
-99	-18	225
-73	-10	174
-33	nan	210
21	99	-183
-52	20	237
-22	-98	57
nan	-58	156
-58	-31	-27
79	-26	nan
12	-81	198

C1	C2	C3	C1	C2	C3
0	0	225	0	0	225
0	0	174	0	0	174
0	nan	210	0	0	210
21	99	-183	21	99	183
0	20	237	0	20	237
0	0	57	0	0	57
nan	0	156	0	0	156
0	0	-27	0	0	27
79	0	nan	79	0	0
12	0	198	12	0	198

	age	weight	max_speed	100_meter_time
0	13	46	16.755530	NaN
1	15	33	15.089844	21.732242
2	13	32	14.584233	22.348161
3	16	45	15.669721	NaN
4	13	39	19.711957	20.169496
...	...	...	...	...
95	14	46	13.231185	NaN
96	16	45	17.572064	23.064310
97	17	38	14.639427	23.142284
98	17	32	15.984808	23.537361
99	14	48	16.072193	20.370838

100 rows × 4 columns

age	weight	max_speed	100_meter_time
13	41	17.9565	23.5558
16	45	13.6646	23.0656
15	44	18.1247	24.7465
15	39	14.621	23.7011
13	45	15.2694	25.8552
16	43	16.9983	24.6807
15	37	13.3174	24.7037
13	40	13.3024	21.8998
14	47	17.4593	26.3073
15	47	18.4853	22.5947

age	weight	max_speed	100_meter_time
13	41	17.9565	22.9804
16	45	13.6646	23.0656
15	44	18.1247	24.7465
15	39	14.621	23.7011
13	45	15.2694	25.8552
16	43	16.9983	24.6807
15	37	13.3174	24.7037
13	40	13.3024	21.8998
14	47	17.4593	26.3073
15	47	18.4853	22.5947

C1	C2	C3
0	6	12
1	8	9
2	2	4
3	1	15
3	5	14
3	3	13
3	3	3
8	7	8
8	6	7
8	6	2

C1	C2	C3
0	6	12
1	8	9
2	2	4
3	1	15
3	5	14
3	3	13
3	3	3
8	7	8
8	6	7
8	6	2

C1	C2	C3
0	6	12
1	8	9
2	2	4
3	1	15
3	3	13
3	3	3
3	5	14
8	6	7
8	6	2
8	6	1

C1	C2	C3
0	6	12
1	8	9
2	2	4
3	5	14
3	3	13
3	3	3
3	1	15
8	7	8
8	6	7
8	6	2

C1	C2	C3
3	1	15
3	5	14
3	3	13
0	6	12
12	8	11
13	6	10
1	8	9
8	7	8
8	6	7
14	5	6

C1
Today we learn AI
Tomorrow AI learns us
Today and Tomorrow are same
Us and AI are same

C1
Today
we
learn
AI
Tomorrow
AI
learns
us

Animals	Mythical
Dragon	0
Unicorn	1
Horse	1
Lizard	1
Goblin	0
Dragon	0
Horse	0
Horse	1
Unicorn	0
Dragon	1
Goblin	0
Lizard	1
Lizard	1
Unicorn	0
Dragon	1

Animals_target_encoded	Target 0 count	Target 1 count
Dragon	2	2
Unicorn	2	1
Horse	1	2
Lizard	0	3
Goblin	2	0

Animals_target_encoded	Target 0 count	Target 1 count	Probability of Target 1 Occurring
Dragon	2	2	0.50
Unicorn	2	1	0.33
Horse	1	2	0.66
Lizard	0	3	1
Goblin	2	0	0

Animals	Mythical
0.50	0
0.33	1
0.66	1
1	0
0	1
0.50	0
0.66	0
0.66	1
0.33	0
0.50	1
0	0
1	1
1	1
0.33	0
0.50	1

Parse progress: (done) 100%

symboling	normalized-losses	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location	wheel-base	length	width	height	curb-weight	engine-type	num-of-cylinders	engine-size	fuel systen
3	nan	alfa-romero	gas	std	two	convertible	rwd	front	88.6	168.8	64.1	48.8	2548	dohc	four	130	mpf
3	nan	alfa-romero	gas	std	two	convertible	rwd	front	88.6	168.8	64.1	48.8	2548	dohc	four	130	mpf
1	nan	alfa-romero	gas	std	two	hatchback	rwd	front	94.5	171.2	65.5	52.4	2823	ohcv	six	152	mpf
2	164	audi	gas	std	four	sedan	fwd	front	99.8	176.6	66.2	54.3	2337	ohc	four	109	mpf
2	164	audi	gas	std	four	sedan	4wd	front	99.4	176.6	66.4	54.3	2824	ohc	five	136	mpf
2	nan	audi	gas	std	two	sedan	fwd	front	99.8	177.3	66.3	53.1	2507	ohc	five	136	mpf
1	158	audi	gas	std	four	sedan	fwd	front	105.8	192.7	71.4	55.7	2844	ohc	five	136	mpf
1	nan	audi	gas	std	four	wagon	fwd	front	105.8	192.7	71.4	55.7	2954	ohc	five	136	mpf
1	158	audi	gas	turbo	four	sedan	fwd	front	105.8	192.7	71.4	55.9	3086	ohc	five	131	mpf
0	nan	audi	gas	turbo	two	hatchback	4wd	front	99.5	178.2	67.9	52	3053	ohc	five	131	mpf

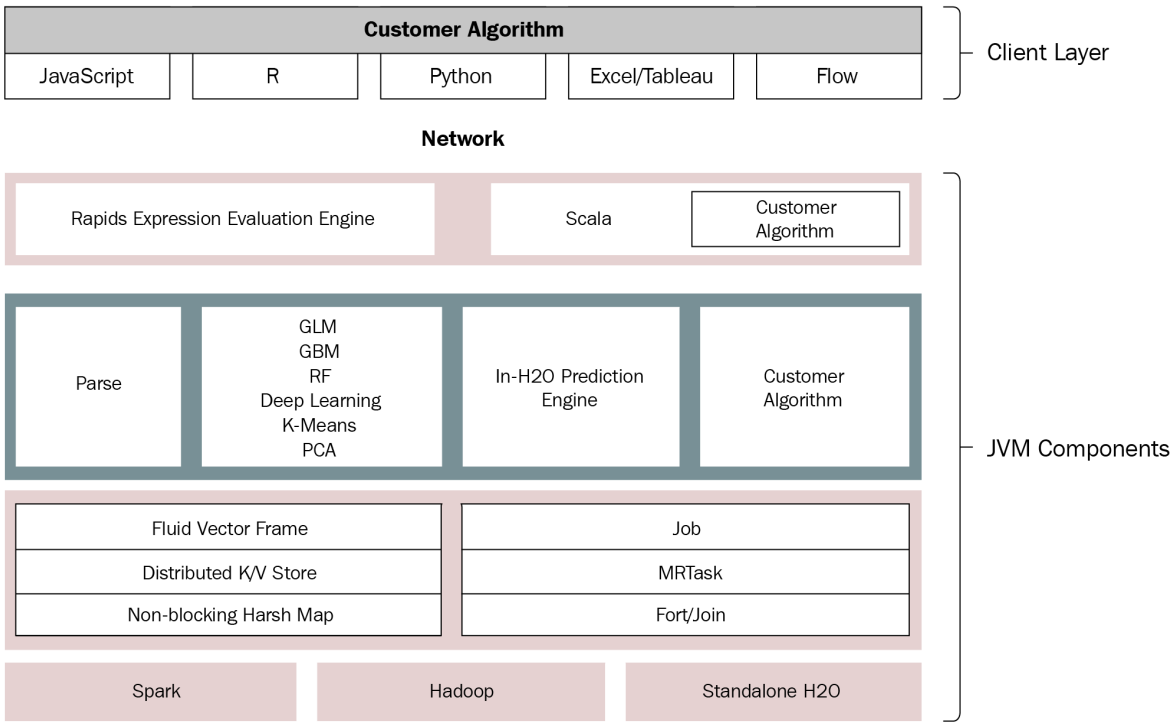
make	fuel-type	body-style	price
alfa-romero	gas	convertible	13495
alfa-romero	gas	convertible	16500
alfa-romero	gas	hatchback	16500
audi	gas	sedan	13950
audi	gas	sedan	17450
audi	gas	sedan	15250
audi	gas	sedan	17710
audi	gas	wagon	18920
audi	gas	sedan	23875
audi	gas	hatchback	nan

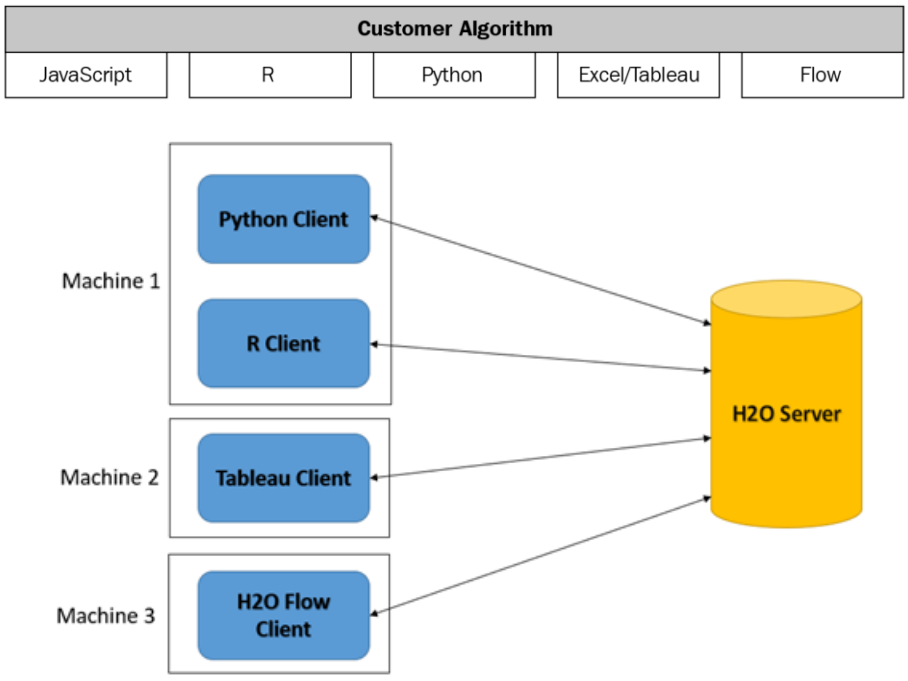
```
targetencoder Model Build progress: |████████████████████████████████████████| (done) 100%
Model Details
=============
H2OTargetEncoderEstimator :  TargetEncoder
Model Key:  TargetEncoder_model_python_1653850929913_2


Target Encoder model summary: Summary for target encoder model
```

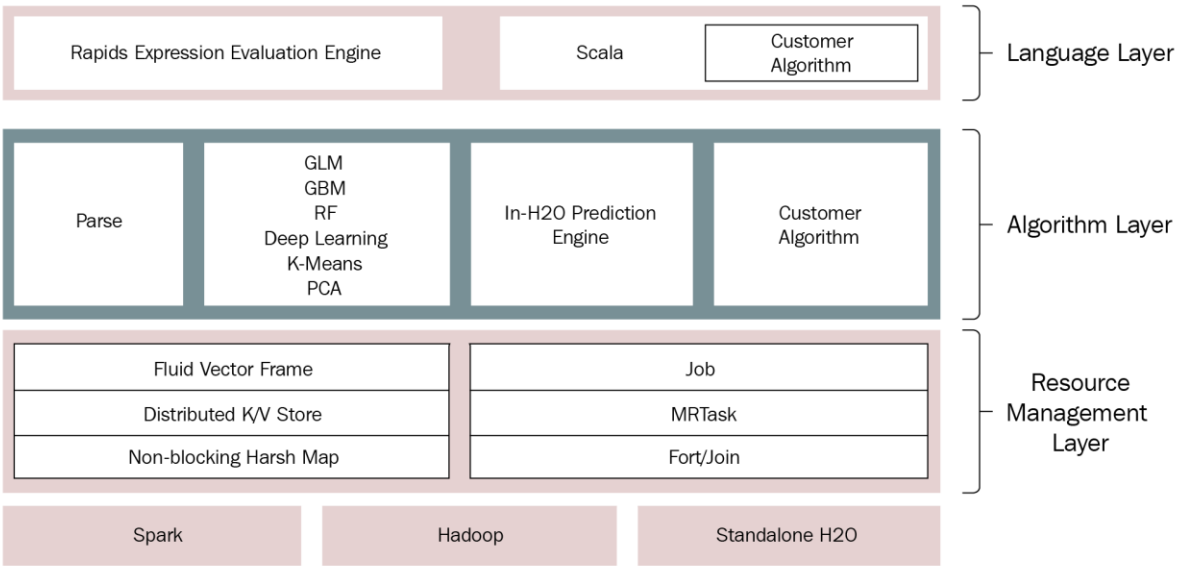
	original_names	encoded_column_names
0	make	make_te
1	fuel-type	fuel-type_te
2	body-style	body-style_te

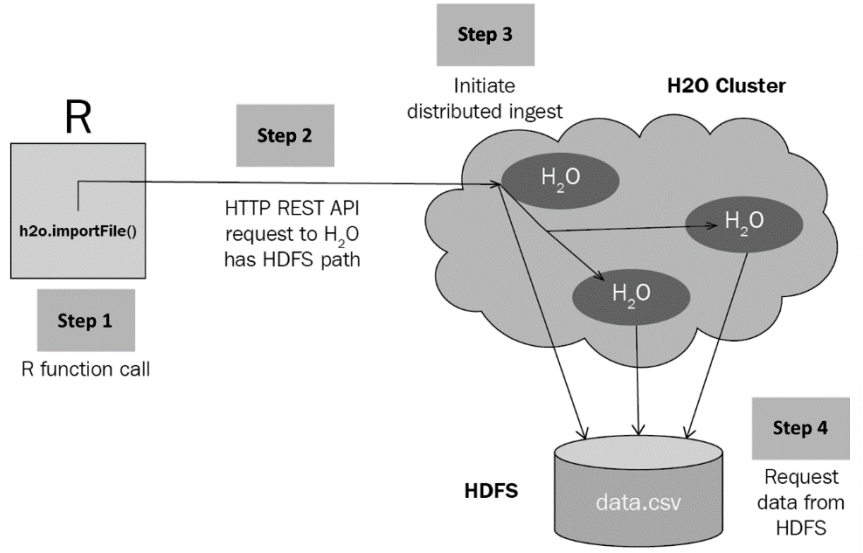
make_te	fuel-type_te	body-style_te	make	fuel-type	body-style	price
15498.3	13074	21890.5	alfa-romero	gas	convertible	13495
15498.3	13074.1	21890.5	alfa-romero	gas	convertible	16500
15498.3	13074.1	9722.24	alfa-romero	gas	hatchback	16500
16656	13074.1	14734.1	audi	gas	sedan	13950
16656	13074.1	14734.1	audi	gas	sedan	17450
16656	13074.1	14734.1	audi	gas	sedan	15250
16656	13074.1	14734.1	audi	gas	sedan	17710
16656	13074.1	11809.9	audi	gas	wagon	18920
26340.7	13074.1	14734.1	bmw	gas	sedan	16430
26340.7	13074.1	14734.1	bmw	gas	sedan	16925
16656	13074.1	14734.1	audi	gas	sedan	23875
16656	13074.1	9722.23	audi	gas	hatchback	nan
26340.7	13074.1	14734.1	bmw	gas	sedan	24565
7790.12	13074.1	14734.1	dodge	gas	sedan	8558
8151	13074.1	9722.23	honda	gas	hatchback	7895
8151	13074.1	14734.1	honda	gas	sedan	8845
11048	13074.1	14734.1	isuzu	gas	sedan	6785
34125	13074.1	14734.1	jaguar	gas	sedan	35550
11260.3	13074.1	9722.23	mazda	gas	hatchback	6795
11260.3	13074.1	9722.23	mazda	gas	hatchback	8845

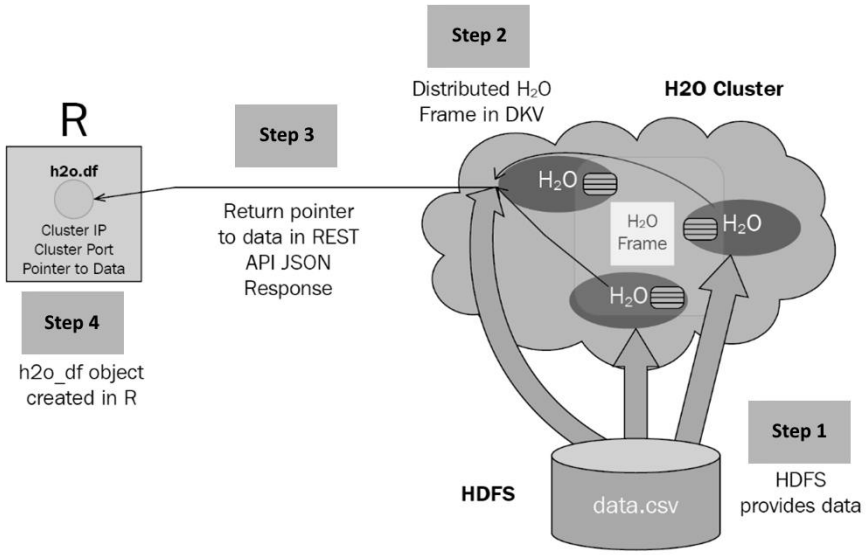
Chapter 4: Understanding H2O AutoML Training and Architecture

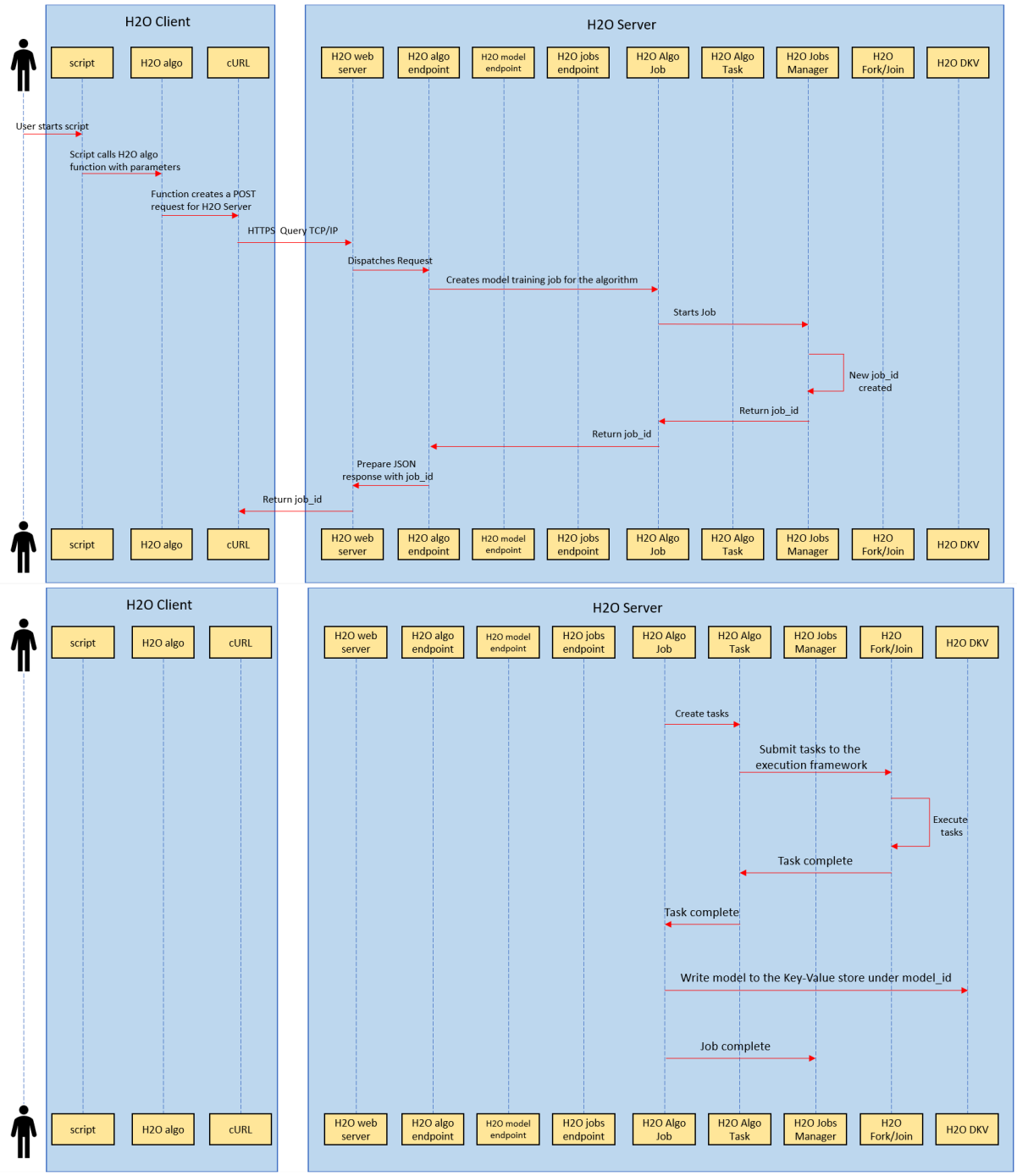


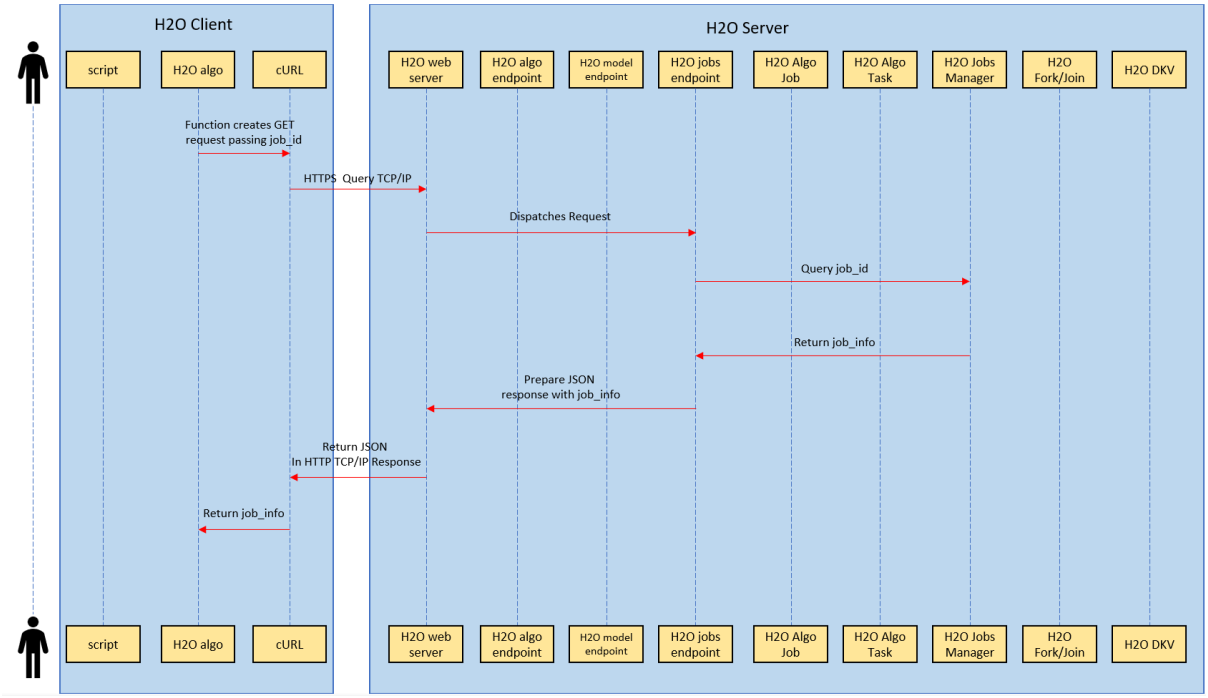


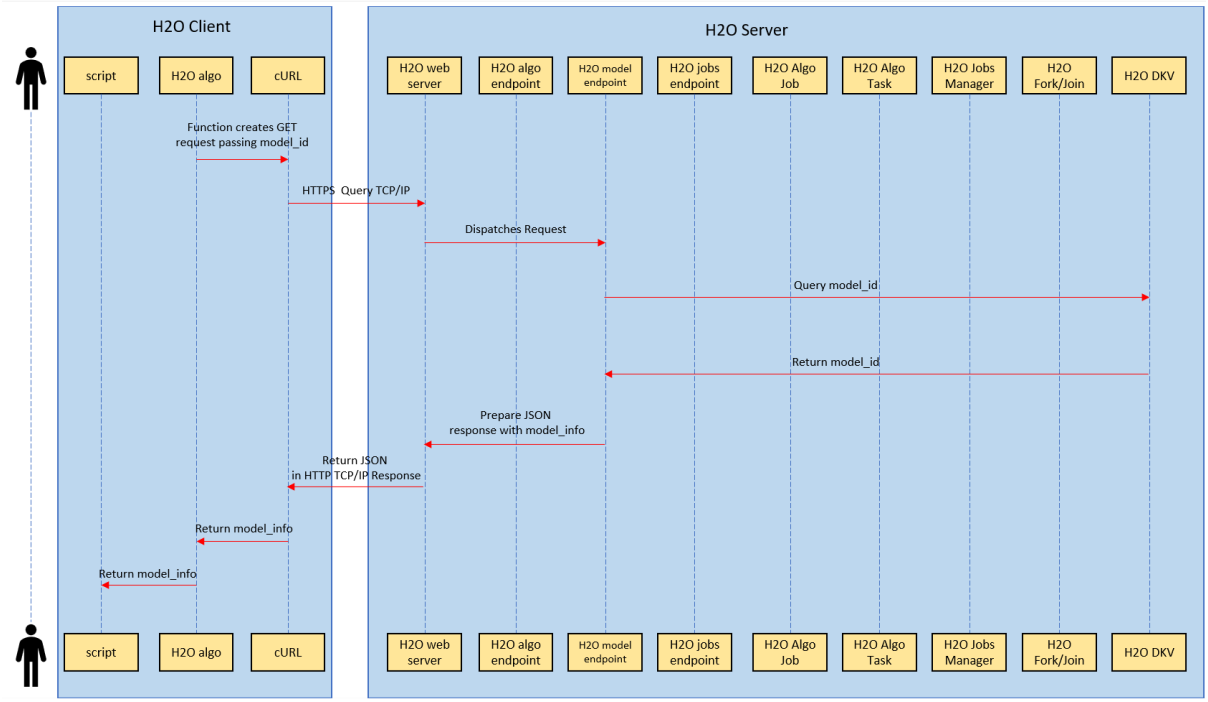

Rapids Expression Evaluation Engine
Scala
Customer Algorithm
Language Layer
Parse
GLM
GBM
RF
Deep Learning
K-Means
PCA
In-H2O Prediction Engine
Customer Algorithm
Algorithm Layer
Fluid Vector Frame
Distributed K/V Store
Non-blocking Harsh Map
Job
MRTask
Fort/Join
Resource Management Layer
Spark
Hadoop
Standalone H2O

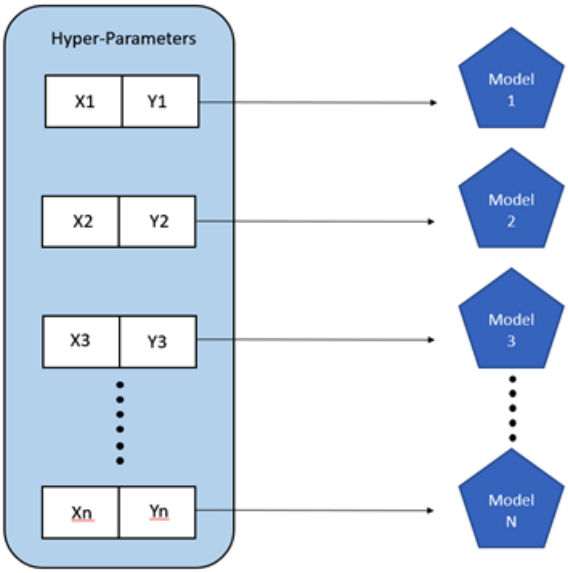

Step 3
Initiate distributed ingest
H2O Cluster
R
Step 2
h2o.importFile()
HTTP REST API request to H2O has HDFS path
H2O
H2O
H2O
Step 1
R function call
Step 4
Request data from HDFS
HDFS
data.csv

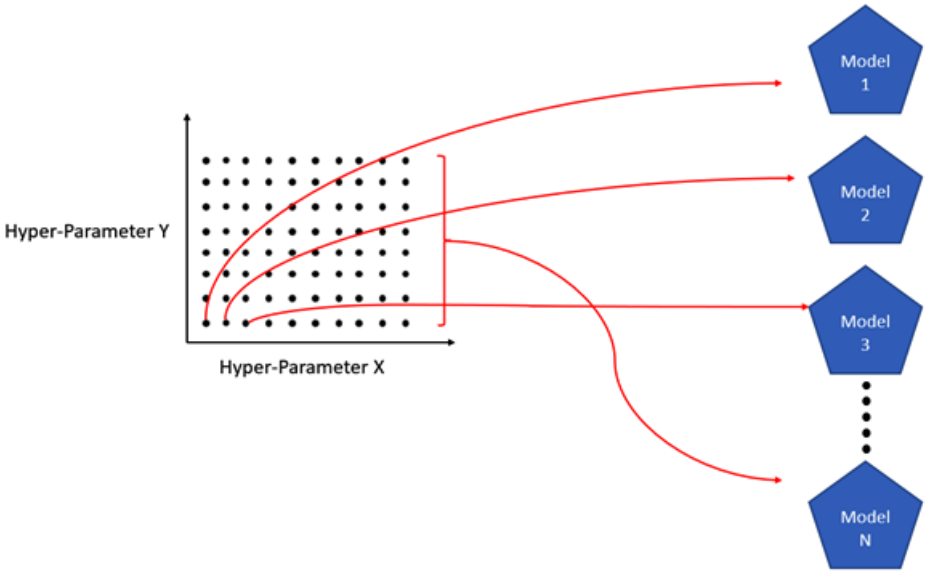

Step 2
Distributed H2O Frame in DKV
H2O Cluster
R
Step 3
h2o.df
Cluster IP
Cluster Port
Pointer to Data
Return pointer to data in REST API JSON Response
H2O
H2O Frame
H2O
H2O
Step 4
h2o_df object created in R
Step 1
HDFS provides data
HDFS
data.csv

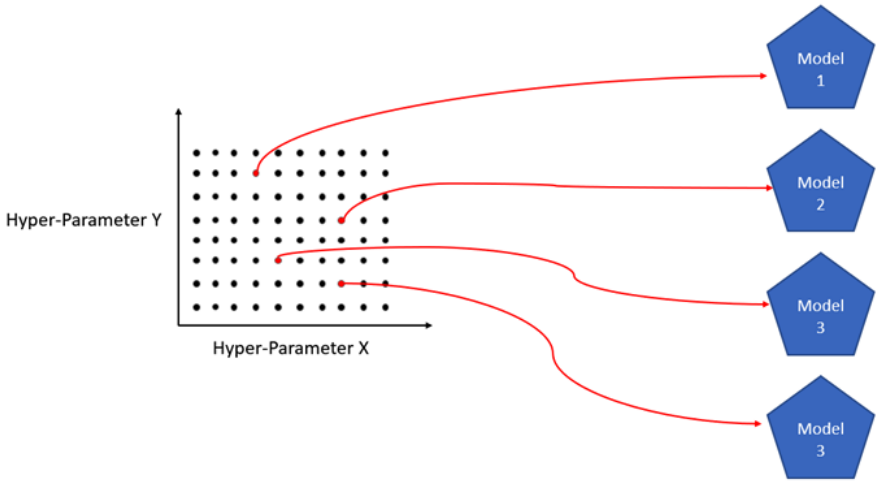

H2O Client
H2O Server
script
H2O algo
cURL
H2O web server
H2O algo endpoint
H2O model endpoint
H2O jobs endpoint
H2O Algo Job
H2O Algo Task
H2O Jobs Manager
H2O Fork/Join
H2O DKV
User starts script
Script calls H2O algo function with parameters
Function creates a POST request for H2O Server
HTTPS Query TCP/IP
Dispatches Request
Creates model training job for the algorithm
Starts Job
New job_id created
Return job_id
Return job_id
Prepare JSON response with job_id
Return job_id
Create tasks
Submit tasks to the execution framework
Execute tasks
Task complete
Task complete
Write model to the Key-Value store under model_id
Job complete


H2O Client
H2O Server
script
H2O algo
cURL
H2O web server
H2O algo endpoint
H2O model endpoint
H2O jobs endpoint
H2O Algo Job
H2O Algo Task
H2O Jobs Manager
H2O Fork/Join
H2O DKV
Function creates GET request passing job_id
HTTPS Query TCP/IP
Dispatches Request
Query job_id
Return job_info
Prepare JSON response with job_info
Return JSON In HTTP TCP/IP Response
Return job_info

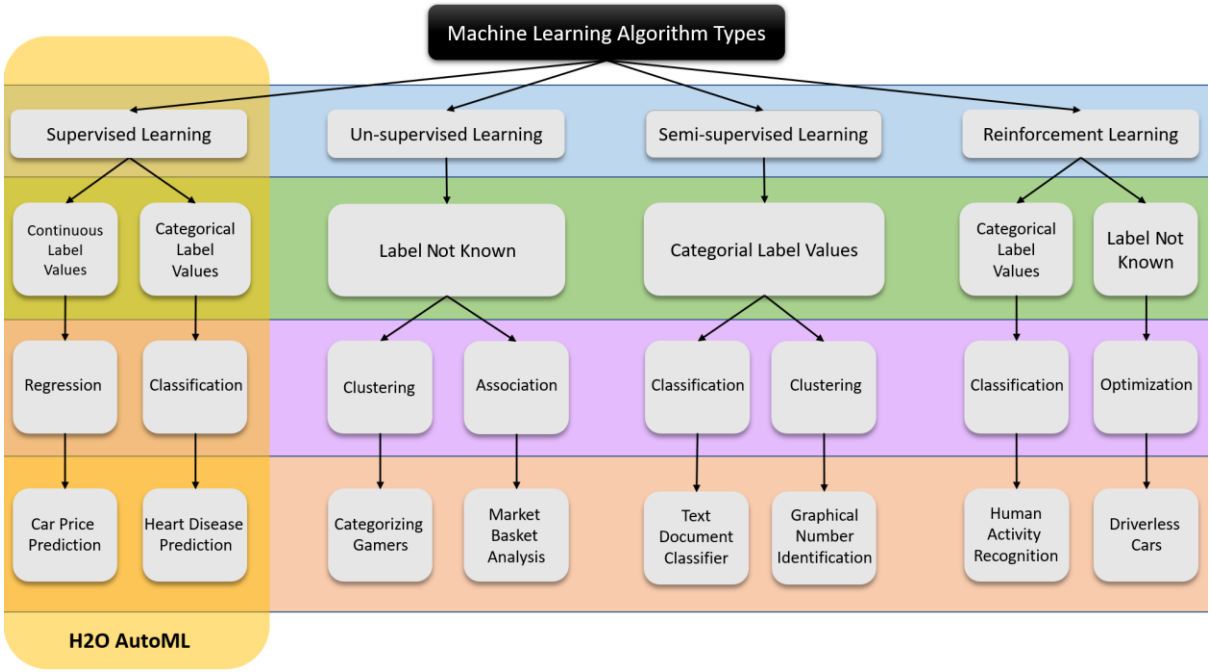

H2O Client
H2O Server
script
H2O algo
cURL
H2O web server
H2O algo endpoint
H2O model endpoint
H2O jobs endpoint
H2O Algo Job
H2O Algo Task
H2O Jobs Manager
H2O Fork/Join
H2O DKV
Function creates GET request passing model_id
HTTPS Query TCP/IP
Dispatches Request
Query model_id
Return model_id
Prepare JSON response with model_info
Return JSON in HTTP TCP/IP Response
Return model_info
Return model_info

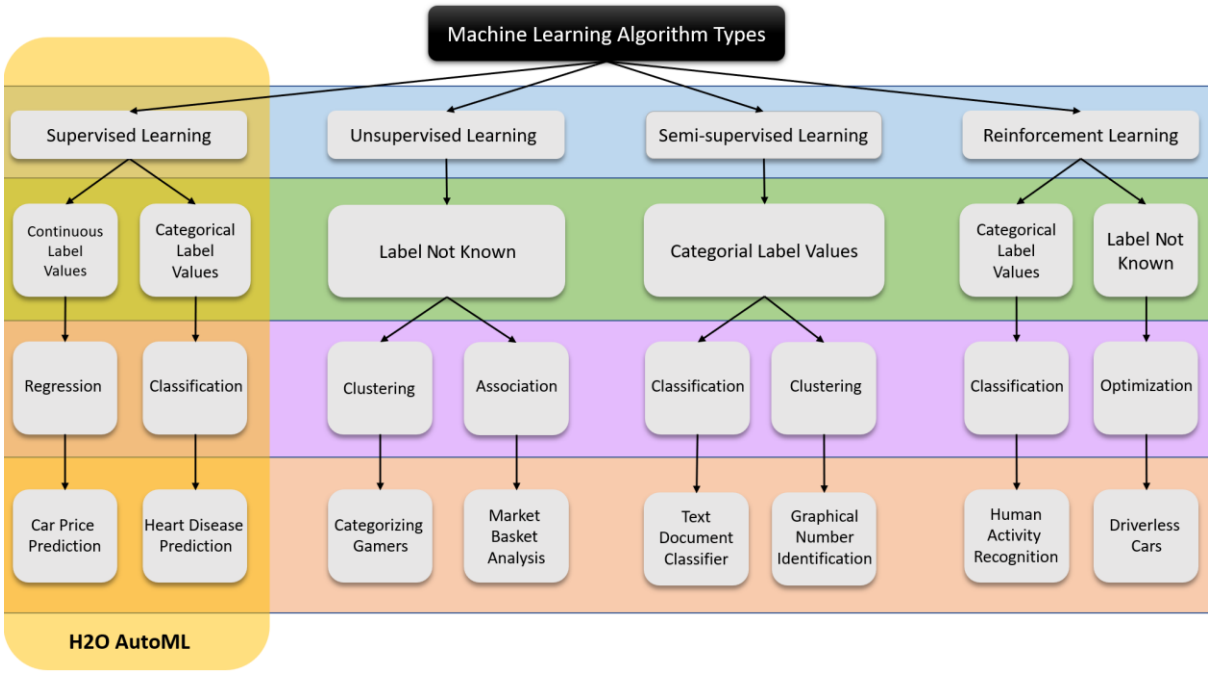

Hyper-Parameters
X1
Y1
X2
Y2
X3
Y3
Xn
Yn
Model 1
Model 2
Model 3
Model N

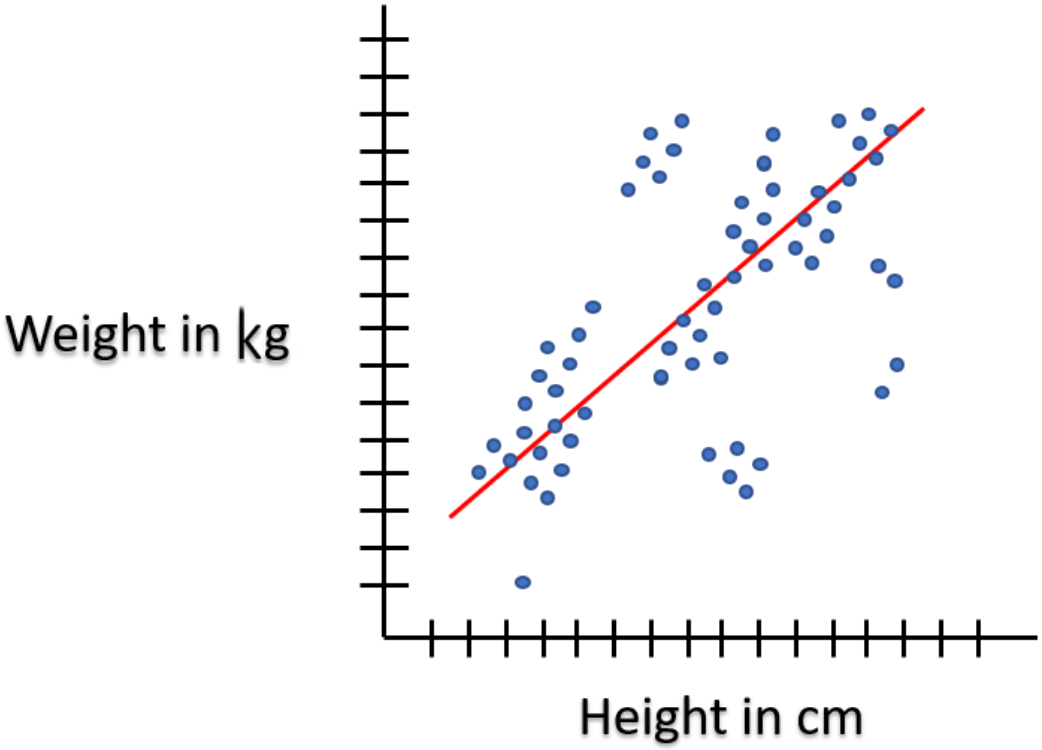

Hyper-Parameter Y
Hyper-Parameter X
Model 1
Model 2
Model 3
Model N

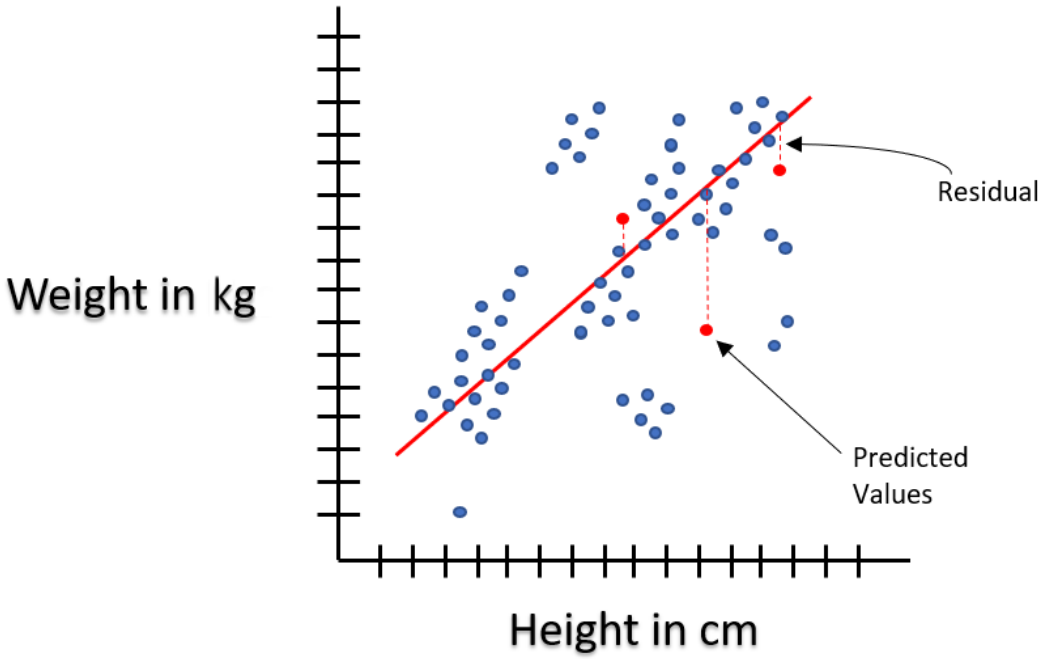

Hyper-Parameter Y
Hyper-Parameter X
Model 1
Model 2
Model 3
Model 3

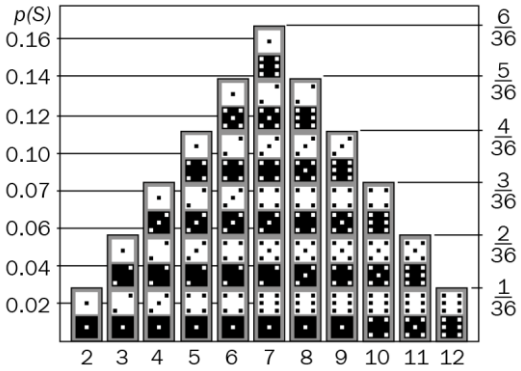
Chapter 5: Understanding AutoML Algorithms

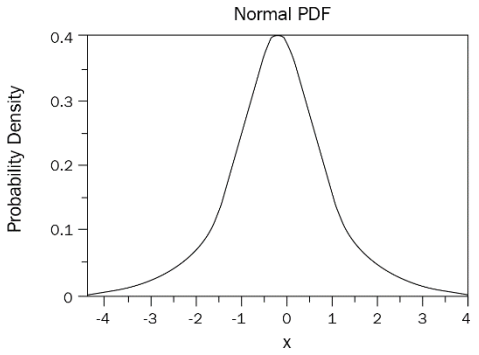


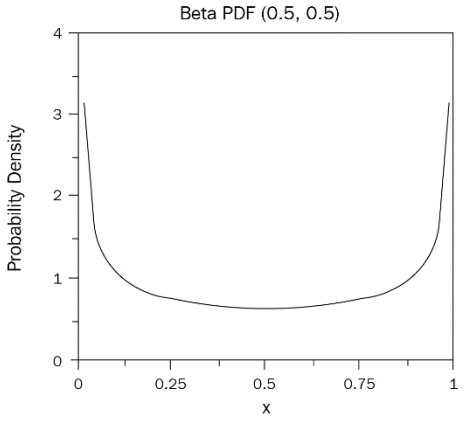


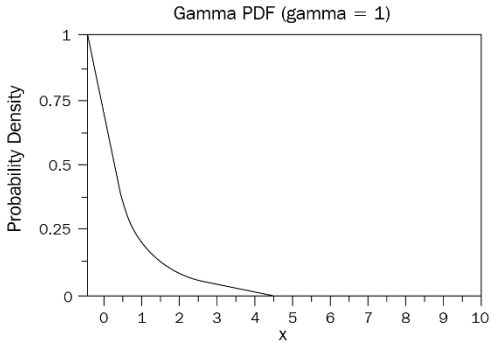

Weight in kg
Height in cm

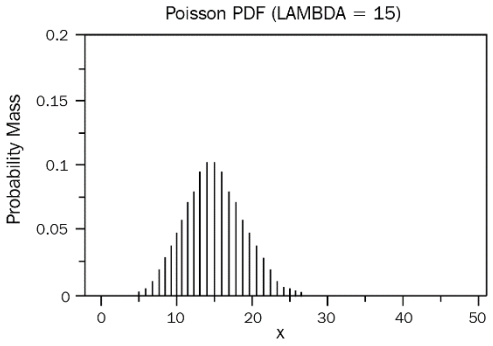

Weight in kg
Residual
Predicted
Values
Height in cm

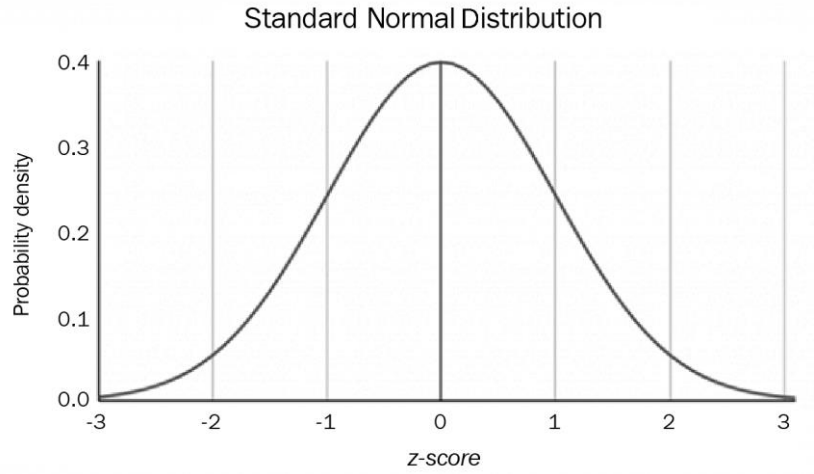

p(S)
0.16
0.14
0.12
0.10
0.07
0.06
0.04
0.02
2 3 4 5 6 7 8 9 10 11 12
6/36
5/36
4/36
3/36
2/36
1/36


Normal PDF
Probability Density
x

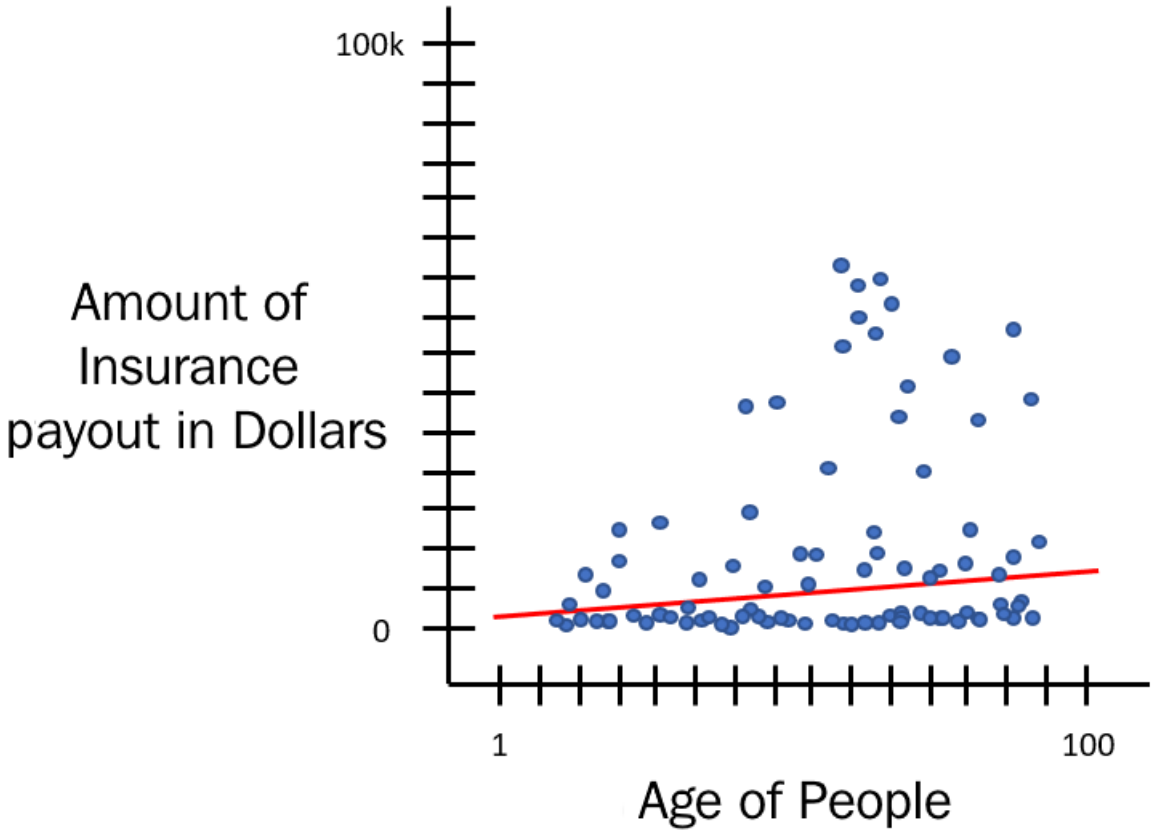

Beta PDF (0.5, 0.5)
Probability Density
x

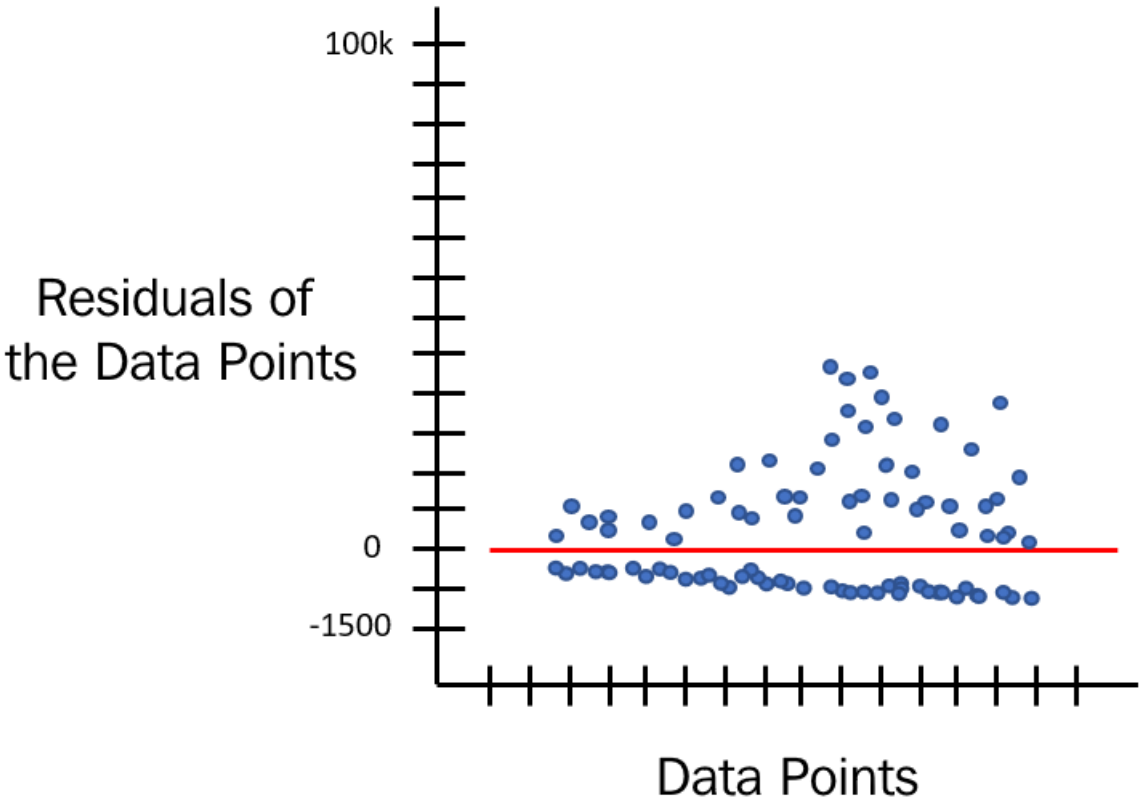

Gamma PDF (gamma = 1)
Probability Density
x

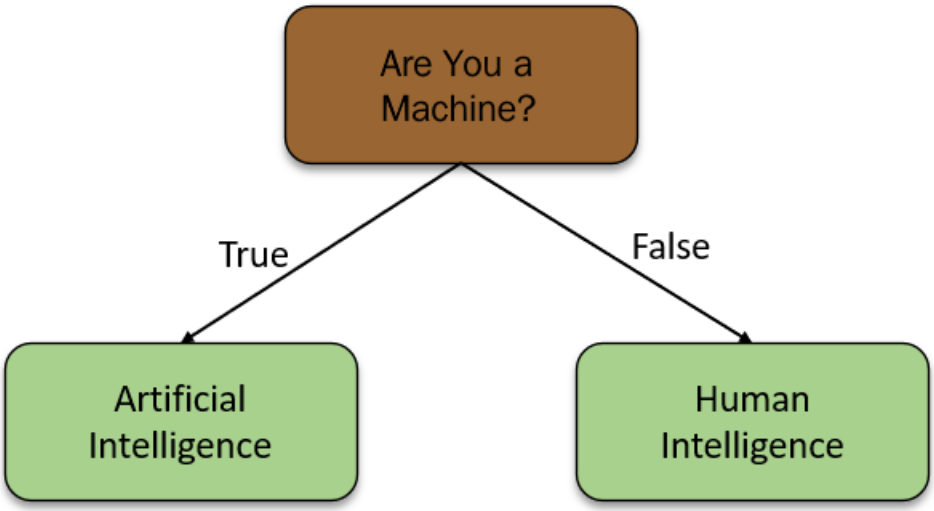

Poisson PDF (LAMBDA = 15)
Probability Mass
x

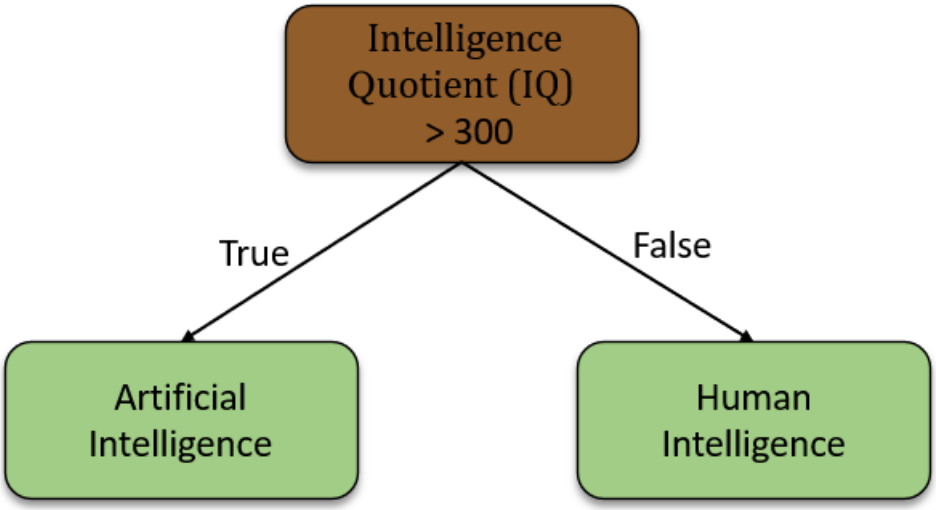

Standard Normal Distribution
Probability density
z-score

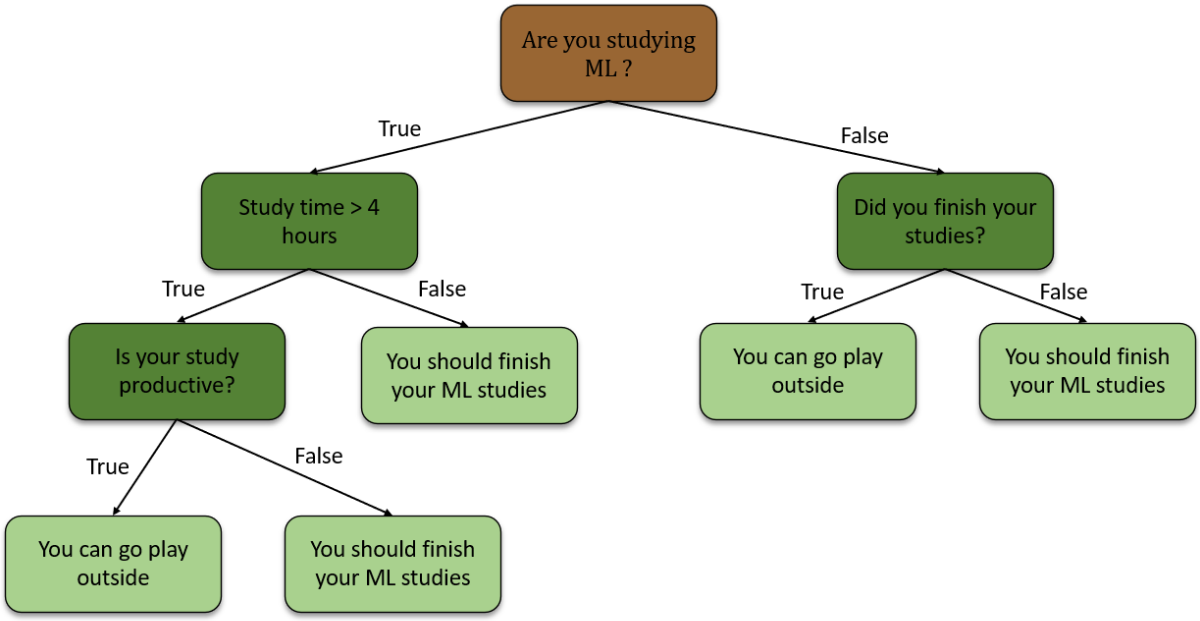

100k
Amount of Insurance payout in Dollars
0
1
100
Age of People


100k
Amount of Insurance payout in Dollars
0
1
100
Age of People

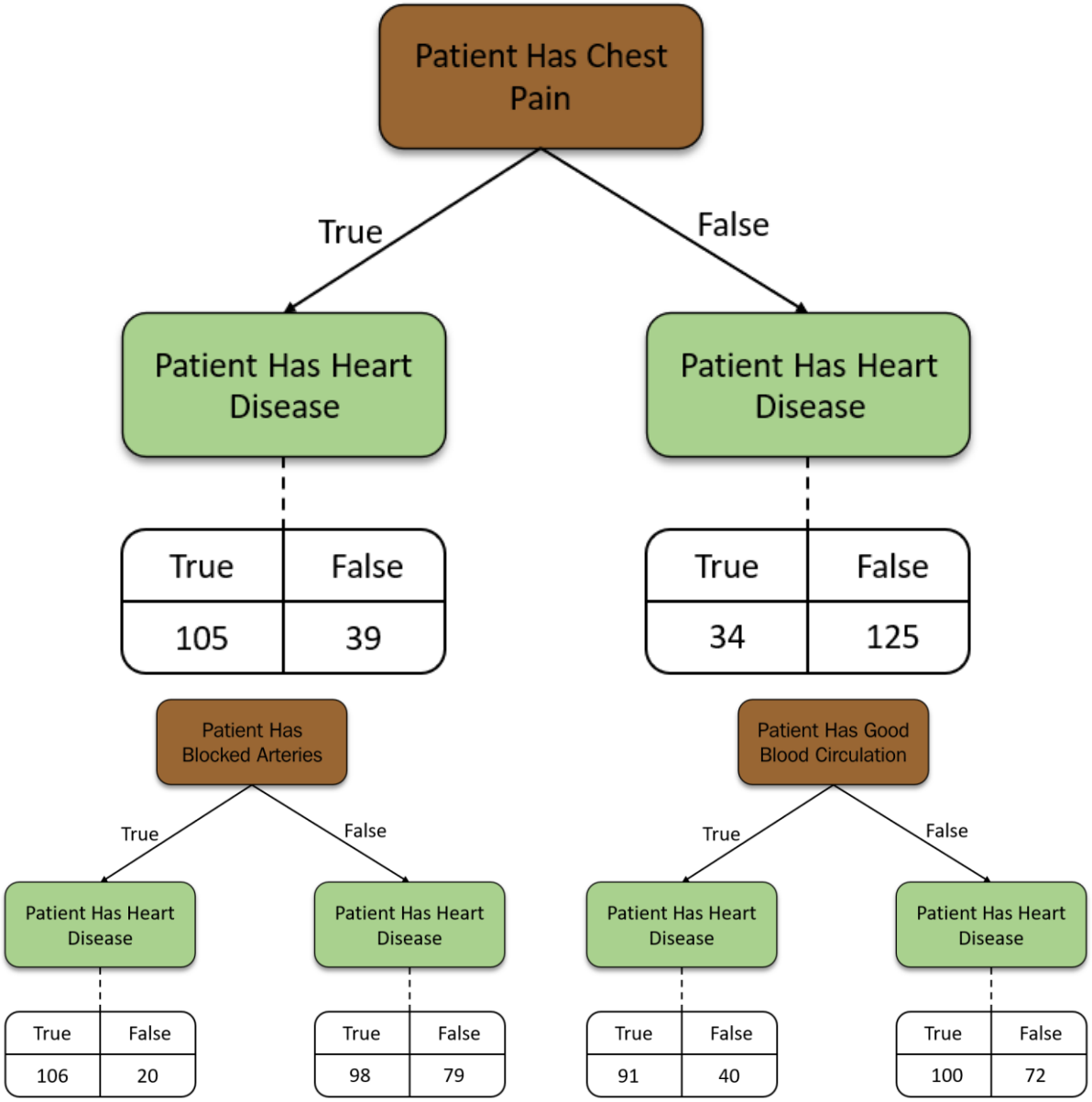

100k
Residuals of
the Data Points
0
-1500
Data Points

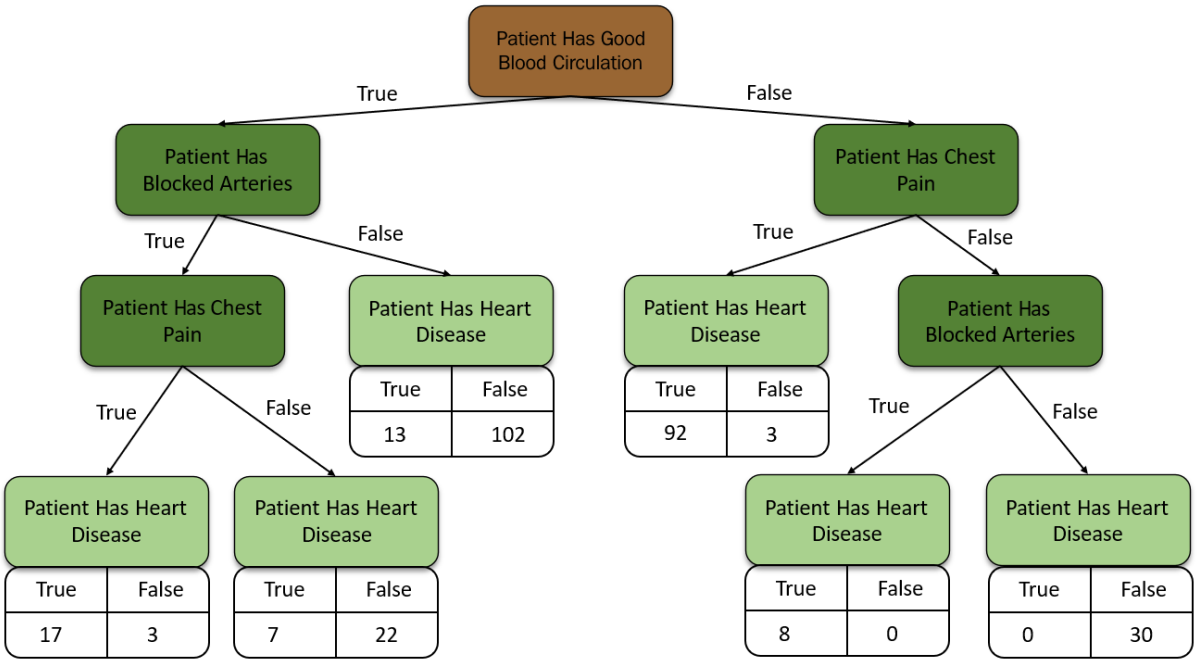

Are You a
Machine?
True
False
Artificial
Intelligence
Human
Intelligence





Chest Pain	Good Blood Circulation	Blocked Arteries	Heart Disease
False	False	False	False
True	True	True	True
True	True	False	False
True	False	N/A	True
False	False	True	True

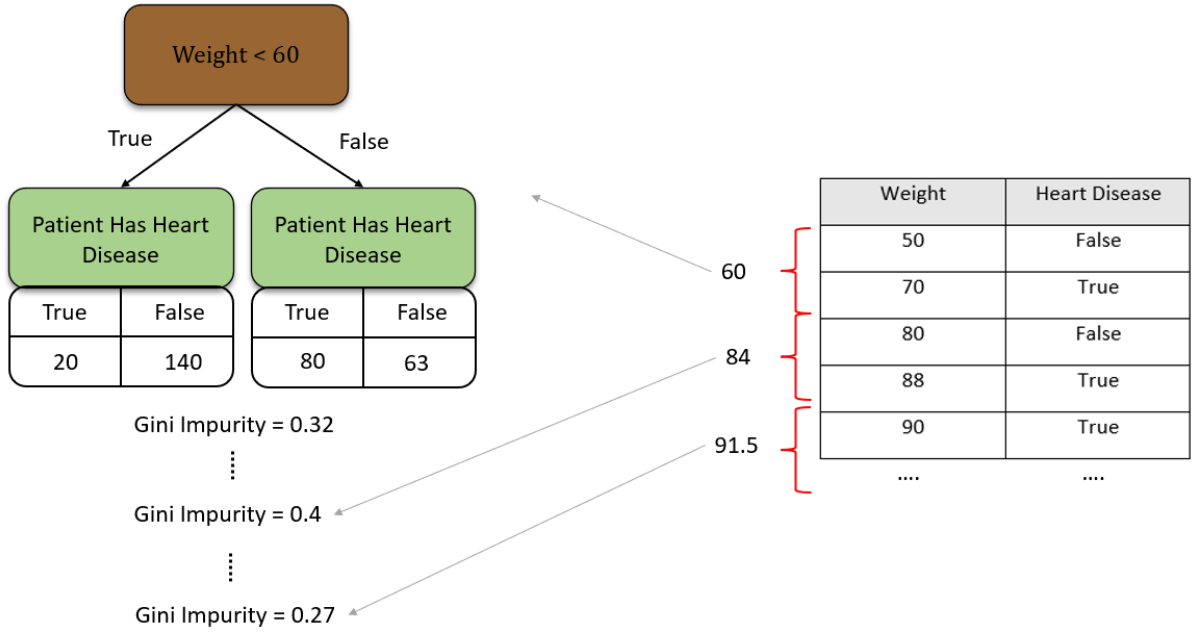

Patient Has Chest Pain
True
False
Patient Has Heart Disease
Patient Has Heart Disease
True
False
105
39
True
False
34
125
Patient Has Blocked Arteries
Patient Has Good Blood Circulation
True
False
True
False
Patient Has Heart Disease
Patient Has Heart Disease
Patient Has Heart Disease
Patient Has Heart Disease
True
False
106
20
True
False
98
79
True
False
91
40
True
False
100
72



Chest Pain	Good Blood Circulation	Blocked Arteries	Weight	Heart Disease
False	False	False	90	False
True	True	True	70	True
True	True	False	88	False
True	False	N/A	50	True
False	False	True	80	True

60
84
91.5

Weight	Heart Disease
50	False
70	True
80	False
88	True
90	True
....	....



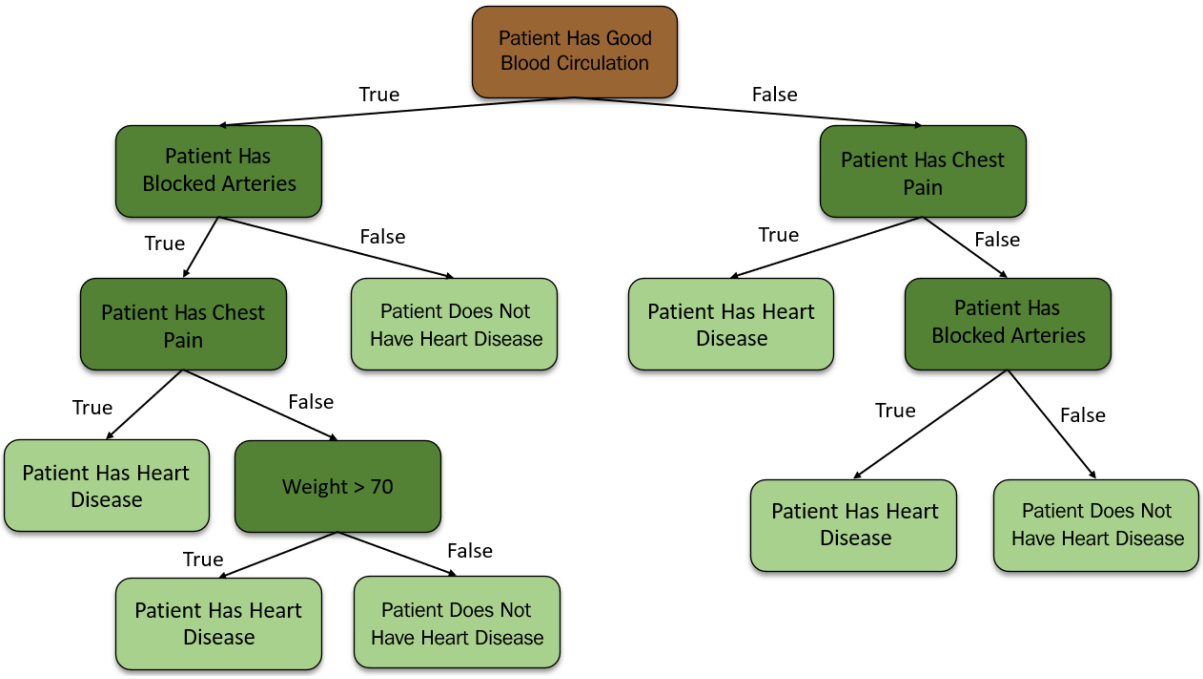
Chest Pain	Good Blood Circulation	Blocked Arteries	Weight	Heart Disease
False	False	False	90	False
True	True	True	70	True
True	True	False	88	False
True	False	N/A	50	True
False	False	True	80	True

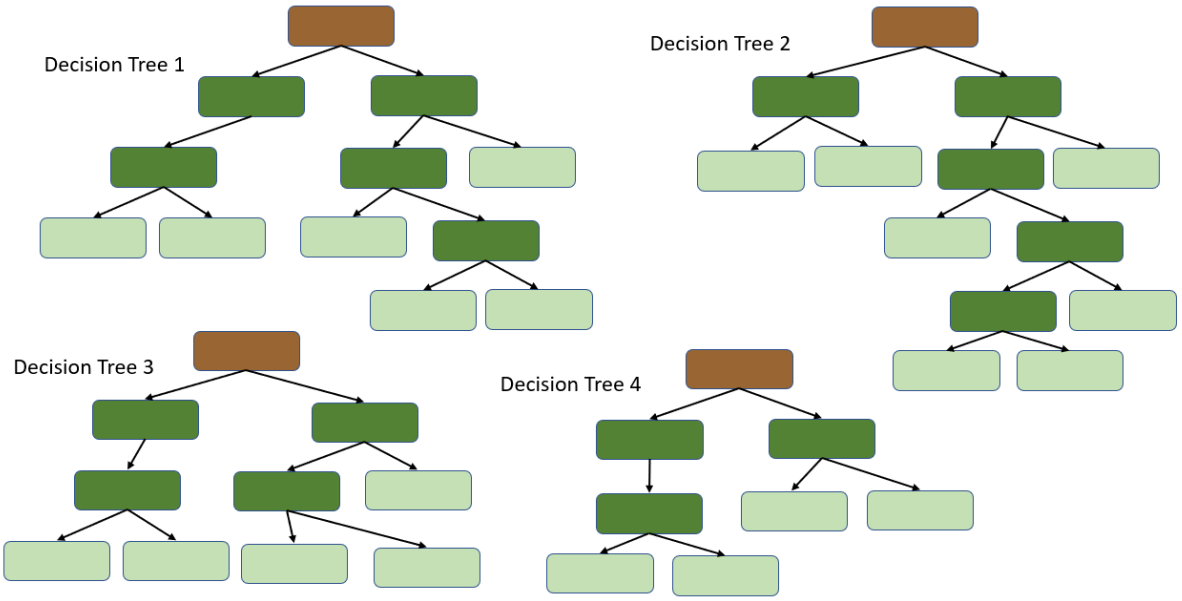
Original Dataset

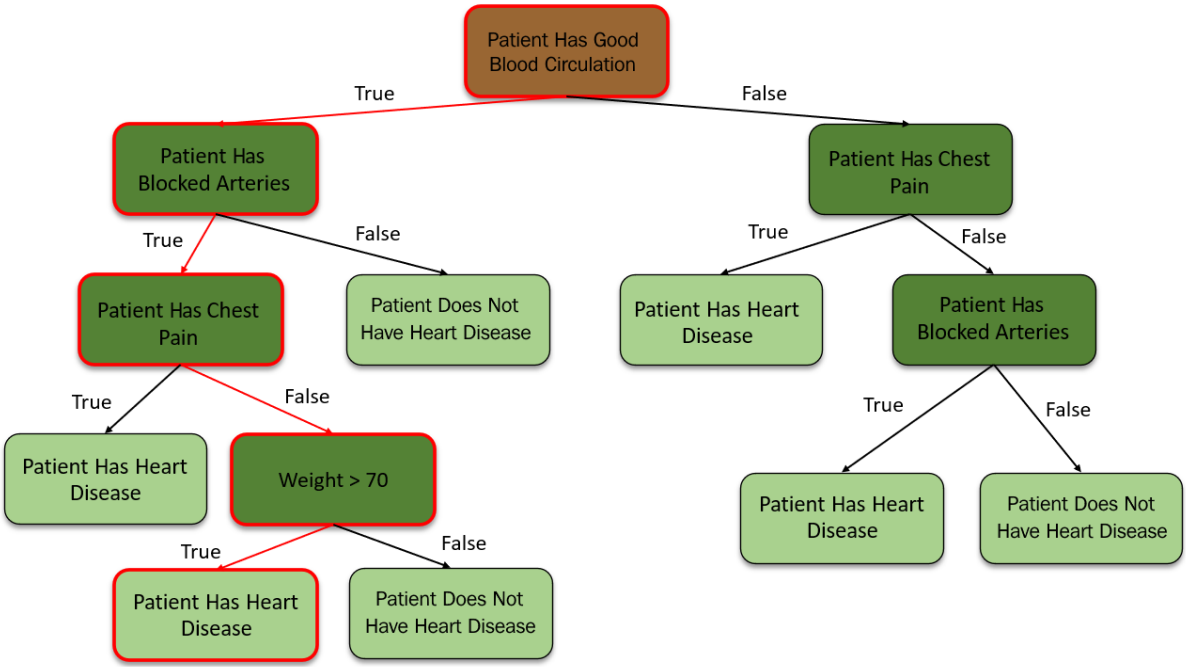
Random Selection

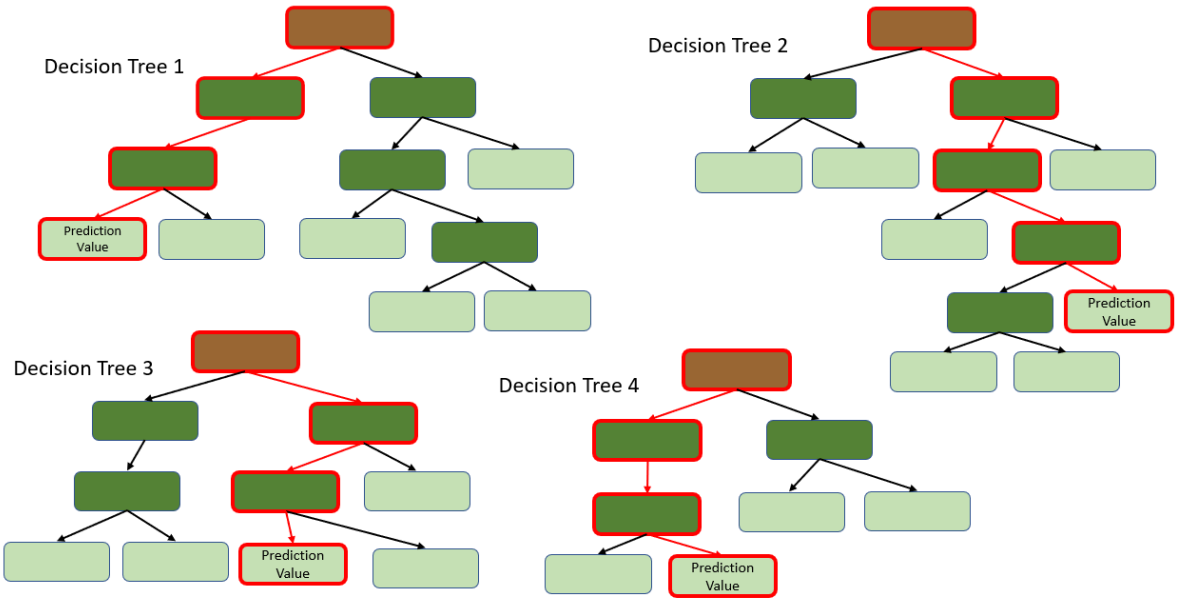
Chest Pain	Good Blood Circulation	Blocked Arteries	Weight	Heart Disease
False	False	False	90	False
False	False	True	80	True
True	True	False	88	False
True	True	False	88	False
True	True	True	70	True

Bootstrapped Dataset




Decision Tree 1
Decision Tree 2
Decision Tree 3
Decision Tree 4


Patient Has Good Blood Circulation
True
False
Patient Has Blocked Arteries
Patient Has Chest Pain
True
False
True
False
Patient Has Chest Pain
Patient Does Not Have Heart Disease
Patient Has Heart Disease
Patient Has Blocked Arteries
True
False
True
False
Patient Has Heart Disease
Weight > 70
Patient Has Heart Disease
Patient Does Not Have Heart Disease
True
False
Patient Has Heart Disease
Patient Does Not Have Heart Disease



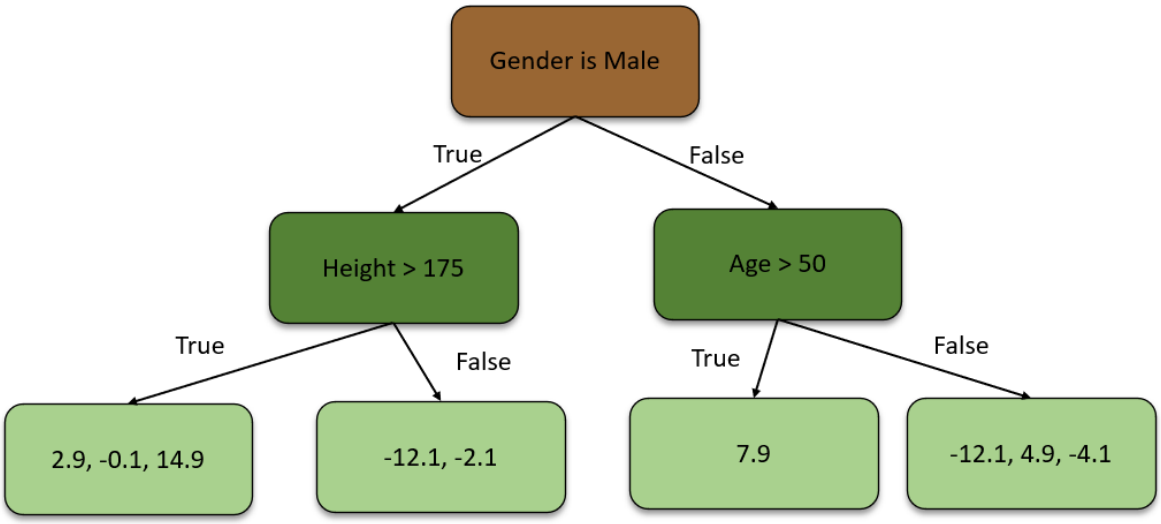
Height	Gender	Age	Weight
170	M	45	50
169	F	26	67
180	M	58	55
185	F	66	45
177	M	45	86
174	M	36	90
182	M	75	77
165	F	43	56
160	F	34	66

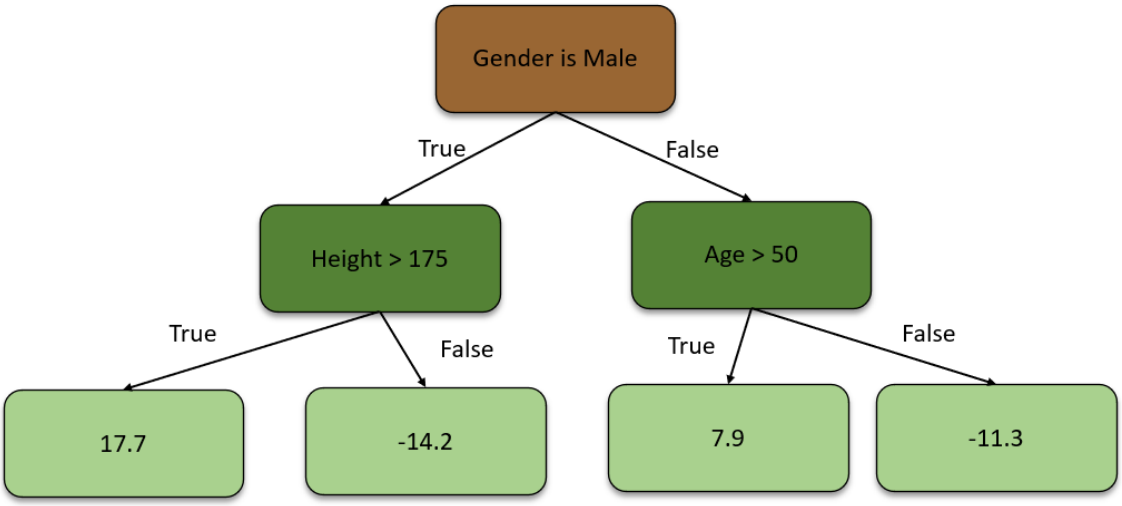
Height	Gender	Age	Weight
170	M	45	50
169	F	26	50
180	M	58	65
185	F	66	70
177	M	45	62
174	M	36	60
182	M	75	77
165	F	43	67
160	F	34	58

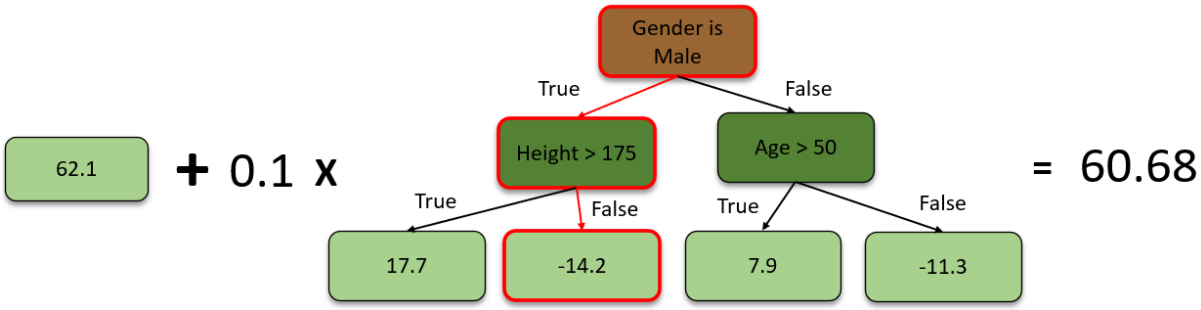
62.1

Average Value of the Weight Column

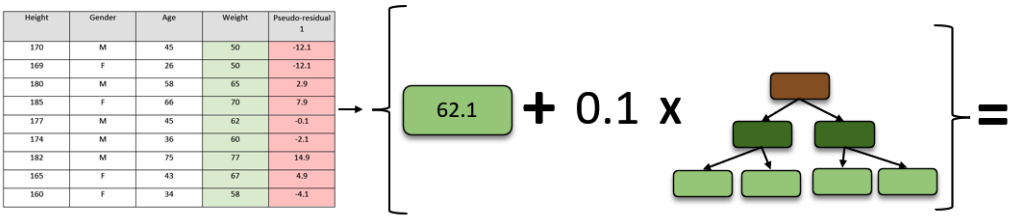
Height	Gender	Age	Weight	Pseudo-residual 1
170	M	45	50	-12.1
169	F	26	50	-12.1
180	M	58	65	2.9
185	F	66	70	7.9
177	M	45	62	-0.1
174	M	36	60	-2.1
182	M	75	77	14.9
165	F	43	67	4.9
160	F	34	58	-4.1



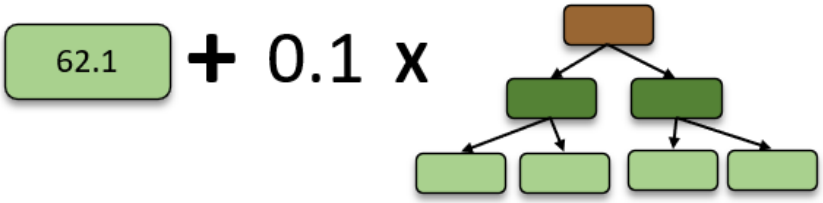




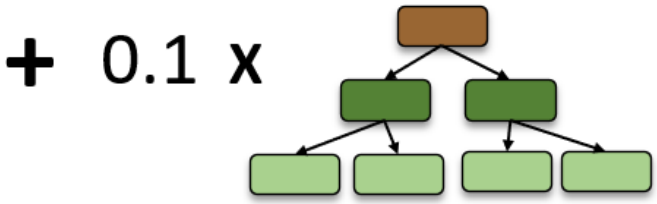
Height	Gender	Age	Weight	Pseudo-residual 1
170	M	45	50	-12.1
169	F	26	50	-12.1
180	M	58	65	2.9
185	F	66	70	7.9
177	M	45	62	-0.1
174	M	36	60	-2.1
182	M	75	77	14.9
165	F	43	67	4.9
160	F	34	58	-4.1



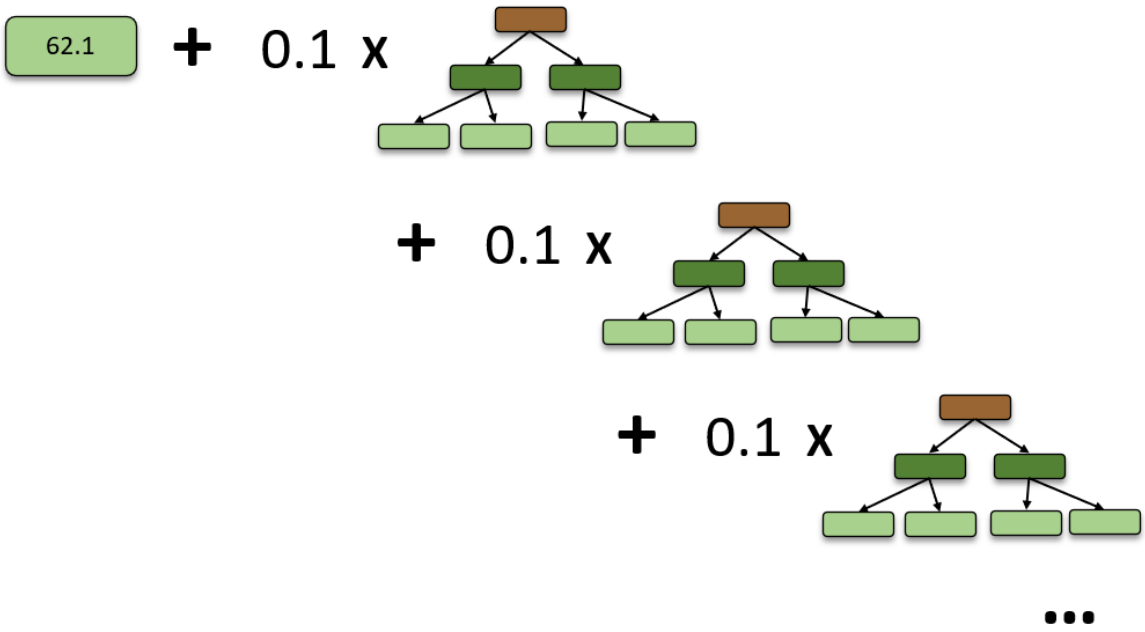
Predicted Weight	Pseudo-residual 2
60.68	-10.68
60.97	-10.97
60.33	4.67
61.31	8.96
62	0
60.33	-0.33
60.68	16.32
60.97	6.03
60.97	-2.97

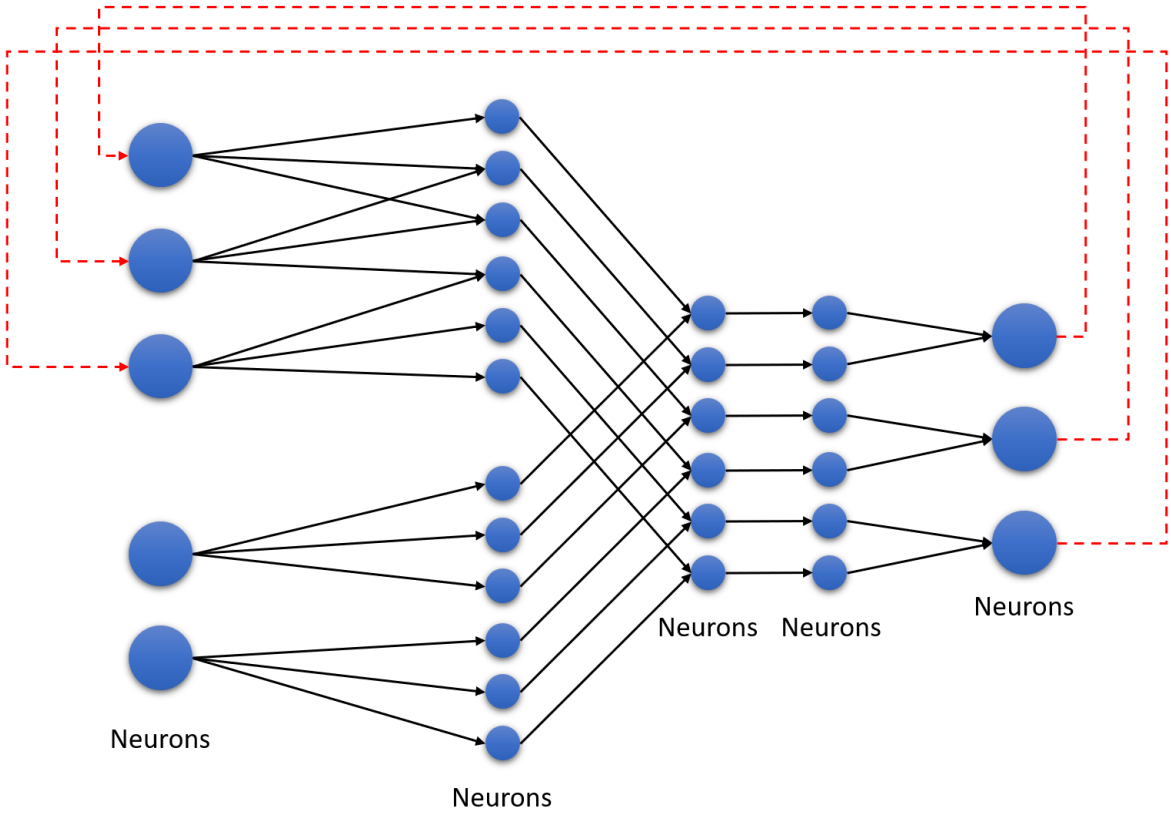


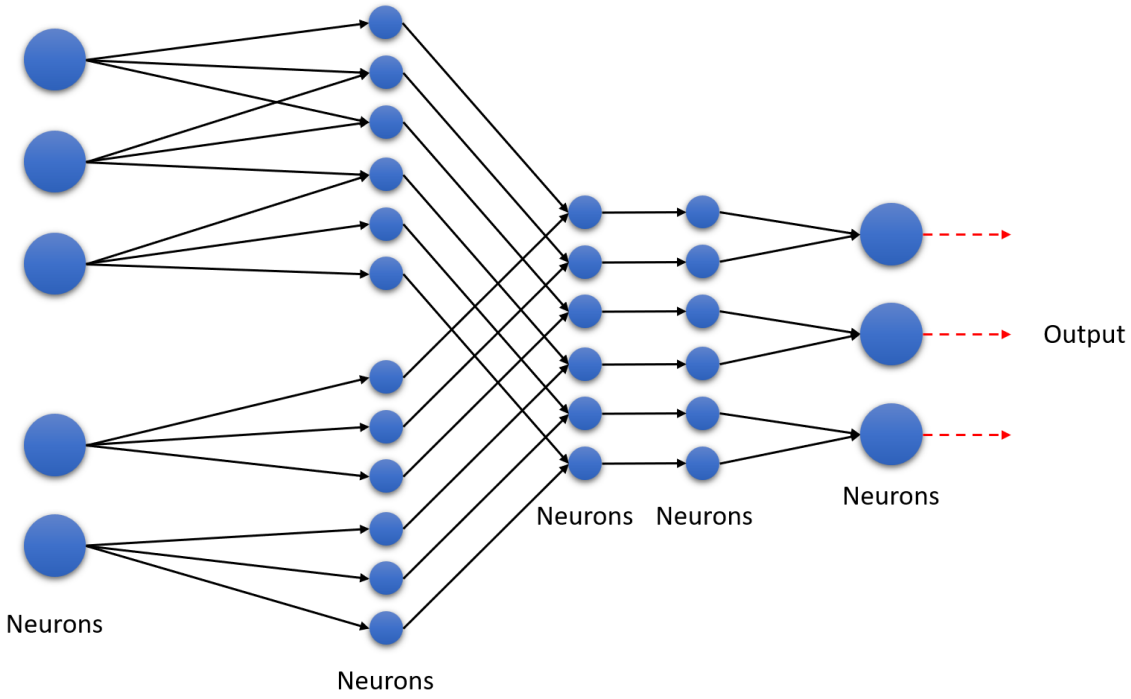
Decision Tree 1

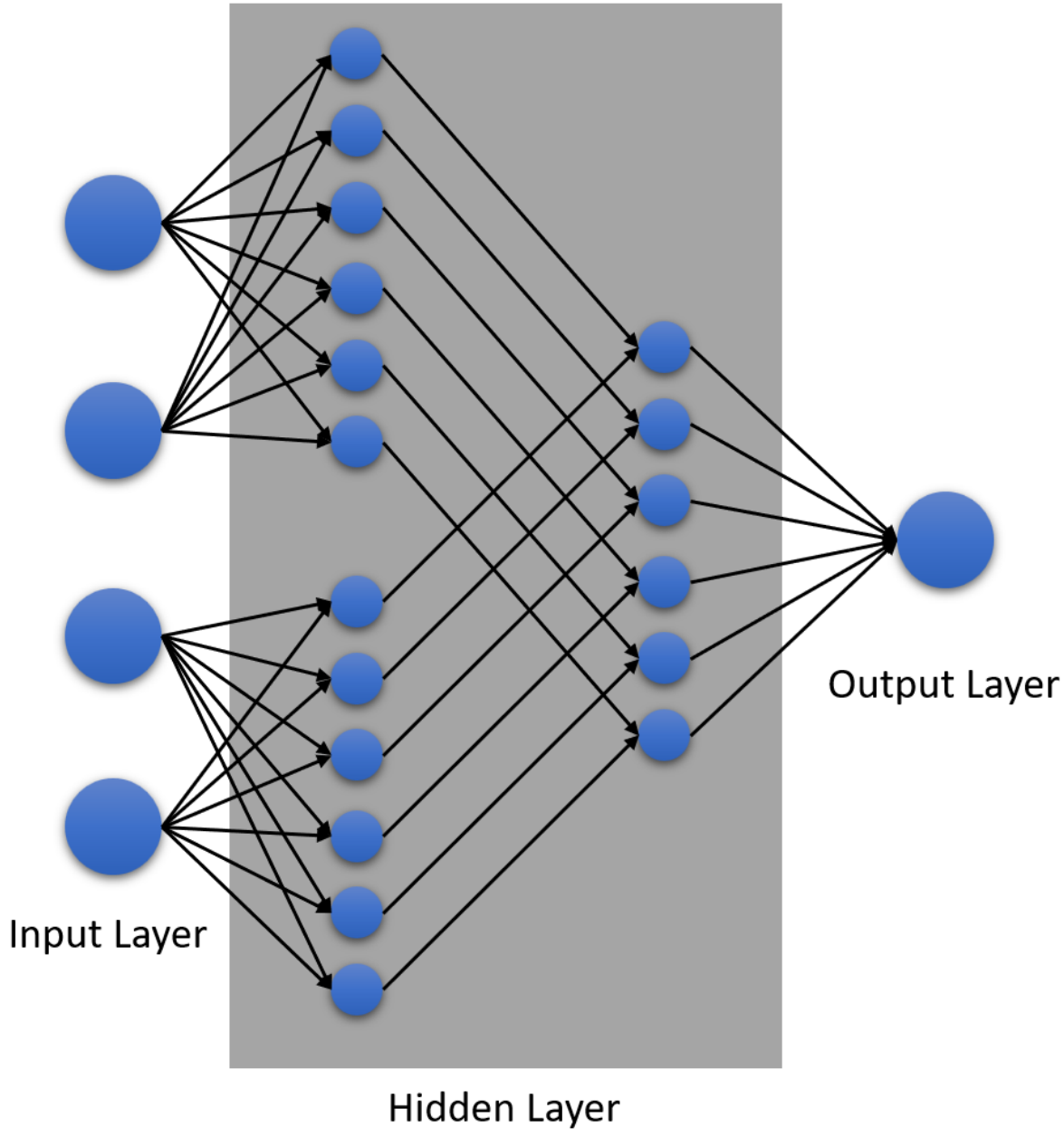


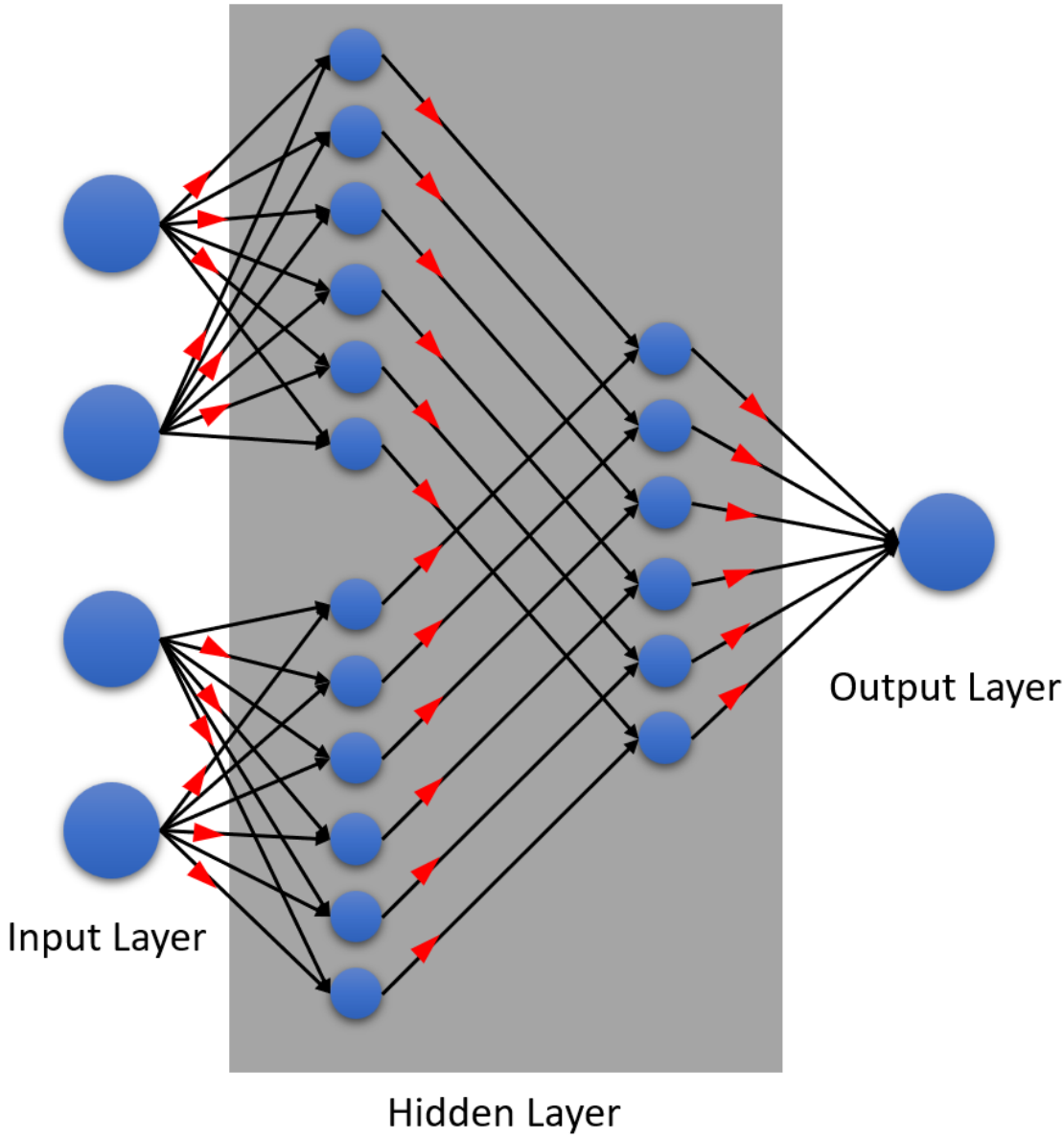
Decision Tree 2

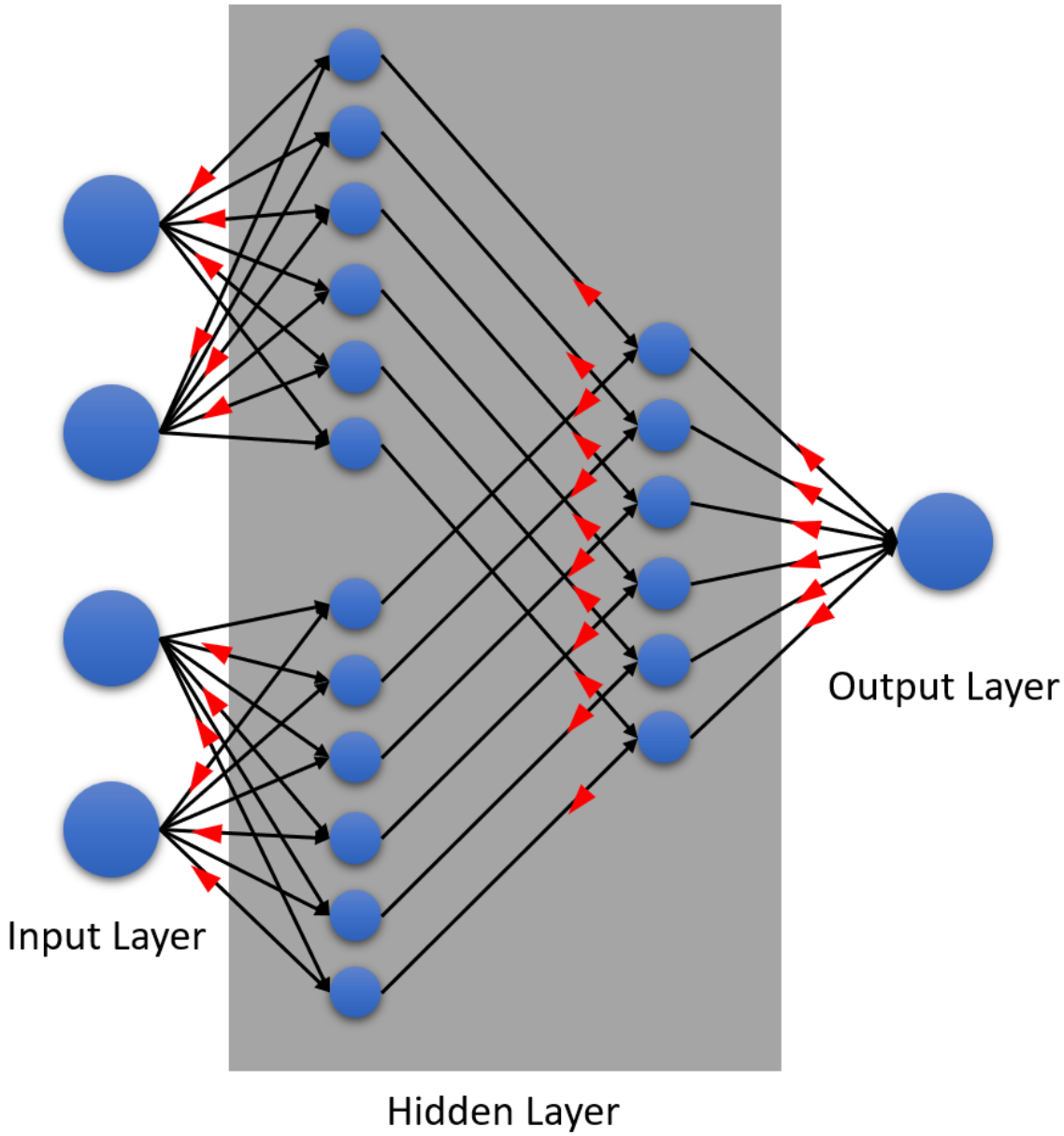

62.1
+ 0.1 x
+ 0.1 x
+ 0.1 x
...


Neurons
Neurons
Neurons
Neurons
Neurons

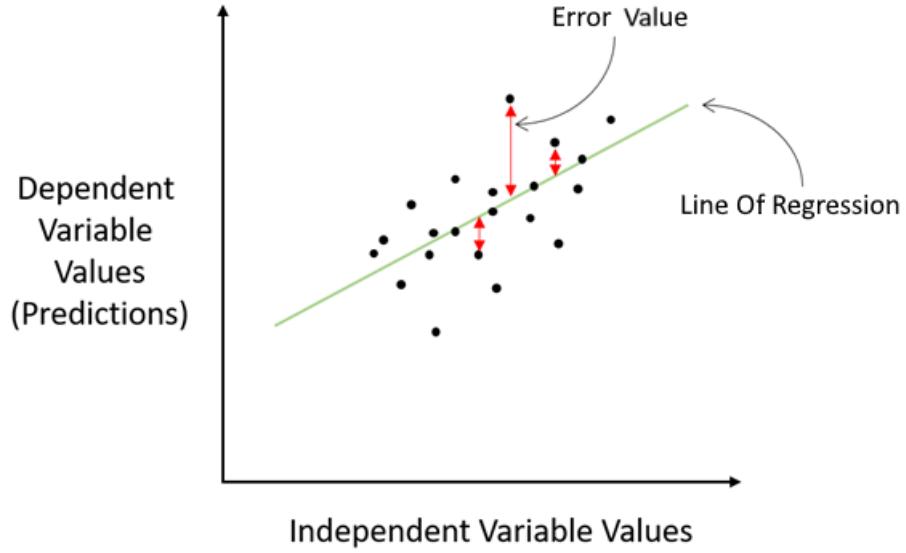

Neurons
Neurons
Neurons
Neurons
Neurons
Output


Input Layer
Hidden Layer
Output Layer


Input Layer
Hidden Layer
Output Layer


Input Layer
Hidden Layer
Output Layer

Chapter 6: Understanding H2O AutoML Leaderboard and Other Performance Metrics

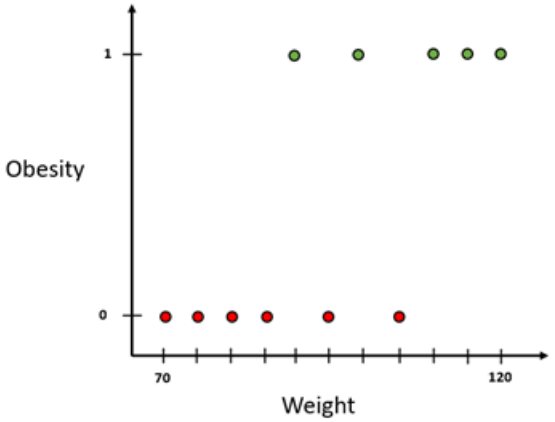


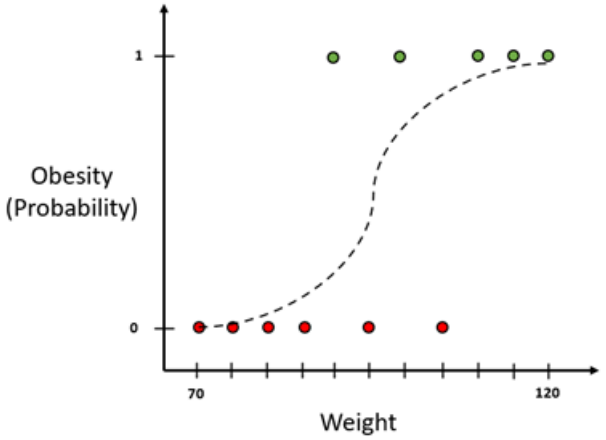
		Actual Values: Patient Has Heart Disease	Actual Values: Patient Does Not Have Heart Disease
Predicted Values	Patient Has Heart Disease	True Positive	False Positive
	Patient Does Not Have Heart Disease	False Negative	True Negative

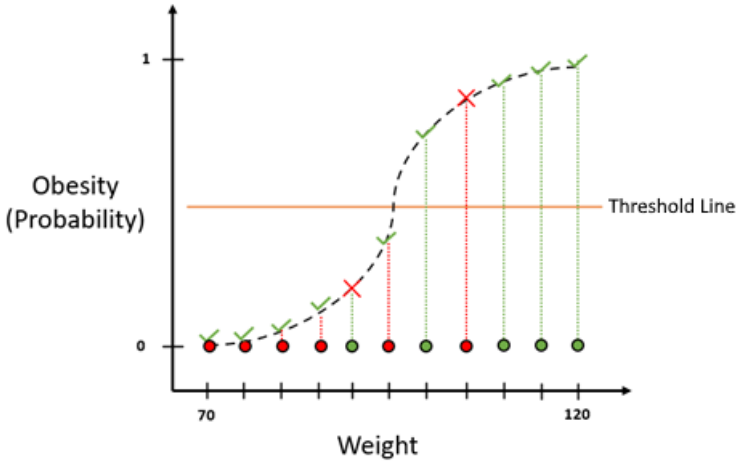
Actual Values

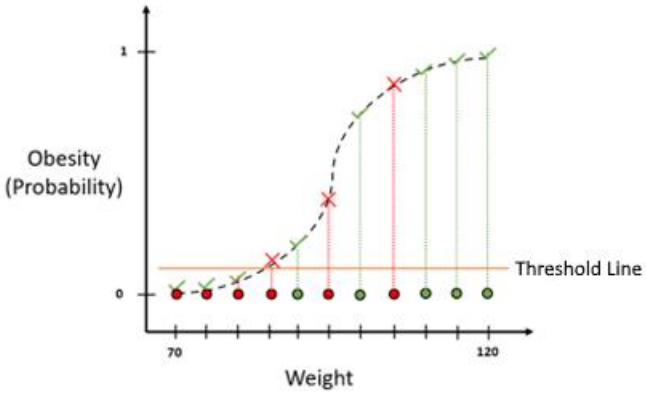
Predicted Values	Class A	Class B	Class C	Class D	Class E	Class F
Class A	Correct Prediction	Incorrect Prediction	Incorrect Prediction	Incorrect Prediction	Incorrect Prediction	Incorrect Prediction
Class B	Incorrect Prediction	Correct Prediction	Incorrect Prediction	Incorrect Prediction	Incorrect Prediction	Incorrect Prediction
Class C	Incorrect Prediction	Incorrect Prediction	Correct Prediction	Incorrect Prediction	Incorrect Prediction	Incorrect Prediction
Class D	Incorrect Prediction	Incorrect Prediction	Incorrect Prediction	Correct Prediction	Incorrect Prediction	Incorrect Prediction
Class E	Incorrect Prediction	Incorrect Prediction	Incorrect Prediction	Incorrect Prediction	Correct Prediction	Incorrect Prediction
Class F	Incorrect Prediction	Incorrect Prediction	Incorrect Prediction	Incorrect Prediction	Incorrect Prediction	Correct Prediction

Weight (kgs)	Obese
70	0
75	0
80	0
85	0
90	1
95	0
100	1
105	0
110	1
115	1
120	1







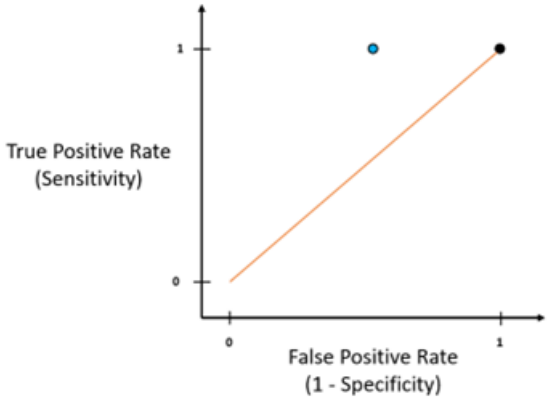


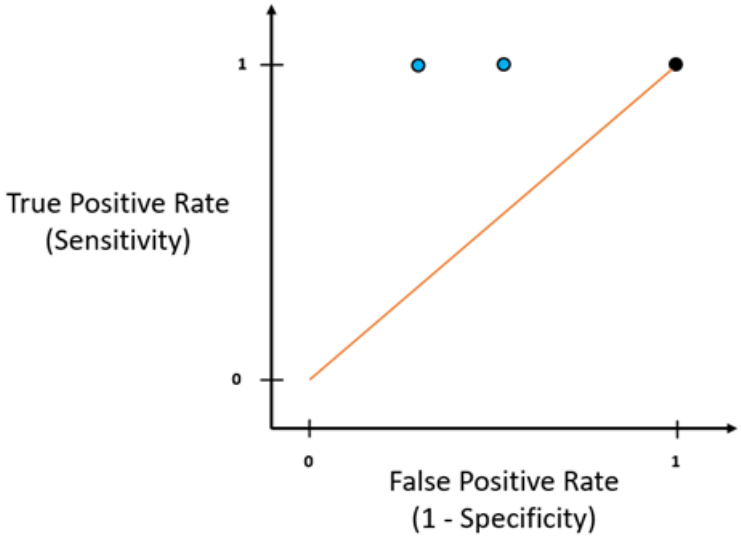
Actual Values

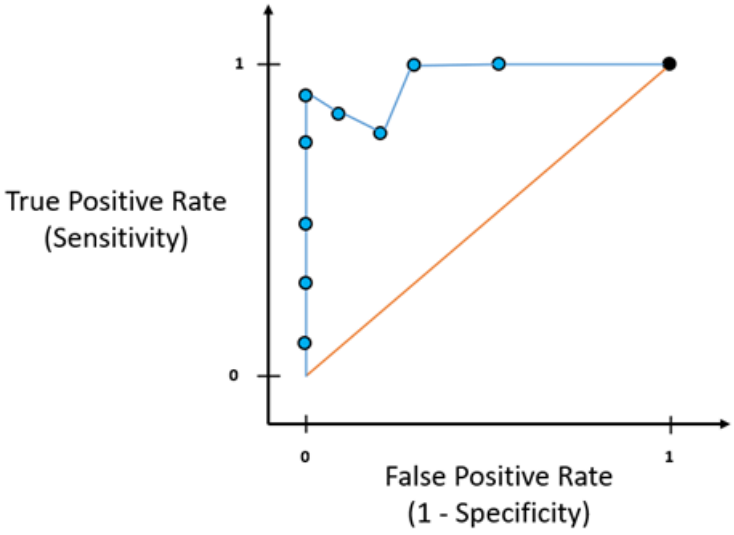
Predicted Values	Patient is Obese	Patient is not Obese
Patient is Obese	5	3
Patient is not Obese	0	3

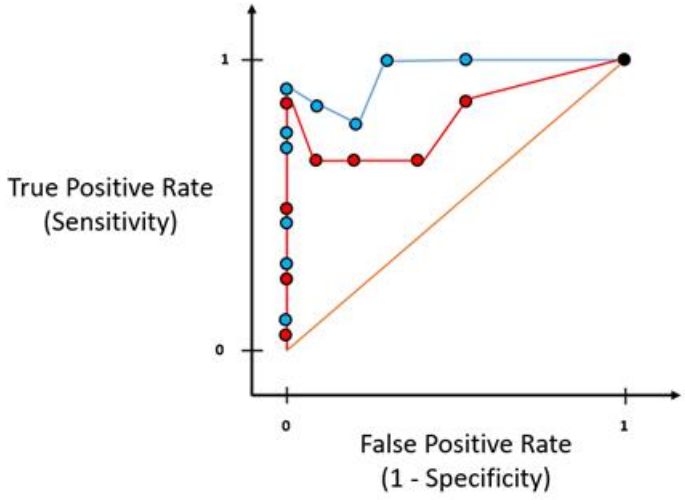
$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{5}{5 + 0} = 1$$

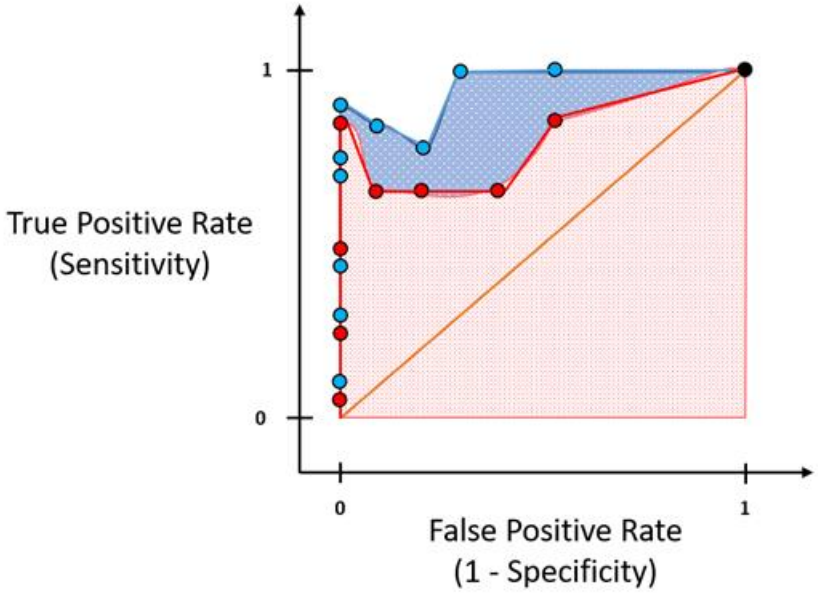
$$1 - \text{Specificity} = 1 - \frac{TN}{TN + FP} = 1 - \frac{3}{3 + 3} = 0.5$$






1
True Positive Rate
(Sensitivity)
0
0
1
False Positive Rate
(1 - Specificity)

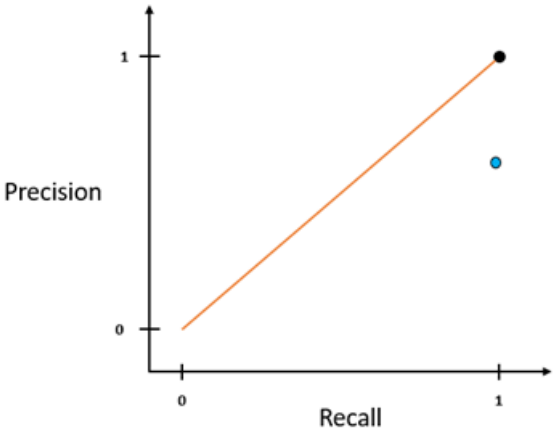

1
True Positive Rate
(Sensitivity)
0
0
1
False Positive Rate
(1 - Specificity)

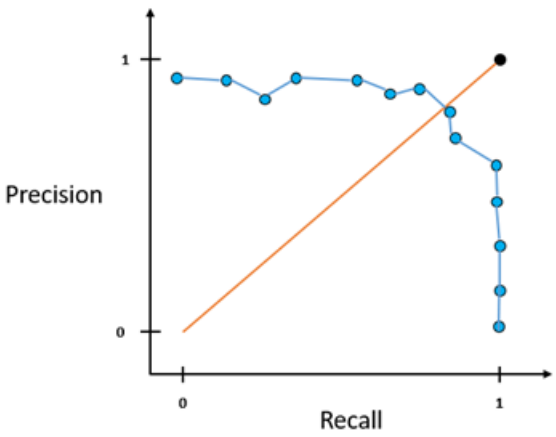

1
True Positive Rate
(Sensitivity)
0
0
1
False Positive Rate
(1 - Specificity)

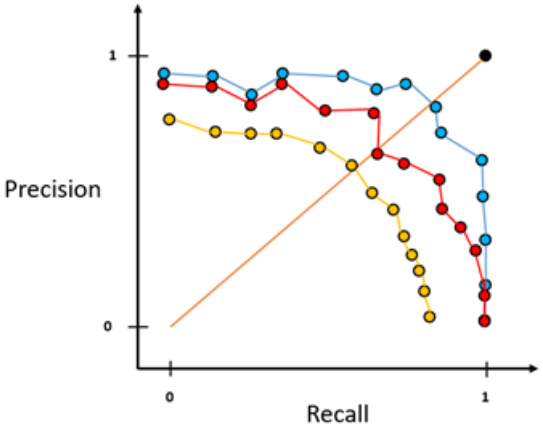
		Actual Values: Patient is Obese	Actual Values: Patient is not Obese
Predicted Values	Patient is Obese	5	3
	Patient is not Obese	0	3

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{5}{5 + 3} = 0.625$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{5}{5 + 0} = 1$$

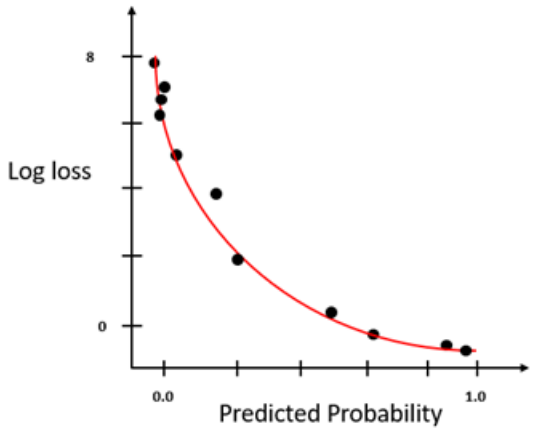


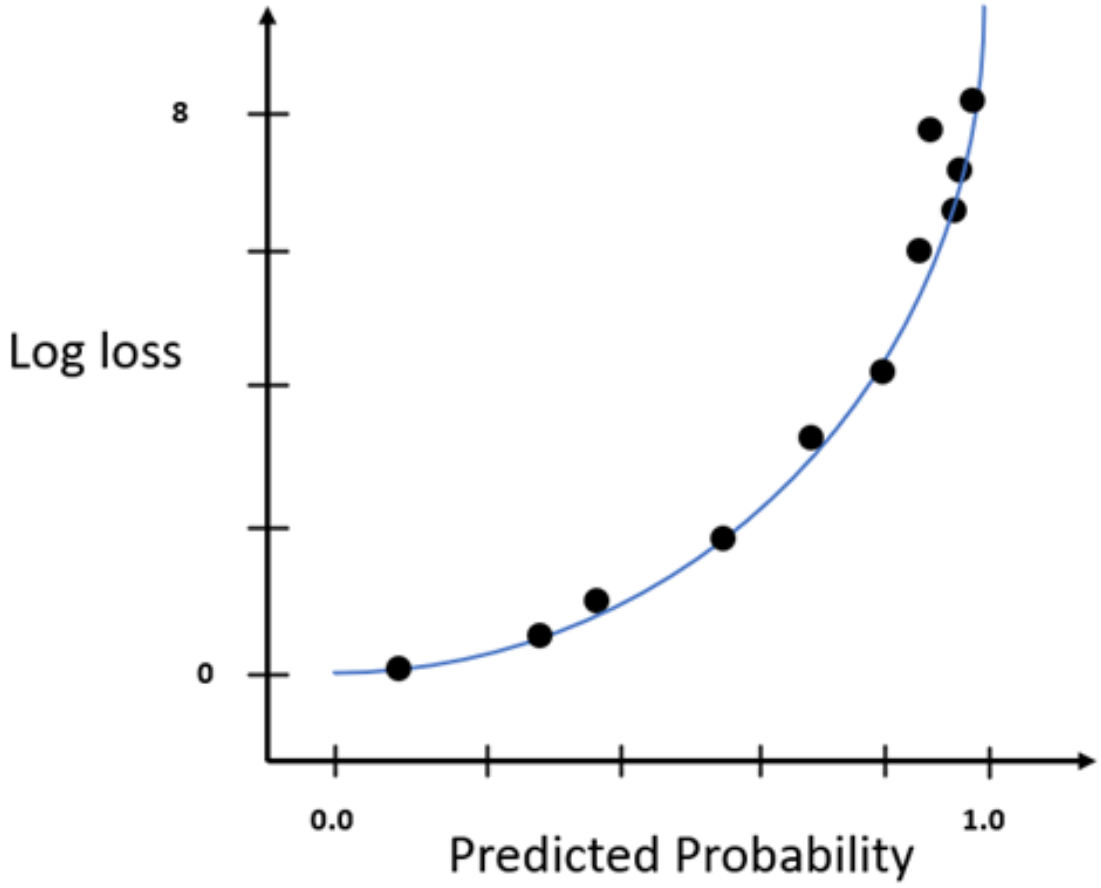




Weight (kgs)	Obese	Prediction Probability
70	0	0.29
75	0	0.35
80	0	0.40
85	0	0.52
90	1	0.54
95	0	0.60
100	1	0.62
105	0	0.69
110	1	0.75
115	1	0.007
120	1	0.0002

Weight (kgs)	Obese	Prediction Probability		Log loss
70	0	0.29	→	0.14
75	0	0.35	→	0.18
80	0	0.40	→	0.22
85	0	0.52	→	0.31
90	1	0.54	→	0.26
95	0	0.60	→	0.39
100	1	0.62	→	0.20
105	0	0.69	→	0.50
110	1	0.75	→	0.12
115	1	0.007	→	2.15
120	1	0.0002	→	3.69





Actual Values

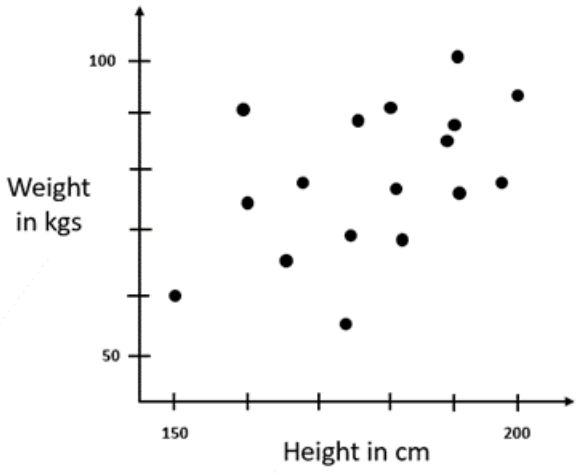
Predicted Values	Patient is Obese	Patient is not Obese
Patient is Obese	75	750
Patient is not Obese	30	6

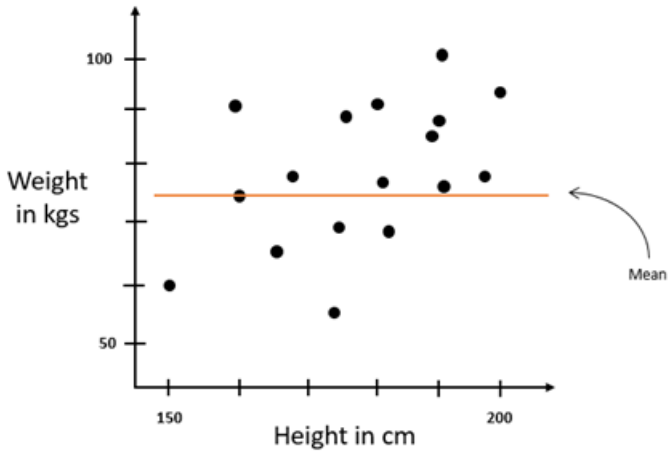
$$\text{Precision} = \frac{TP}{TP + FP} = \frac{75}{75 + 750} = 0.09$$

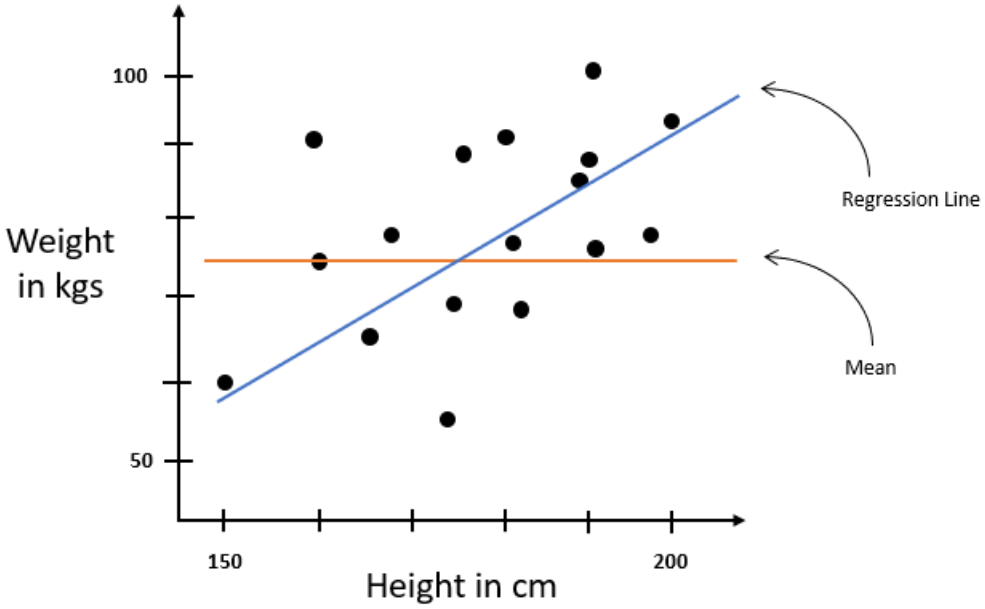
$$\text{Recall} = \frac{TP}{TP + FN} = \frac{75}{75 + 30} = 0.71$$

		Actual Values	
		Fruit is grape	Fruit is not a grape
Predicted Values	Fruit is grape	163	17
	Fruit is not a grape	12	8

		Actual Values	
		Fruit is watermelon	Fruit is not a watermelon
Predicted Values	Fruit is watermelon	8	12
	Fruit is not a watermelon	17	163




100
Weight
in kgs
50
Mean
150
200
Height in cm


100
Weight
in kgs
50
Regression Line
Mean
150
200
Height in cm

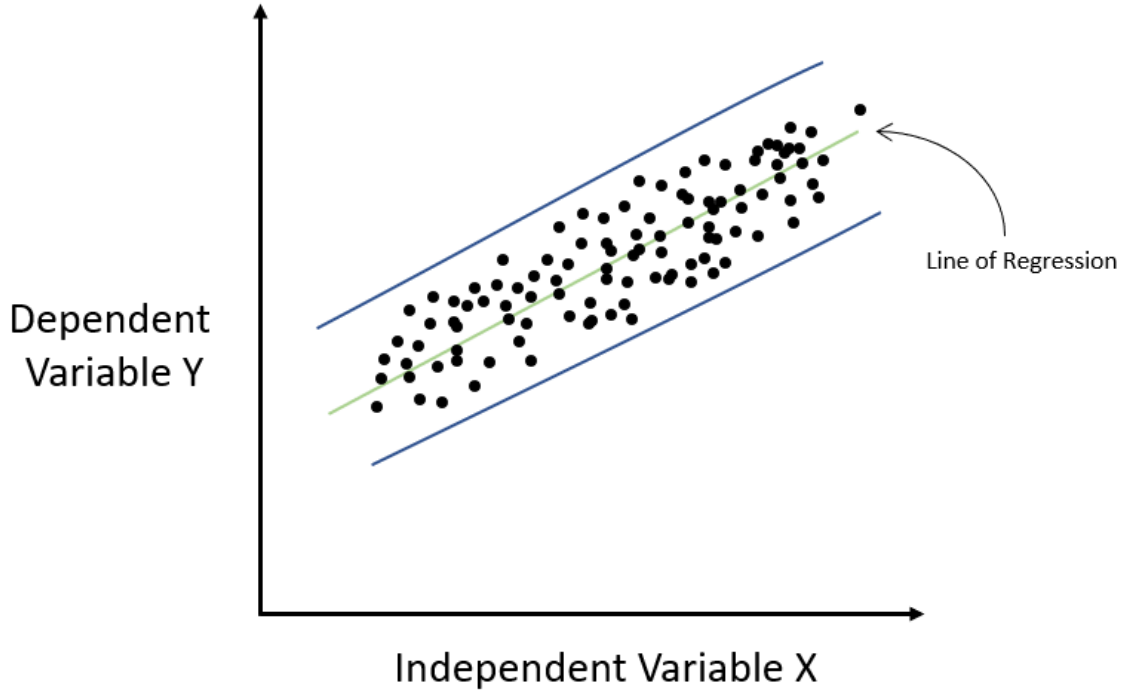
Chapter 7: Working with Model Explainability

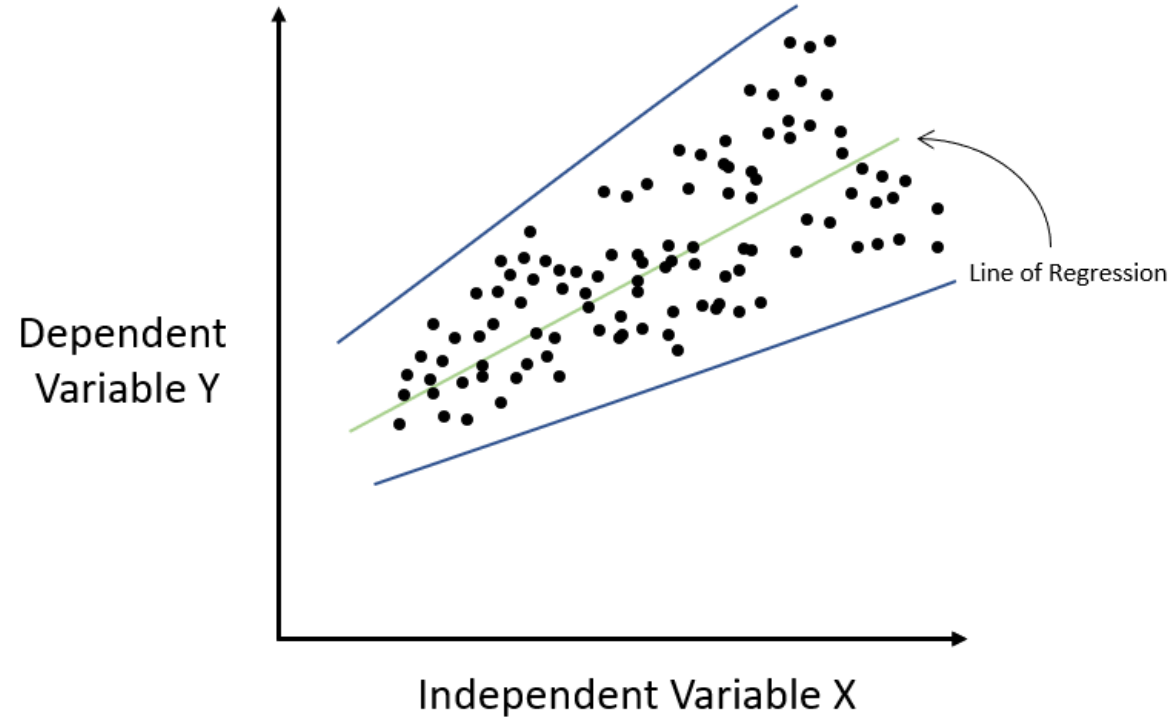
```
In [30]: aml.explain(test_dataframe)
```

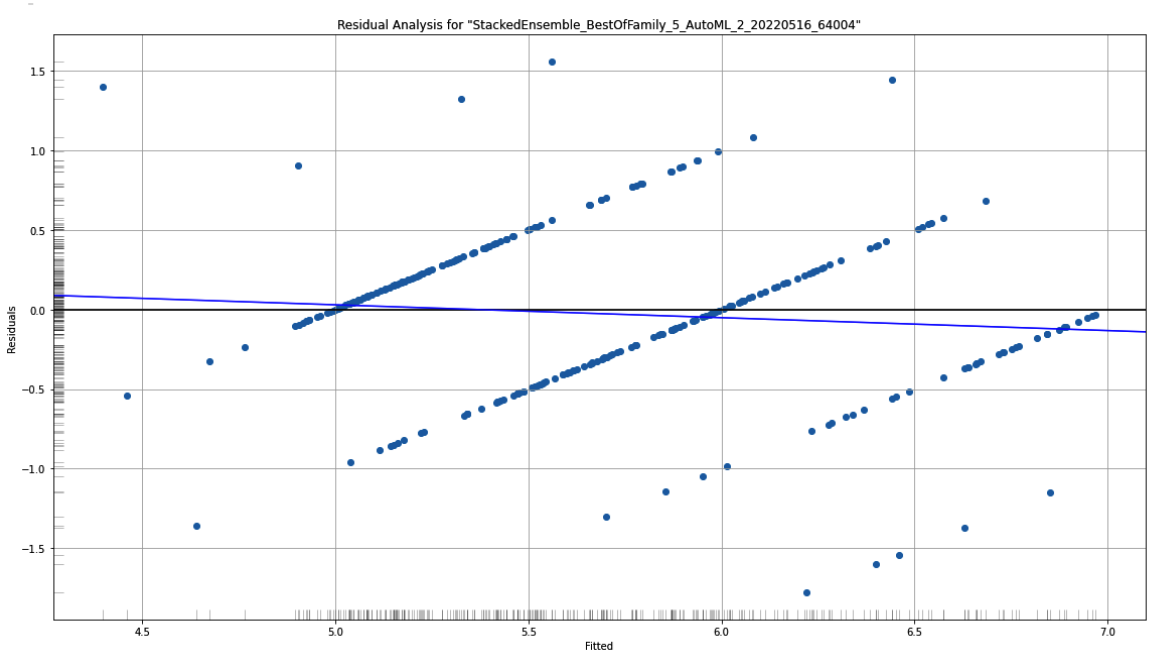
Leaderboard

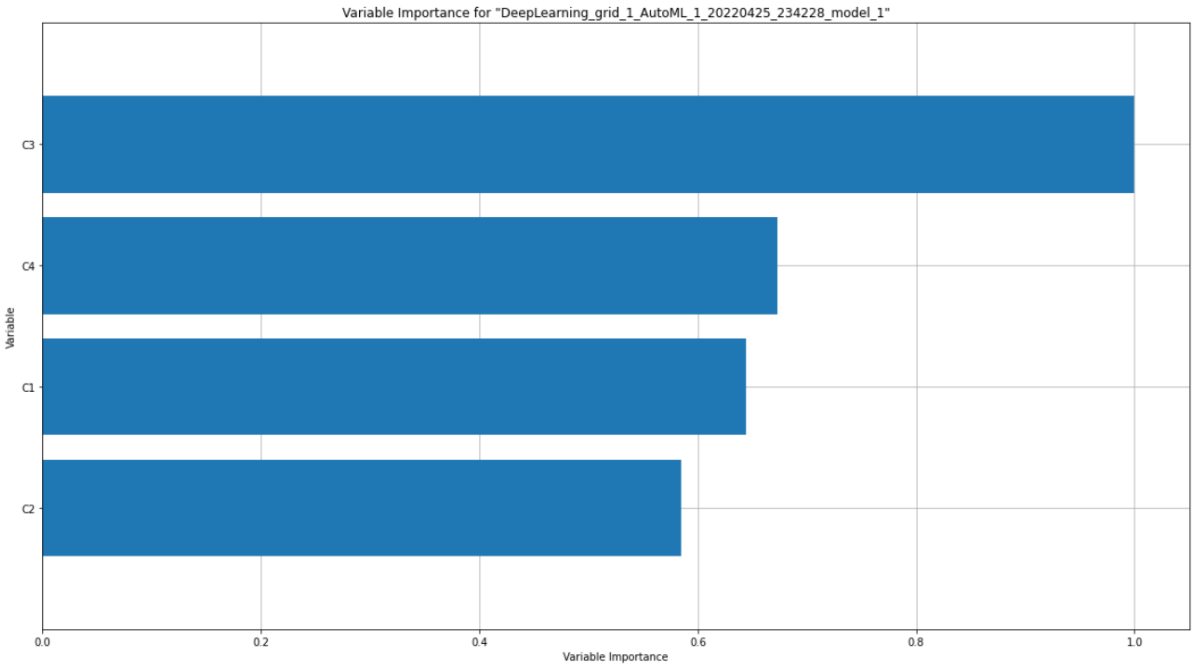
> Leaderboard shows models with their metrics. When provided with H2OAutoML object, the leaderboard shows 5-fold cross-validated metrics by default (depending on the H2OAutoML settings), otherwise it shows metrics computed on the frame. At most 20 models are shown by default.

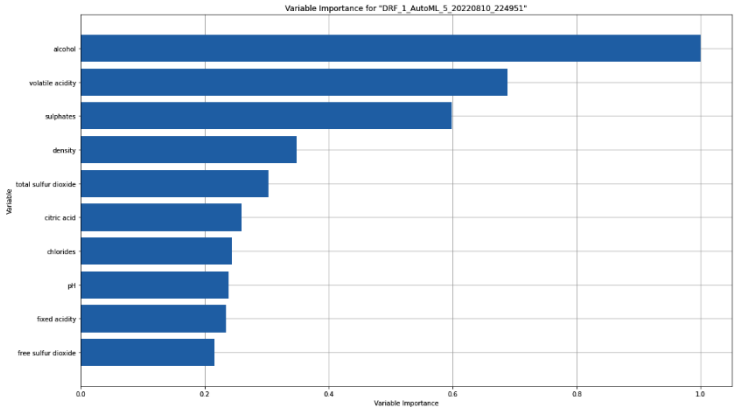
model_id	mean_per_class_error	logloss	rmse	mse	training_time_ms	predict_time_per_row_ms
DeepLearning_grid_1_AutoML_4_20220810_222639_model_1	0.0238359	0.0778748	0.150391	0.0226175	5017	0.006268
StackedEnsemble_AllModels_1_AutoML_4_20220810_222639	0.0314117	0.0972614	0.164382	0.0270214	1345	0.071765
GLM_1_AutoML_4_20220810_222639	0.0314117	0.0759111	0.157322	0.0247502	252	0.001568
StackedEnsemble_BestOfFamily_1_AutoML_4_20220810_222639	0.0389874	0.105026	0.176591	0.0311843	2339	0.044059

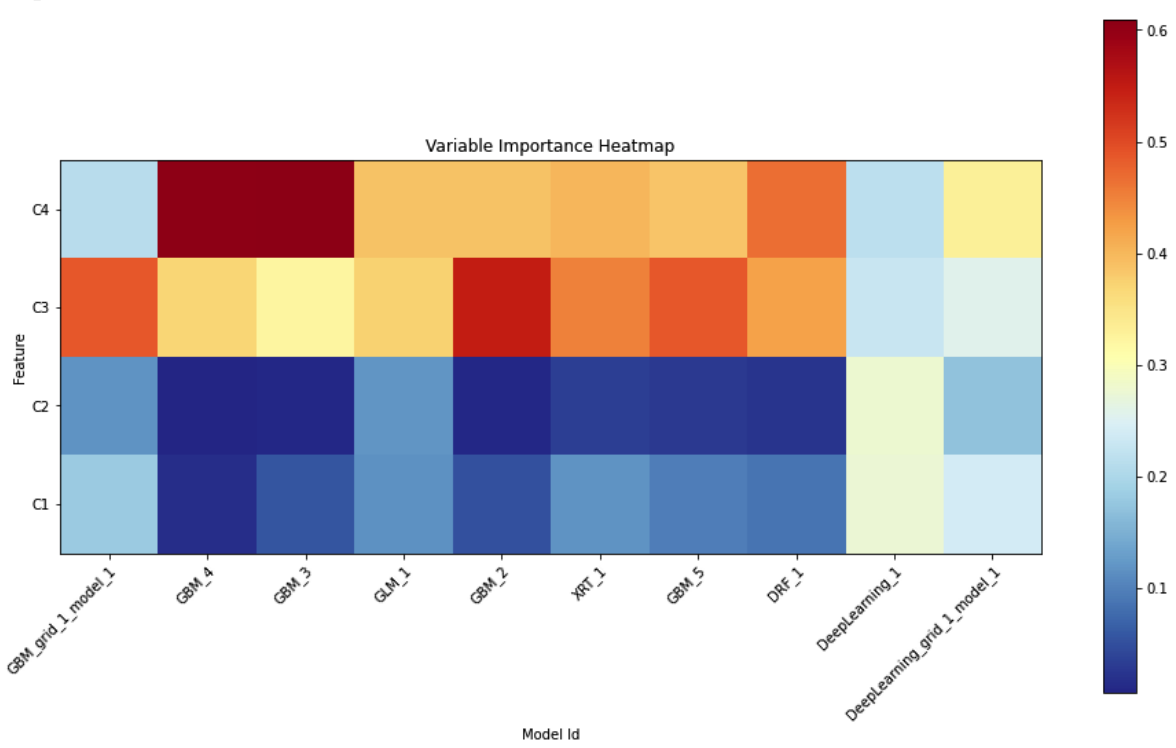


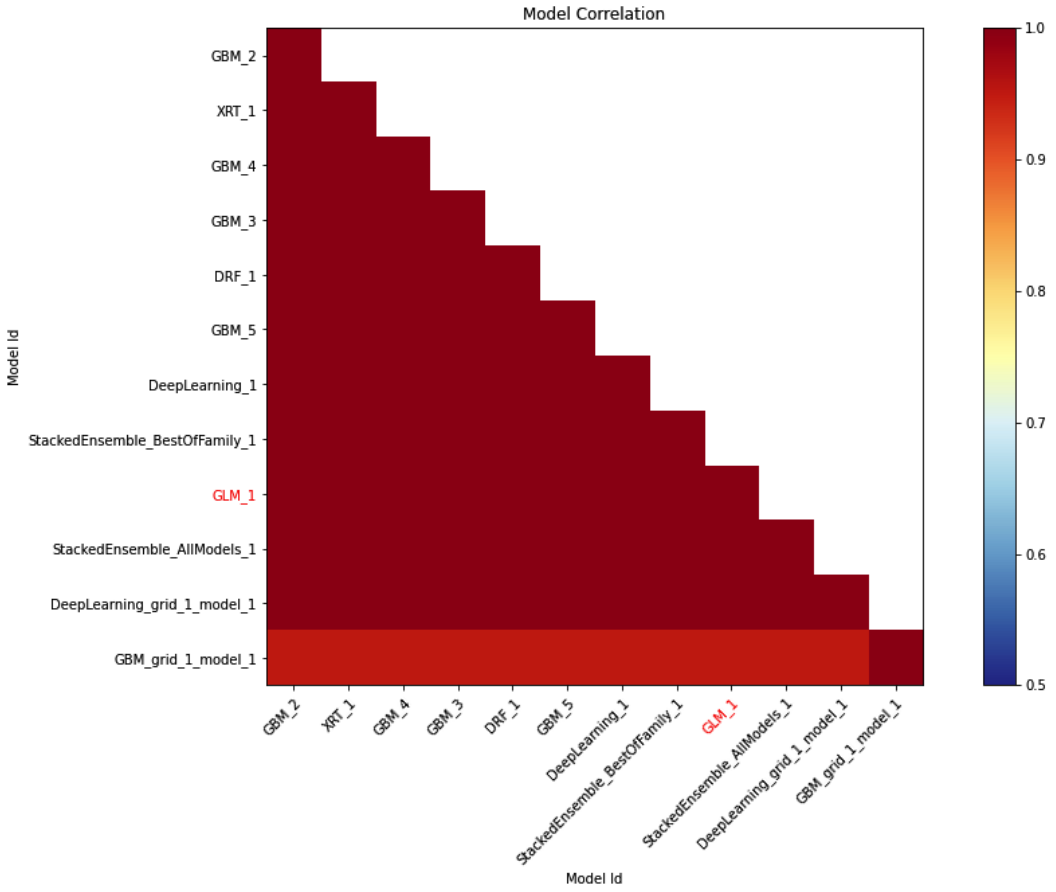

Dependent
Variable Y
Line of Regression
Independent Variable X

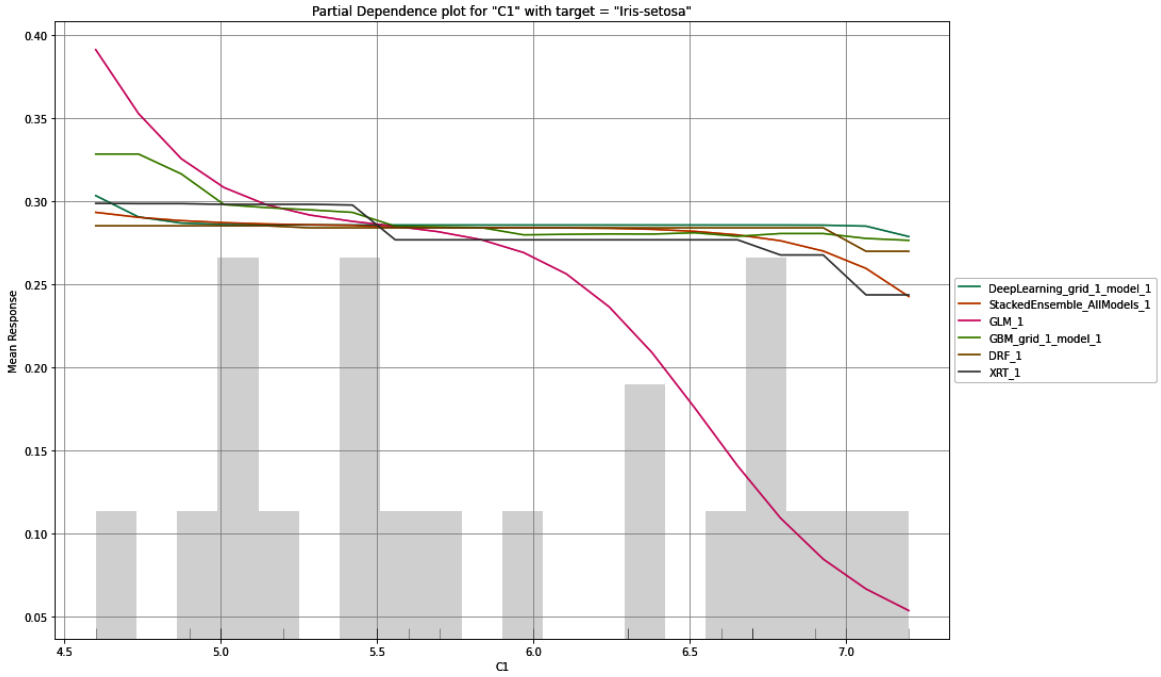

Residual Analysis for "StackedEnsemble_BestOfFamily_5_AutoML_2_20220516_64004"
Residuals
Fitted
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
4.5
5.0
5.5
6.0
6.5
7.0

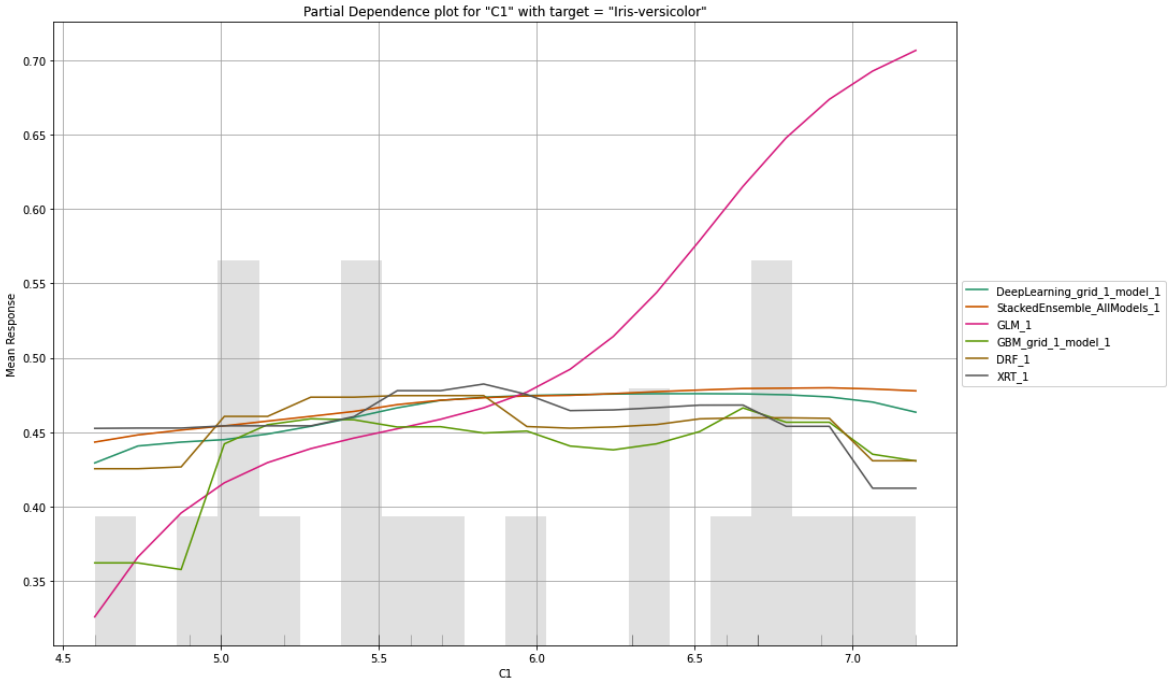

Variable Importance for "DeepLearning_grid_1_AutoML_1_20220425_234228_model_1"
C3
C4
C1
C2
Variable
0.0
0.2
0.4
0.6
0.8
1.0
Variable Importance

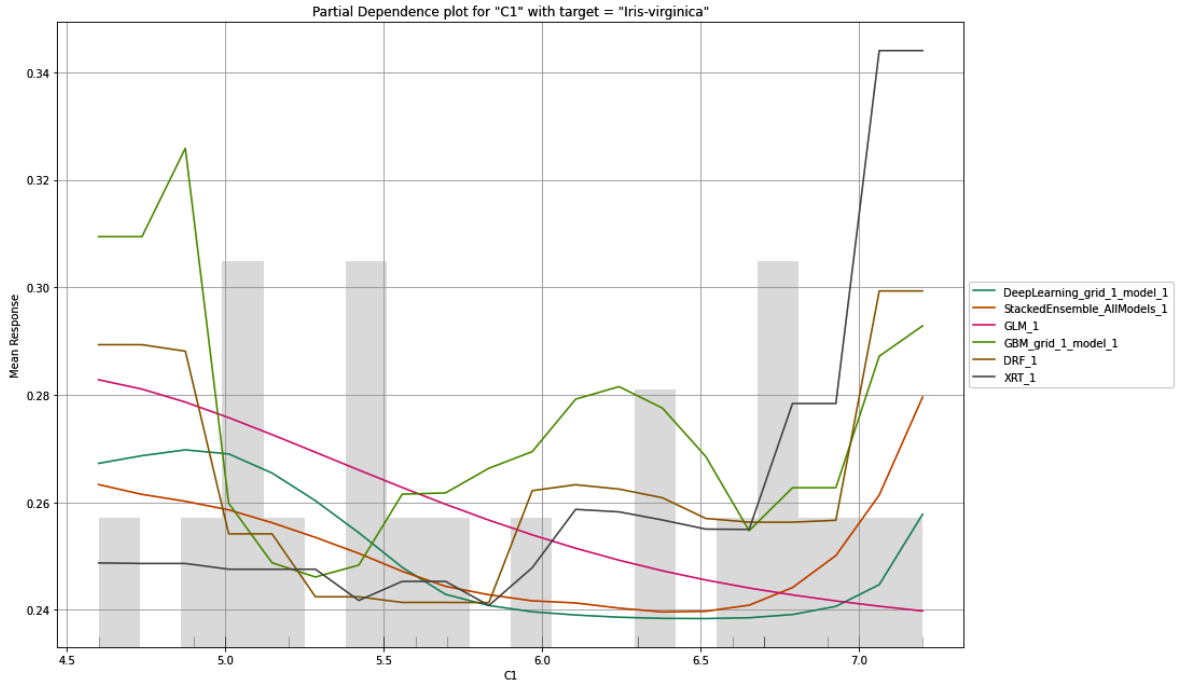

Variable Importance for "DRF_1_AutoML_5_20220810_224951"
alcohol
volatile acidity
sulphates
density
total sulfur dioxide
citric acid
chlorides
pH
fixed acidity
free sulfur dioxide
Variable
0.0
0.2
0.4
0.6
0.8
1.0
Variable Importance

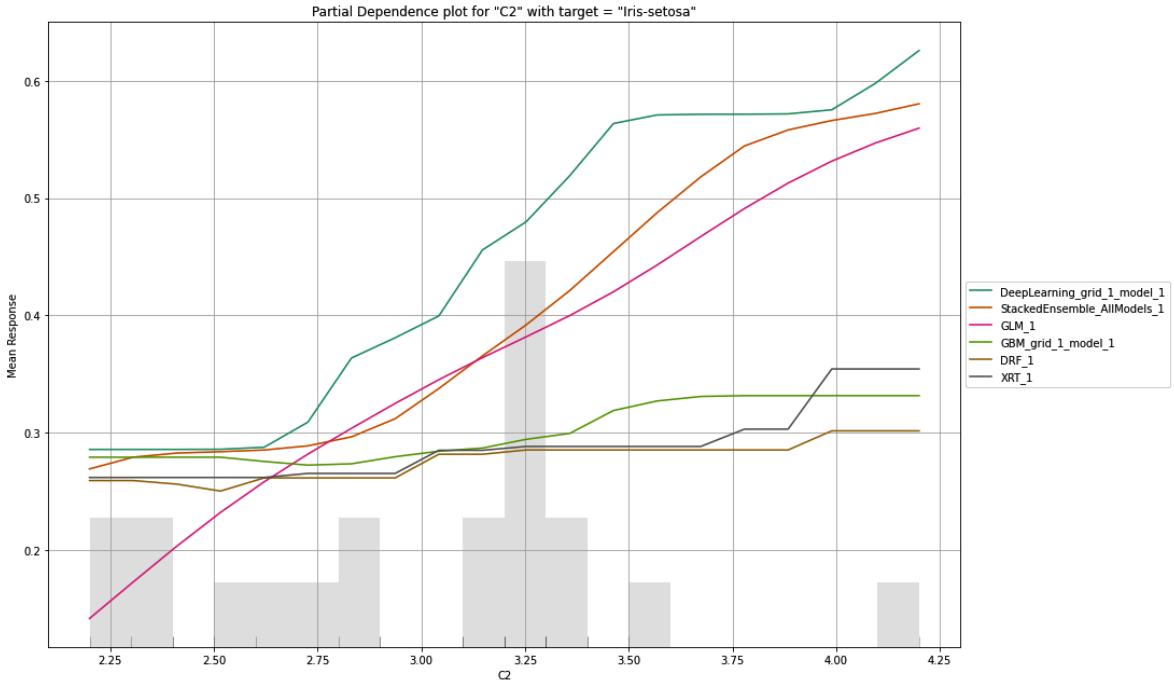

Variable Importance Heatmap
Feature
C4
C3
C2
C1
GBM_grid_1_model_1
GBM_4
GBM_3
GLM_1
GBM_2
XRT_1
GBM_5
DRF_1
DeepLearning_1
DeepLearning_grid_1_model_1
Model Id
0.6
0.5
0.4
0.3
0.2
0.1

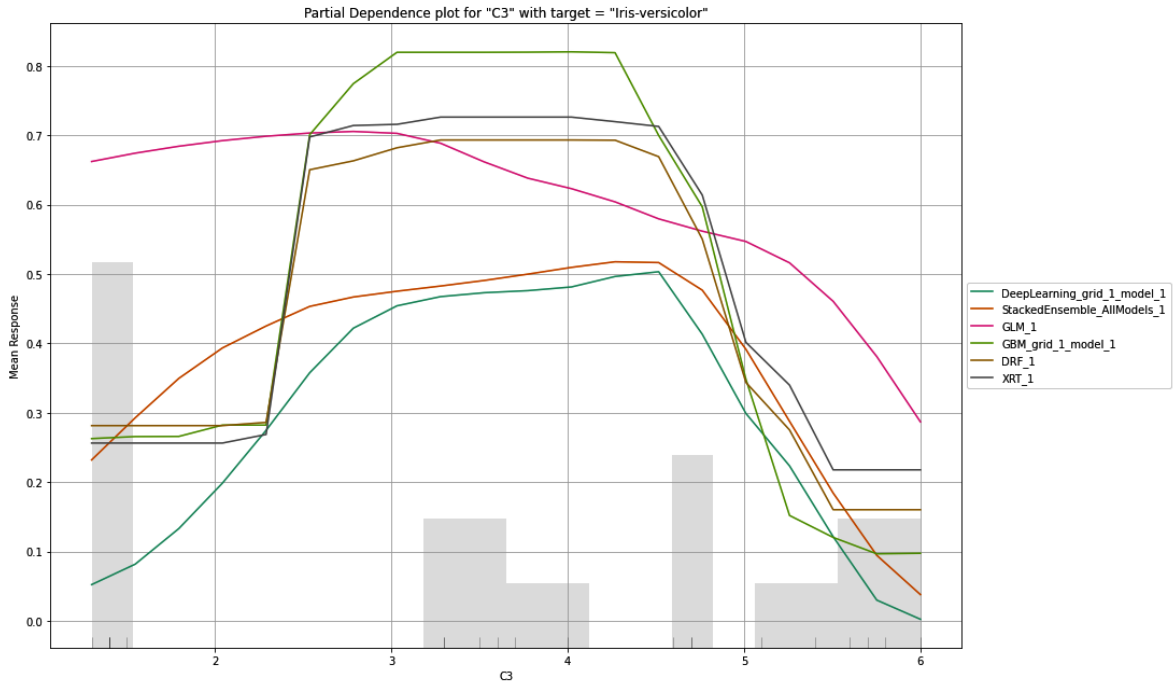

Model Correlation
Model Id
GBM_2
XRT_1
GBM_4
GBM_3
DRF_1
GBM_5
DeepLearning_1
StackedEnsemble_BestOfFamily_1
GLM_1
StackedEnsemble_AllModels_1
DeepLearning_grid_1_model_1
GBM_grid_1_model_1
Model Id
1.0
0.9
0.8
0.7
0.6
0.5

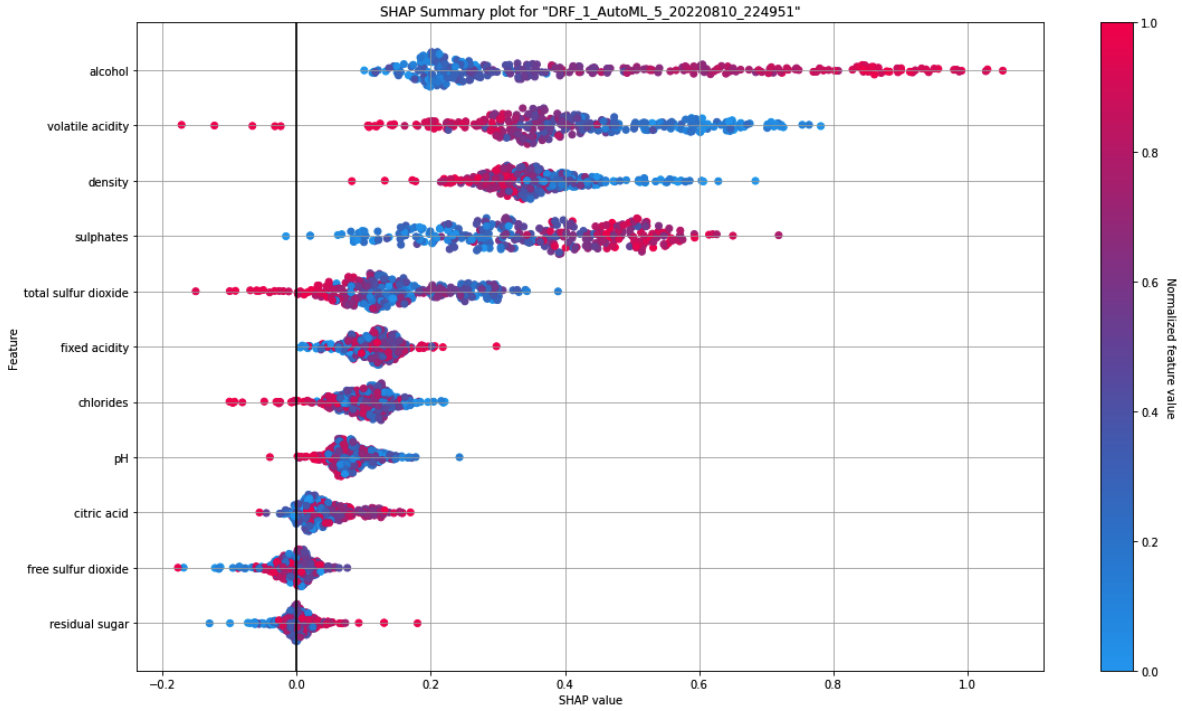

Partial Dependence plot for "C1" with target = "Iris-setosa"
Mean Response
C1
DeepLearning_grid_1_model_1
StackedEnsemble_AllModels_1
GLM_1
GBM_grid_1_model_1
DRF_1
XRT_1


Partial Dependence plot for "C1" with target = "Iris-versicolor"
Mean Response
C1
DeepLearning_grid_1_model_1
StackedEnsemble_AllModels_1
GLM_1
GBM_grid_1_model_1
DRF_1
XRT_1


Partial Dependence plot for "C1" with target = "Iris-virginica"
Mean Response
C1
DeepLearning_grid_1_model_1
StackedEnsemble_AllModels_1
GLM_1
GBM_grid_1_model_1
DRF_1
XRT_1


Partial Dependence plot for "C2" with target = "Iris-setosa"
Mean Response
C2
DeepLearning_grid_1_model_1
StackedEnsemble_AllModels_1
GLM_1
GBM_grid_1_model_1
DRF_1
XRT_1



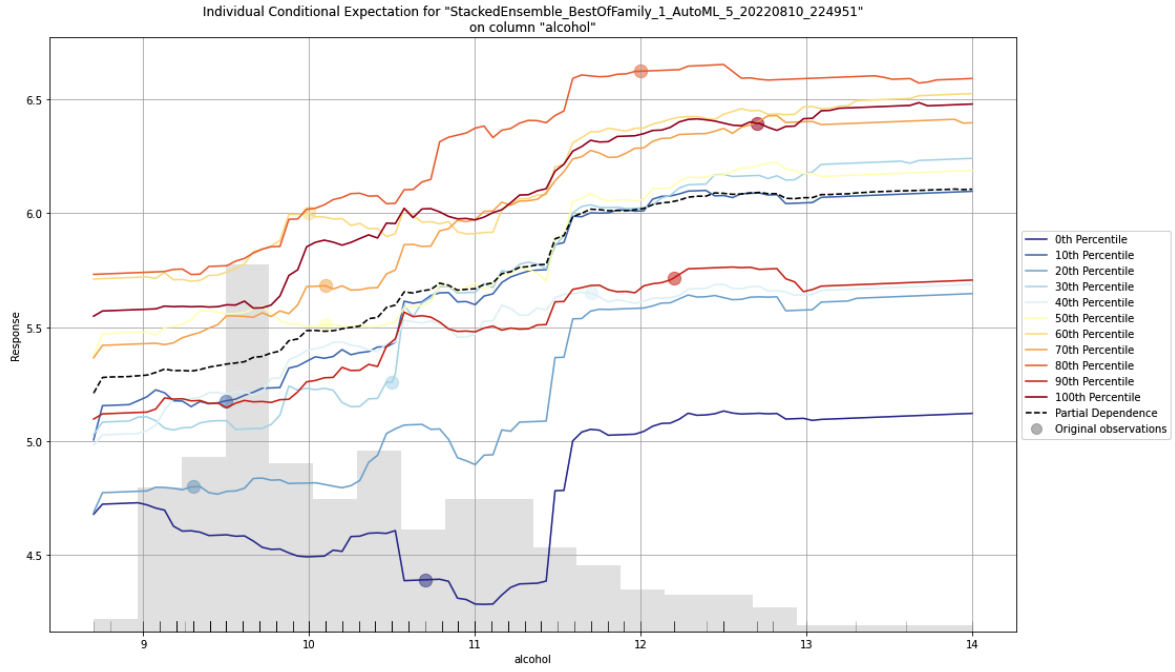


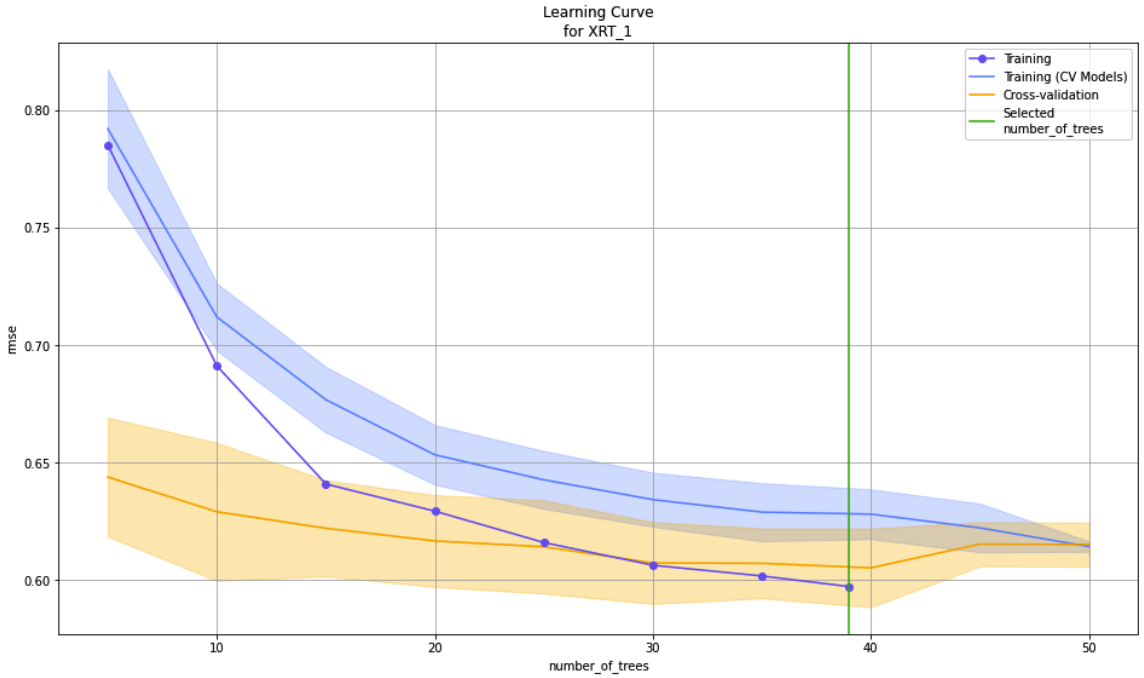
	Feature 1	Feature 2	Feature 3	Feature 4	…	Target
Observation 1	5	57	8	4	…	1
Observation 2	2	11	47	67	…	1
Observation 3	6	43	84	8	…	0
Observation 4	7	3	46	457	…	1
…	…	…	…	…	…	…
Observation N	3	5	27	37	…	1

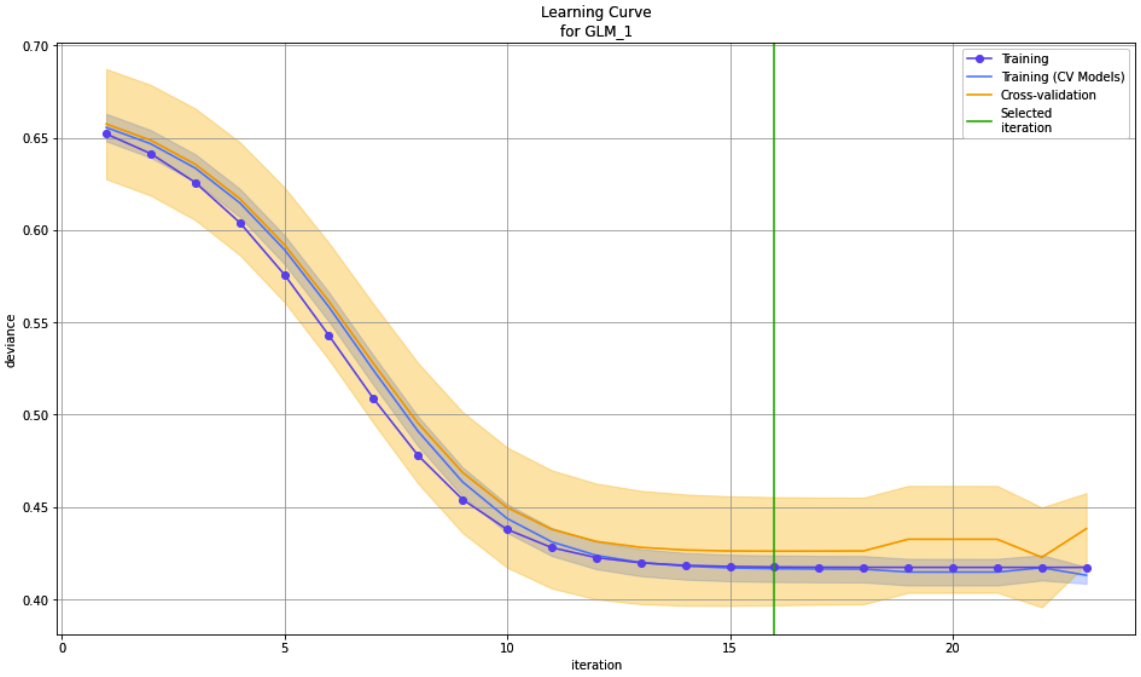
	Feature 1	Feature 2	Feature 3	Feature 4	…	Target
Observation 1	5	57	8	4	…	1
Observation 1	2	57	8	4	…	0.34
Observation 1	6	57	8	4	…	0.72
Observation 1	7	57	8	4	…	0.21
…	…	…	…	…	…	…
Observation 1	3	57	8	4	…	0.71

	Feature 1	Feature 2	Feature 3	Feature 4	…	Target
Observation 1	5	57	8	4	…	1
Observation 2	2	11	47	67	…	1
Observation 3	6	43	84	8	…	0
Observation 4	7	3	46	457	…	1
…	…	…	…	…	…	…
Observation N	3	5	27	37	…	1

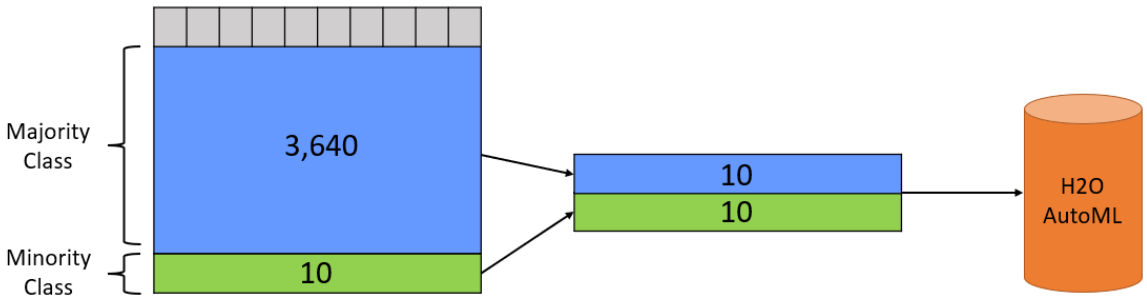
	Feature 1	Feature 2	Feature 3	Feature 4	…	Target
Observation 2	5	11	47	67	…	0.34
Observation 2	2	11	47	67	…	1
Observation 2	6	11	47	67	…	0.77
Observation 2	7	11	47	67	…	0.84
…	…	…	…	…	…	…
Observation 2	3	11	47	67	…	0.75

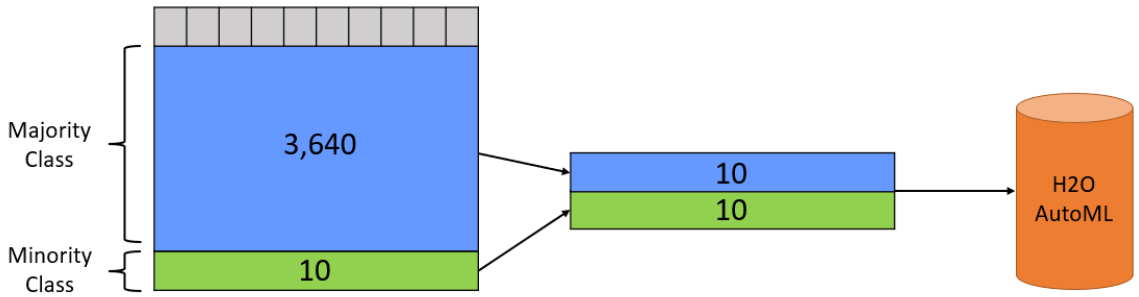



Learning Curve
for XRT_1
Training
Training (CV Models)
Cross-validation
Selected
number_of_trees
rmse
0.80
0.75
0.70
0.65
0.60
10
20
30
40
50
number_of_trees


Learning Curve
for GLM_1
Training
Training (CV Models)
Cross-validation
Selected
iteration
deviance
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0
5
10
15
20
iteration

Chapter 8: Exploring Optional Parameters for H2O AutoML

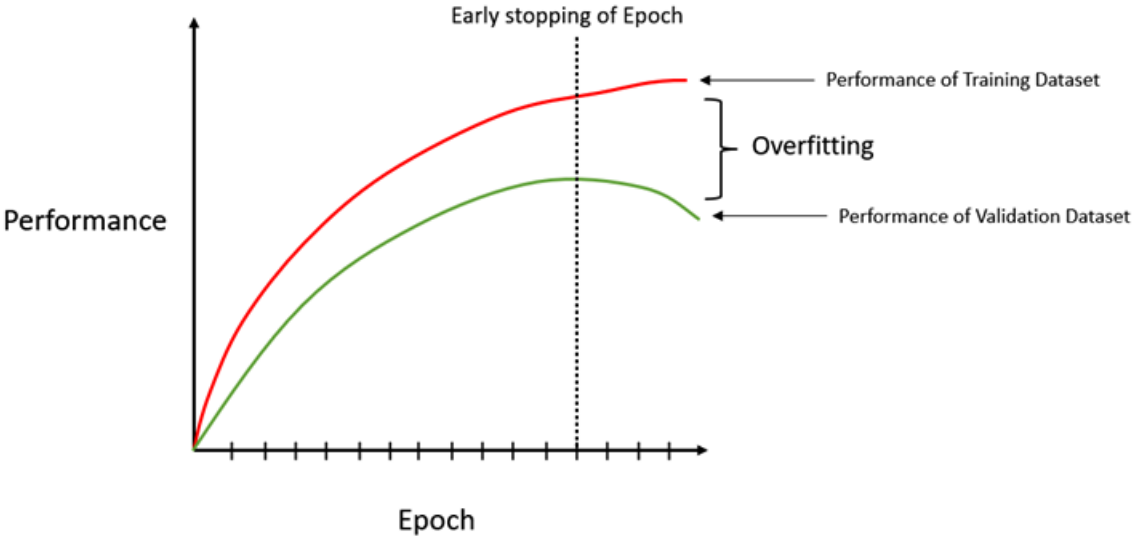




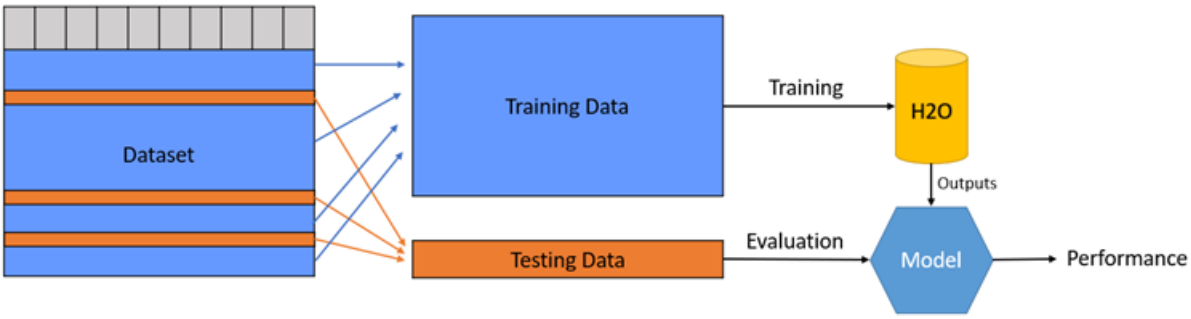
balance_classes ☑ Balance training data class counts via over/under-sampling (for imbalanced data).

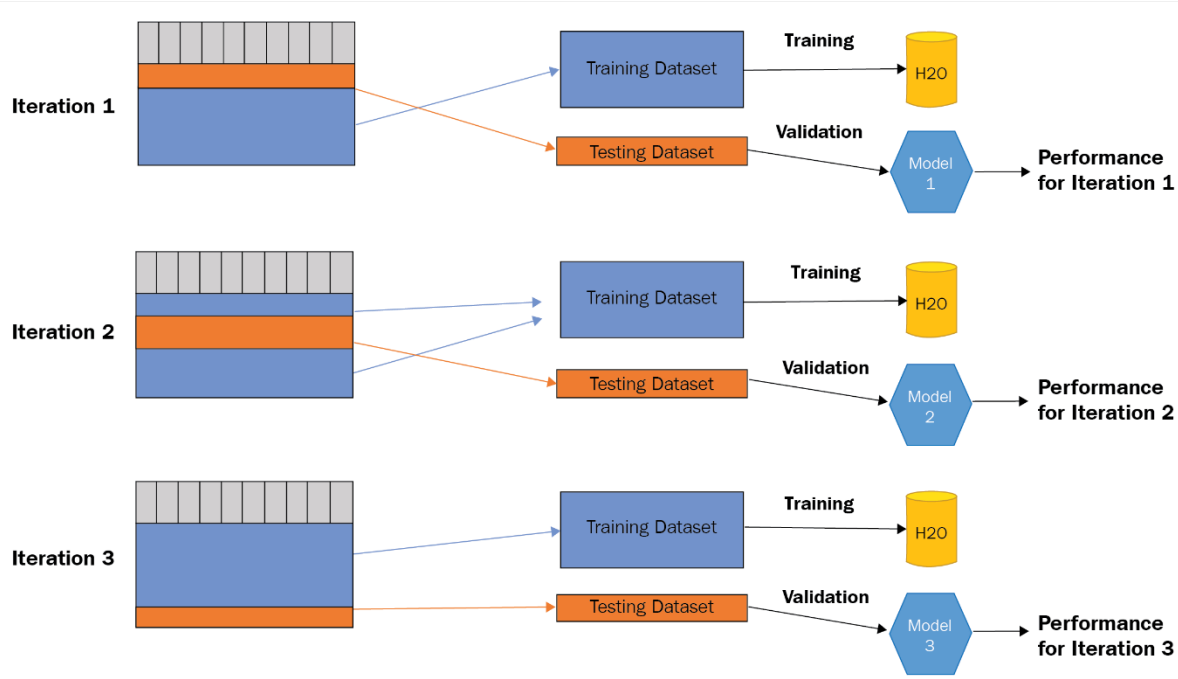
EXPERT

class_sampling_factors		Desired over/under-sampling ratios per class (in lexicographic order). If not specified, sampling factors will be automatically computed to obtain class balance during training. Requires balance_classes.
max_after_balance_size	5	Maximum relative size of the training data after balancing class counts (defaults to 5.0 and can be less than 1.0). Requires balance_classes.



stopping_rounds	3	Early stopping based on convergence of stopping_metric. Stop if simple moving average of length k of the stopping_metric does not improve for k:=stopping_rounds scoring events (0 to disable)
stopping_metric	AUTO	Metric to use for early stopping (AUTO: logloss for classification, deviance for regression)
stopping_tolerance	-1	Relative tolerance for metric-based stopping criterion (stop if relative improvement is not at least this much)





ADVANCED

nfolds	5	Number of folds for k-fold cross-validation (defaults to -1 (AUTO), otherwise it must be >=2 or use 0 to disable). Disabling prevents Stacked Ensembles from being built.
fold_column	(Choose...)	Fold column (contains fold IDs) in the training frame. These assignments are used to create the folds for cross-validation of the models.

EXPERT

keep_cross_validation_predictions	☑	Whether to keep the predictions of the cross-validation predictions. This needs to be set to TRUE if running the same AutoML object for repeated runs because CV predictions are required to build additional Stacked Ensemble models in AutoML.
keep_cross_validation_models	☑	Whether to keep the cross-validated models. Keeping cross-validation models may consume significantly more memory in the H2O cluster.
keep_cross_validation_fold_assignment	☑	Whether to keep cross-validation assignments.

Chapter 9: Exploring Miscellaneous Features in H2O AutoML

```
[1 2 2 0 2 1 0 2 0 1 1 2 2 2 0 0 2 2 0 0 1 2 0 1 1 2 1 1 1 2 0 1 1 0 1 0 0
 2 0 2 2 1 0 0 1]

Parse progress: |████████████████████████████████████████████████| (done) 100%
Parse progress: |████████████████████████████████████████████████| (done) 100%
AutoML progress: |
22:50:58.301: AutoML: XGBoost is not available; skipping it.

|
22:50:59.943: _min_rows param, The dataset size is too small to split for min_rows=100.0: must have at least 200.0 (weighted) r
ows, but have only 105.0.

|████████████████████████████████████████████████| (done) 100%
```

Out[17]:

```
H2OAutoMLClassifier
H2OAutoMLClassifier(max_models=10, seed=5, sort_metric='logloss')
```

model_id	logloss	mean_per_class_error	rmse	mse
GLM_1_AutoML_2_20220613_224223	0.0860788	0.0479303	0.162001	0.0262443
DeepLearning_grid_1_AutoML_2_20220613_224223_model_1	0.112033	0.038671	0.170577	0.0290966
StackedEnsemble_AllModels_1_AutoML_2_20220613_224223	0.11419	0.0378462	0.176666	0.0312109
StackedEnsemble_BestOfFamily_1_AutoML_2_20220613_224223	0.120517	0.0476502	0.18068	0.0326452
XRT_1_AutoML_2_20220613_224223	0.124044	0.0473856	0.190803	0.0364056
GBM_2_AutoML_2_20220613_224223	0.132214	0.0473856	0.200067	0.040027
DRF_1_AutoML_2_20220613_224223	0.133123	0.0473856	0.189798	0.0360231
GBM_3_AutoML_2_20220613_224223	0.147124	0.0473856	0.202676	0.0410774
GBM_4_AutoML_2_20220613_224223	0.155193	0.0473856	0.20303	0.041221
GBM_5_AutoML_2_20220613_224223	0.184402	0.0473856	0.210164	0.0441689

```
Parse progress: |████████████████████████████████████████████████| (done) 100%
glm prediction progress: |████████████████████████████████████████| (done) 100%
[1 2 2 0 2 1 0 2 0 1 1 1 2 2 0 0 2 2 0 0 1 2 0 1 1 2 1 1 1 2 0 1 1 0 1 0 0
 2 0 2 2 1 0 0 1]

Parse progress: |████████████████████████████████████████████████| (done) 100%
Parse progress: |████████████████████████████████████████████████| (done) 100%
AutoML progress: |
23:41:40.934: AutoML: XGBoost is not available; skipping it.

|████████████████████████████████████████████████| (done) 100%
```

```
H2OAutoMLRegressor
H2OAutoMLRegressor(max_models=10, max_runtime_secs_per_model=30, seed=5)
```

```
Parse progress: |████████████████████████████████████████████████| (done) 100%
stackedensemble prediction progress: |████████████████████████████| (done) 100%
[5.7050527  4.71579041 4.99625922 6.88799804 5.30901742 5.08507949
 5.22402198 6.80312639 5.73784549 5.05679221 6.60598185 5.16579878
 5.14112093 5.8436239  6.03287821 5.95839962 6.33995783 6.05639611
 5.76648316 5.75068268 4.90408778 5.29082171 5.17198907 5.217705
 5.97337232 5.43873863 5.19827529 5.31316769 6.2793359  6.38171303
 5.58378112 5.36376057 5.42070109 4.84603301 5.9211492  6.17345933
 6.09528051 5.68300139 5.10375214 5.33199332 6.11587125 5.05826719
 5.52626887 5.12766708 6.54879736 5.67341307 5.24367939 5.7050527
 5.72492442 5.20803101 5.72238569 5.23847352 4.94162759 5.72515368
 5.99710415 5.54065627 5.96809731 5.41532481 6.51548553 6.43069508
 5.41738151 5.50117357 6.77200607 5.60270773 5.02891041 6.637265
 6.61085364 5.51616503 5.77759467 5.88147359 5.12760658 6.87250045
 5.61076927 5.45198623 5.4550793  5.02793331 4.86960596 5.24755881
 5.14942367 5.03725484 5.36312455 4.98812026 5.49151424 5.16758049
 5.04899969 5.48321742 5.63502592 5.01813754 5.63146962 5.06783828
 6.08415337 6.10573536 5.89912238 5.51682489 5.8902318  5.61761149
 5.83461759 5.45310272 5.81230101 6.56012283 5.68300139 5.72515368
 5.42014689 6.37008337 5.44917672 5.65400694 5.37099615 5.05670911
 5.50945522 4.7857324  6.07391301 5.6309158  5.38736267 5.94324949
 6.12127789 5.25270546 5.09663222 4.94148322 5.24268557 5.76740135
 5.02992971 4.98601981 5.23574556 6.31680286 4.96510575 5.52742118
 5.05826719 5.69985666 5.16785547 6.83281514 4.81951631 4.98821234
```

timestamp	level	stage	message	name	value
00:13:05.18	INFO	Workflow	Project: AutoML_2_20220615_01305		
00:13:05.19	INFO	Validation	5-fold cross-validation will be used.		
00:13:05.19	INFO	Validation	Setting stopping tolerance adaptively based on the training frame: 0.05		
00:13:05.19	INFO	Validation	Build control seed: 5		
00:13:05.19	INFO	DataImport	training frame: Frame key: AutoML_2_20220615_01305_training_iris_data1.hex cols: 5 rows: 150 chunks: 1 size: 1982 checksum: -5547756281708519194		
00:13:05.19	INFO	DataImport	validation frame: NULL		
00:13:05.19	INFO	DataImport	leaderboard frame: NULL		
00:13:05.19	INFO	DataImport	blending frame: NULL		
00:13:05.19	INFO	DataImport	response column: C5		
00:13:05.19	INFO	DataImport	fold column: null		

```
{'creation_epoch': '1655248385',
 'start_epoch': '1655248385',
 'start_GLM_def_1': '1655248385',
 'start_GBM_def_5': '1655248386',
 'start_DRF_def_1': '1655248386',
 'start_GBM_def_2': '1655248386',
 'start_GBM_def_3': '1655248386',
 'start_GBM_def_4': '1655248386',
 'start_DRF_XRT': '1655248386',
 'start_GBM_def_1': '1655248387',
 'start_DeepLearning_def_1': '1655248387',
 'start_GBM_grid_1': '1655248387',
 'start_DeepLearning_grid_1': '1655248387',
 'start_StackedEnsemble_best_of_family_xglm': '1655248395',
 'start_StackedEnsemble_all_xglm': '1655248396',
 'stop_epoch': '1655248397',
 'duration_secs': '12'}
```

Chapter 10: Working with Plain Old Java Objects (POJOs)

```
CS getModel "DRF_1_AutoML_1_20220619_141718"
```

Model

Model ID: DRF_1_AutoML_1_20220619_141718
Algorithm: Distributed Random Forest
Actions: Refresh | Predict... | Download POJO | Download Model Deployment Package (MOJO) | Export | Inspect | Delete | Download Gen Model

```
Predicted Class of Iris flower is: Iris-setosa
Class probabilities are:
0.9975669100660882
0.0
0.002433089933911768

Process finished with exit code 0
```

Chapter 11: Working with Model Object

model_id	mean_per_class_error	logloss	rmse	mse
GLM_1_AutoML_4_20220801_225630	0.0266667	0.067412	0.148337	0.0220039
GBM_lr_annealing_selection_AutoML_4_20220801_225630_select_model	0.0333333	0.140917	0.183485	0.0336669
GBM_grid_1_AutoML_4_20220801_225630_model_3	0.0466667	0.164239	0.206905	0.0428097
DRF_1_AutoML_4_20220801_225630	0.0533333	0.121106	0.193738	0.0375344
GBM_2_AutoML_4_20220801_225630	0.0533333	0.162145	0.210702	0.0443953
XRT_1_AutoML_4_20220801_225630	0.0533333	0.123873	0.197378	0.0389581
GBM_5_AutoML_4_20220801_225630	0.06	0.203821	0.224769	0.0505213
GBM_3_AutoML_4_20220801_225630	0.06	0.165065	0.210117	0.0441491
GBM_4_AutoML_4_20220801_225630	0.0666667	0.177578	0.221646	0.0491269

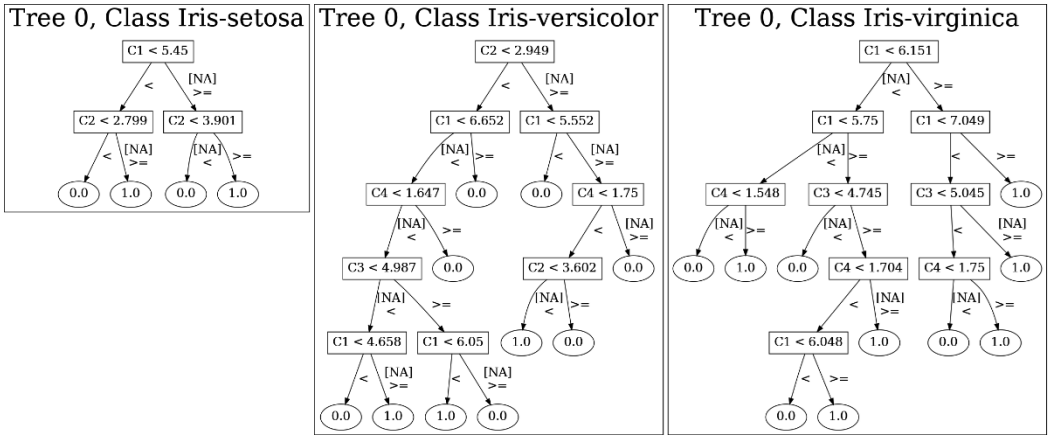
```
getModel "DRF_1_AutoML_1_20220619_141718"
```

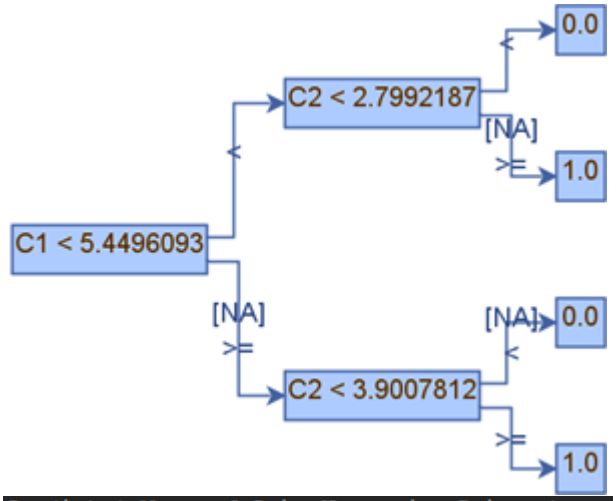
Model

Model ID: DRF_1_AutoML_1_20220619_141718

Algorithm: Distributed Random Forest

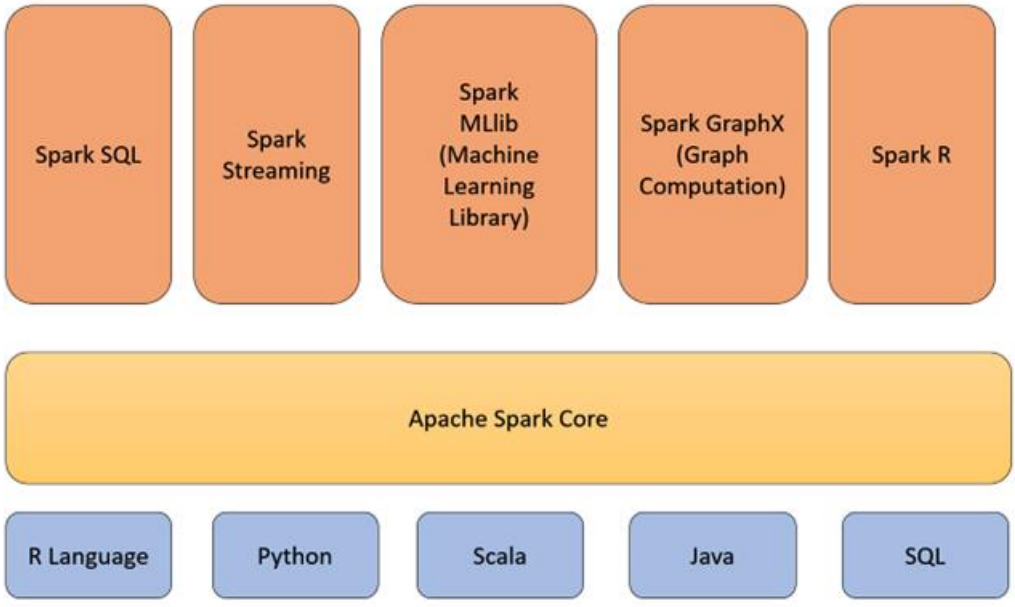
Actions: Refresh | Predict... | Download POJO | Download Model Deployment Package (MOJO) | Export | Inspect | Delete | Download Gen Model

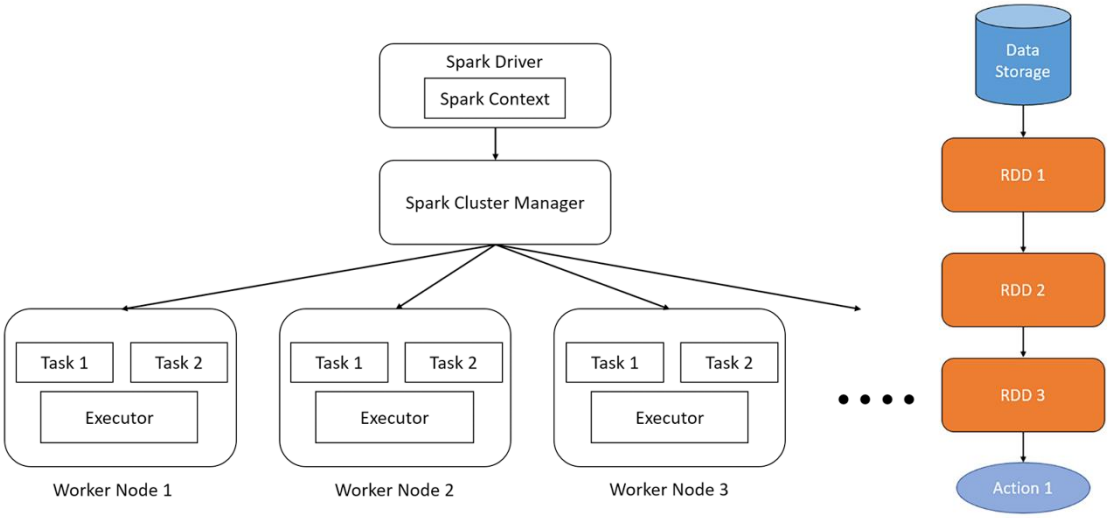


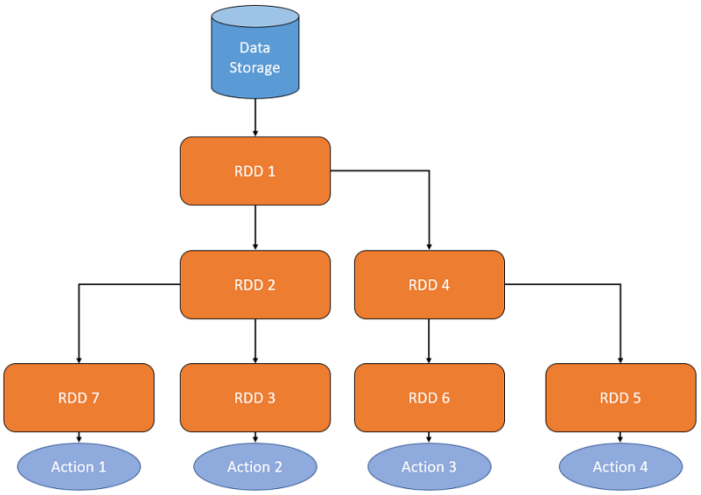


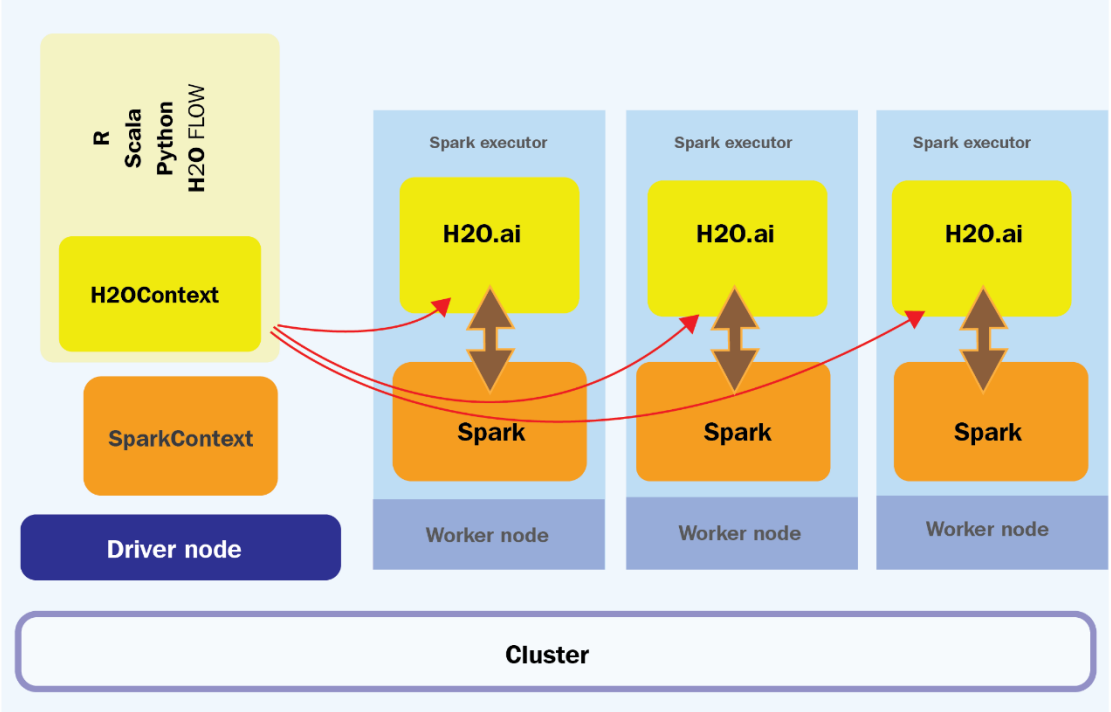

```
Predicted Class of Iris flower is: Iris-setosa
Class probabilities are:
0.9971308697337545
0.0
0.0028691302662455114

Process finished with exit code 0
```

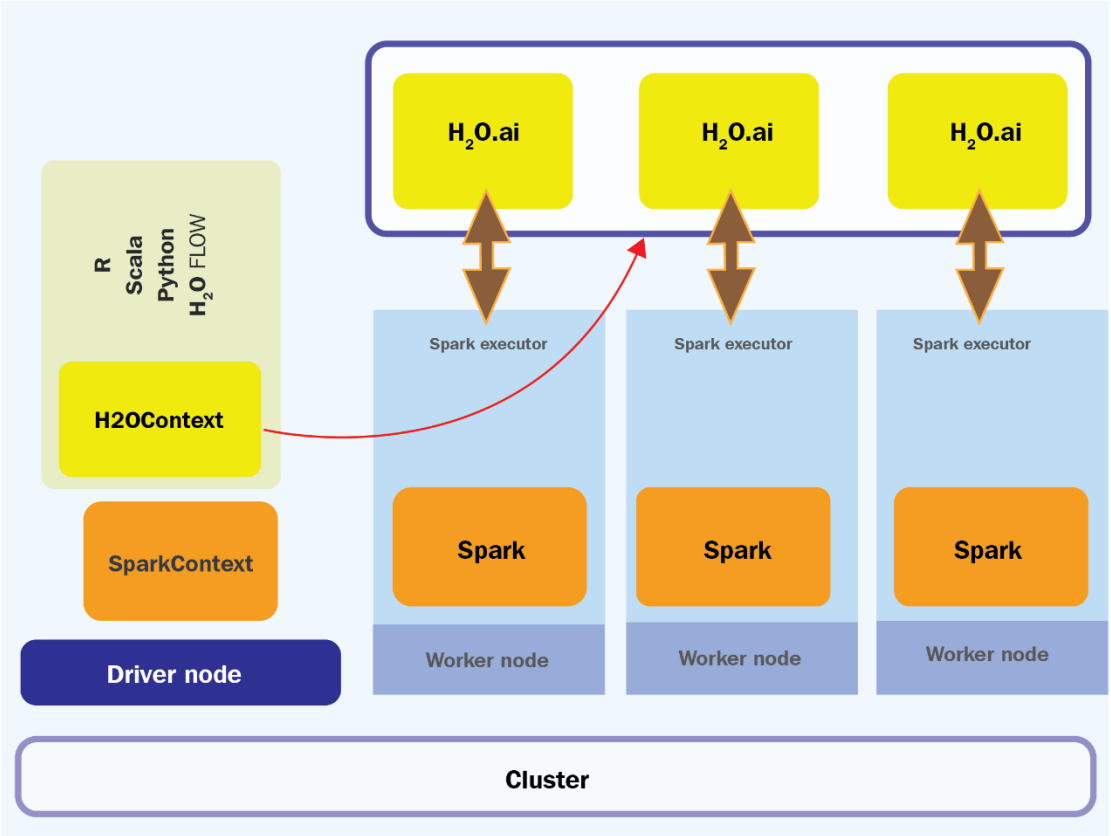
Chapter 12: Working with H2O AutoML and Apache Spark

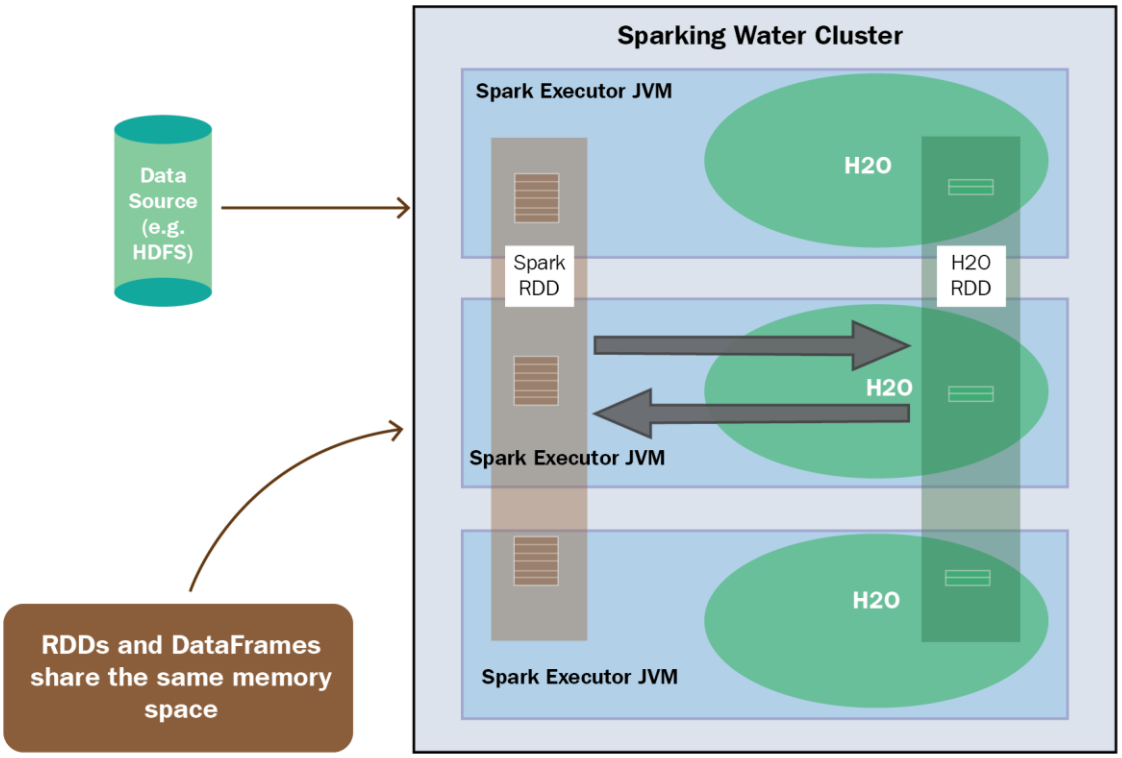


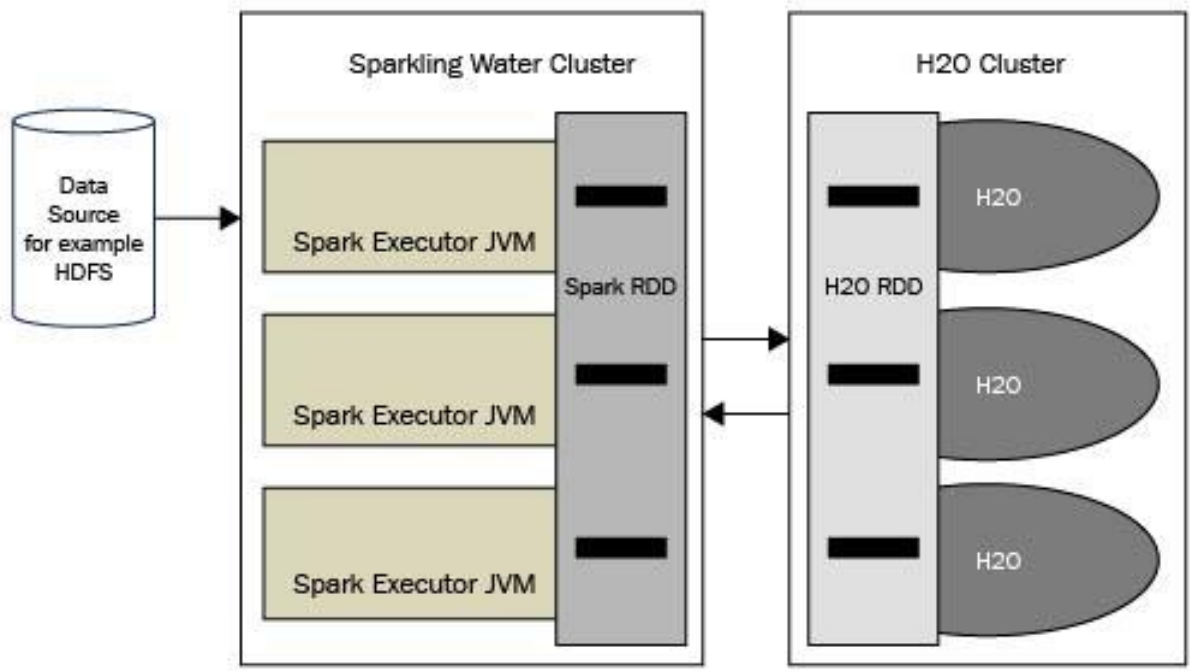





R
Scala
Python
H2O FLOW
H2OContext
SparkContext
Driver node
Spark executor
H2O.ai
Spark
Worker node
Spark executor
H2O.ai
Spark
Worker node
Spark executor
H2O.ai
Spark
Worker node
Cluster


R
Scala
Python
H2O FLOW
H2OContext
SparkContext
Driver node
H2O.ai
H2O.ai
H2O.ai
Spark executor
Spark
Worker node
Spark executor
Spark
Worker node
Spark executor
Spark
Worker node
Cluster


Sparking Water Cluster
Spark Executor JVM
H2O
Data Source (e.g. HDFS)
Spark RDD
H2O RDD
H2O
Spark Executor JVM
H2O
Spark Executor JVM
RDDs and DataFrames share the same memory space


Sparkling Water Cluster
H2O Cluster
Data Source for example HDFS
Spark Executor JVM
Spark RDD
H2O RDD
H2O
Spark Executor JVM
H2O
Spark Executor JVM
H2O

```
Sparkling Water Context:
 * Sparkling Water Version: 3.36.1.2-1-3.2
 * H2O name: sparkling-water-salil_local-1657573153324
 * cluster size: 1
 * list of used nodes:
  (executorId, host, port)
  ------------------------
  (0,10.0.2.15,54321)
  ------------------------

  Open H2O Flow in browser: http://10.0.2.15:54323 (CMD + click in Mac OSX)


h2oContext: ai.h2o.sparkling.H2OContext =

Sparkling Water Context:
 * Sparkling Water Version: 3.36.1.2-1-3.2
 * H2O name: sparkling-water-salil_local-1657573153324
 * cluster size: 1
 * list of used nodes:
  (executorId, host, port)
  ------------------------
  (0,10.0.2.15,54321)
  ------------------------

  Open H2O Flow in browser: http://10.0.2.15:54323 (CMD + click in Mac OSX)


scala>
```

Cement	Blast Furnace Slag	Fly Ash	Water	Superplasticizer	Coarse Aggregate	Fine Aggregate	Age	Concrete compressive strength
540.0	0.0	0.0	162.0	2.5	1040.0	676.0	28	79.99
540.0	0.0	0.0	162.0	2.5	1055.0	676.0	28	61.89
332.5	142.5	0.0	228.0	0.0	932.0	594.0	270	40.27
332.5	142.5	0.0	228.0	0.0	932.0	594.0	365	41.05
540.0	0.0	0.0	162.0	2.5	1040.0	676.0	28	79.99
540.0	0.0	0.0	162.0	2.5	1055.0	676.0	28	61.89
...	...	...	...	...	...	...	...	...

```
salil@salil-VirtualBox:~/softwares/sparkling-water-3.36.1.2-1-3.2$ ./bin/sparkling-shell

Using Spark defined in the SPARK_HOME=/home/salil/softwares/spark-3.2.1-bin-hadoop3.2 environmental property


-----
  Spark master (MASTER)     : local[*]
  Spark home   (SPARK_HOME) : /home/salil/softwares/spark-3.2.1-bin-hadoop3.2
  H2O build version         : 3.36.1.2 (zumbo)
  Sparkling Water version   : 3.36.1.2-1-3.2
  Spark build version       : 3.2.1
  Scala version             : 2.12
----

22/07/07 23:13:52 WARN Utils: Your hostname, salil-VirtualBox resolves to a loopback address: 127.0.1.1; using 10.0.2.15 instead (on interface enp0s3)
22/07/07 23:13:52 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform (file:/home/salil/softwares/spark-3.2.1-bin-hadoop3.2/jars/spark-unsafe_2.12-3.2.1.jar) to constructor ja
va.nio.DirectByteBuffer(long,int)
WARNING: Please consider reporting this to the maintainers of org.apache.spark.unsafe.Platform
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
22/07/07 23:13:54 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Spark context Web UI available at http://10.0.2.15:4040
Spark context available as 'sc' (master = local[*], app id = local-1657232049726).
Spark session available as 'spark'.
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /___/ .__/\_,_/_/ /_/\_\   version 3.2.1
      /_/

Using Scala version 2.12.15 (OpenJDK 64-Bit Server VM, Java 11.0.15)
Type in expressions to have them evaluated.
Type :help for more information.

scala>
```

```
salil@salil-VirtualBox:~/softwares/sparkling-water-3.36.1.2-1-3.2$ bin/pysparkling

Using Spark defined in the SPARK_HOME=/home/salil/softwares/spark-3.2.1-bin-hadoop3.2 environmental property

Python 3.10.4 (main, Apr  2 2022, 09:04:19) [GCC 11.2.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
22/07/12 01:02:05 WARN Utils: Your hostname, salil-VirtualBox resolves to a loopback address: 127.0.1.1; using 10.0.2.15 instead (on interface enp0s3)
22/07/12 01:02:05 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform (file:/home/salil/softwares/spark-3.2.1-bin-hadoop3.2/jars/spark-unsafe_2.12-3.2.1.jar) to constructor ja
va.nio.DirectByteBuffer(long,int)
WARNING: Please consider reporting this to the maintainers of org.apache.spark.unsafe.Platform
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
22/07/12 01:02:06 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /___/ .__/\_,_/_/ /_/\_\   version 3.2.1
      /_/

Using Python version 3.10.4 (main, Apr  2 2022 09:04:19)
Spark context Web UI available at http://10.0.2.15:4040
Spark context available as 'sc' (master = local[*], app id = local-1657584129365).
SparkSession available as 'spark'.
>>>
```

```
h2oContext: ai.h2o.sparkling.H2OContext =

Sparkling Water Context:
 * Sparkling Water Version: 3.36.1.2-1-3.2
 * H2O name: sparkling-water-salil_local-1657232049726
 * cluster size: 1
 * list of used nodes:
  (executorId, host, port)
  ------------------------
  (0,10.0.2.15,54321)
  ------------------------

  Open H2O Flow in browser: http://10.0.2.15:54323 (CMD + click in Mac OSX)
```

```
Returning leader model and printing info about it below.
Model Details
==============
H2OStackedEnsemble
Model Key: StackedEnsemble_AllModels_1_AutoML_2_20220709_155143_e0a6c5512727

Model summary

Training metrics
RMSLE: 0.06461275612484682
Nobs: 711.0
RMSE: 1.664316734598618
ResidualDeviance: 1969.4345872692202
NullDeviance: 201217.88337876744
MAE: 1.180679697214416
MeanResidualDeviance: 2.7699501930650072
ScoringTime: 1.657378357095E12
MSE: 2.7699501930650072
R2: 0.9902124276719378
NullDegreesOfFreedom: 710.0
AIC: 2760.118254384428
ResidualDegreesOfFreedom: 703.0

Cross validation metrics
RMSLE: 0.1460912336455651
Nobs: 711.0
RMSE: 4.324493096662628
ResidualDeviance: 13296.582026131815
NullDeviance: 201708.75554137037
MAE: 2.9346239510372185
MeanResidualDeviance: 18.701240543082722
ScoringTime: 1.657378356816E12
MSE: 18.701240543082722
R2: 0.9339194816938675
NullDegreesOfFreedom: 710.0
AIC: 4119.957985290266
ResidualDegreesOfFreedom: 702.0
```

	model_id	rmse	mse	mae	rmsle	mean_residual_deviance
0	StackedEnsemble_AllModels_1_AutoML_2_20220709_155143	4.324493096662628	18.701240543082722	2.9346239510372185	0.1460912336455651	18.701240543082722
1	StackedEnsemble_BestOfFamily_1_AutoML_2_20220709_155143	4.384260179975179	19.221737325715992	2.9686839361381865	0.14812221198292888	19.221737325715992
2	GBM_4_AutoML_2_20220709_155143	4.573584379622189	20.917674077524083	3.1189841752490692	0.14510419291520354	20.917674077524083
3	XGBoost_3_AutoML_2_20220709_155143	4.594722541302787	21.11147523155594	3.14733816048264	0.16959885368039 04	21.11147523155594
4	GBM_3_AutoML_2_20220709_155143	4.602836781243628	21.186106434769197	3.216627946091281	0.14846442209051083	21.186106434769197
5	XGBoost_2_AutoML_2_20220709_155143	4.795345366837029	22.995337187245365	3.2627855644950383	0.163751109572822	22.995337187245365
6	GBM_2_AutoML_2_20220709_155143	4.805228403927745	23.09022001391398	3.357554479064718	0.15569704766858766	23.09022001391398
7	XGBoost_1_AutoML_2_20220709_155143	4.883657455766908	23.850110145267706	3.4043382679527485	0.16722619939781377	23.850110145267706
8	XRT_1_AutoML_2_20220709_155143	5.925199691510438	35.107991384275394	4.320396072254607	0.2074160378479092	35.107991384275394
9	DRF_1_AutoML_2_20220709_155143	5.946360562106173	35.35920393457163	4.299032302631552	0.2060104350171061	35.35920393457163
10	GBM_1_AutoML_2_20220709_155143	6.787876516135789	46.07526759830773	5.291743594342693	0.22030051984937776	46.07526759830773
11	GLM_1_AutoML_2_20220709_155143	10.520585472187229	110.68271867759698	8.317403167170115	0.3410797993409181	110.68271867759698

Cement	Blast Furnace Slag	Fly Ash (component 3)(kg in a m^3 mixture)	Water	Superplasticizer	Coarse Aggregate	Fine Aggregate	Age	Concrete compressive strength	detailed_prediction	prediction
102.0	153.0	0.0	192.0	0.0	887.0	942.0	28.0	17.28	{18.702647076809}	18.70264707680947
108.3	162.4	0.0	203.5	0.0	938.2	849.0	28.0	20.59	{21.148178536290}	21.1481785362902
116.0	173.0	0.0	192.0	0.0	909.8	891.9	7.0	10.09	{10.275402528409}	10.275402528409455
122.6	183.9	0.0	203.5	0.0	958.2	800.1	3.0	3.32	{3.3479052271691}	3.3479052271691123
122.6	183.9	0.0	203.5	0.0	958.2	800.1	7.0	10.35	{9.946223757837}	9.94622375783771
133.0	200.0	0.0	192.0	0.0	927.4	839.2	3.0	6.88	{7.9060281149653}	7.9060281149653555
133.1	210.2	0.0	195.7	3.1	949.4	795.3	28.0	28.94	{29.909670454107}	29.90967045410782
136.0	162.0	126.0	172.0	10.0	923.0	764.0	28.0	29.07	{30.297083753127}	30.297083753127474
136.0	196.0	98.0	199.0	6.0	847.0	783.0	28.0	26.97	{33.626668830798}	33.626668830798884
139.6	209.4	0.0	192.0	0.0	1047.0	806.9	7.0	14.59	{11.280078862592}	11.280078862592134
139.6	209.4	0.0	192.0	0.0	1047.0	806.9	28.0	28.24	{26.423179241134}	26.423179241134303
139.6	209.4	0.0	192.0	0.0	1047.0	806.9	90.0	39.36	{38.775039769988}	38.77503976998816
139.7	163.9	127.7	236.7	5.8	868.6	655.6	28.0	35.23	{33.813219065518}	33.81321906551879
140.0	133.0	103.0	200.0	7.0	916.0	753.0	28.0	36.44	{33.075507366166}	33.07550736616611
141.9	166.6	129.7	173.5	10.9	882.6	785.3	28.0	44.61	{34.034988038101}	34.03498803810175
142.0	167.0	130.0	174.0	11.0	883.0	785.0	28.0	44.61	{34.091406666564}	34.09140666656469
144.0	136.0	106.0	178.0	7.0	941.0	774.0	28.0	26.14	{27.963026614734}	27.963026614734147
145.0	0.0	134.0	181.0	11.0	979.0	812.0	28.0	13.2	{14.164433423972}	14.164433423972408
145.0	116.0	119.0	184.0	5.7	833.0	880.0	28.0	29.16	{27.405954825524}	27.405954825524
145.4	0.0	178.9	201.7	7.8	824.0	868.7	28.0	10.54	{10.957437233230}	10.957437233230017

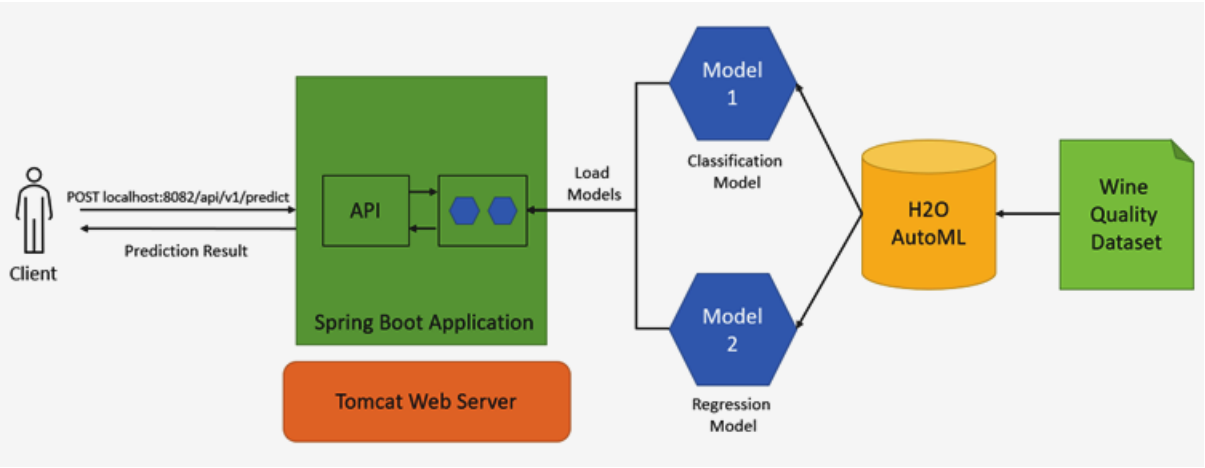
only showing top 20 rows

Cement	Blast Furnace Slag	Fly Ash	Water	Superplasticizer	Coarse Aggregate	Fine Aggregate	Age	Concrete compressive strength	detailed_prediction	prediction
102.0	153.0	0.0	192.0	0.0	887.0	942.0	28.0	17.28	{19.20426780893517}	19.20426780893517
108.3	162.4	0.0	203.5	0.0	938.2	849.0	28.0	20.59	{21.78240293412489}	21.78240293412489
116.0	173.0	0.0	192.0	0.0	909.8	891.9	7.0	10.09	{9.418119576417453}	9.418119576417453
122.6	183.9	0.0	203.5	0.0	958.2	800.1	3.0	3.32	{4.788849427699202}	4.788849427699202
122.6	183.9	0.0	203.5	0.0	958.2	800.1	7.0	10.35	{10.959896897256051}	10.959896897256051
133.0	200.0	0.0	192.0	0.0	927.4	839.2	3.0	6.88	{8.052752331426289}	8.052752331426289
133.1	210.2	0.0	195.7	3.1	949.4	795.3	28.0	28.94	{30.209503295883003}	30.209503295883003
136.0	162.0	126.0	172.0	10.0	923.0	764.0	28.0	29.07	{31.24308884928773}	31.24308884928773
136.0	196.0	98.0	199.0	6.0	847.0	783.0	28.0	26.97	{34.16633337130581}	34.16633337130581
139.6	209.4	0.0	192.0	0.0	1047.0	806.9	7.0	14.59	{12.15639740930616}	12.15639740930616
139.6	209.4	0.0	192.0	0.0	1047.0	806.9	28.0	28.24	{27.173735465958444}	27.173735465958444
139.6	209.4	0.0	192.0	0.0	1047.0	806.9	90.0	39.36	{39.92498779377876}	39.92498779377876
139.7	163.9	127.7	236.7	5.8	868.6	655.6	28.0	35.23	{33.70633527711549}	33.70633527711549
140.0	133.0	103.0	200.0	7.0	916.0	753.0	28.0	36.44	{32.804805244797436}	32.804805244797436
141.9	166.6	129.7	173.5	10.9	882.6	785.3	28.0	44.61	{32.74609436008333}	32.74609436008333
142.0	167.0	130.0	174.0	11.0	883.0	785.0	28.0	44.61	{32.884240085206514}	32.884240085206514
144.0	136.0	106.0	178.0	7.0	941.0	774.0	28.0	26.14	{27.641907236852298}	27.641907236852298
145.0	0.0	134.0	181.0	11.0	979.0	812.0	28.0	13.2	{13.215764253351152}	13.215764253351152
145.0	116.0	119.0	184.0	5.7	833.0	880.0	28.0	29.16	{26.058585272574064}	26.058585272574064
145.4	0.0	178.9	201.7	7.8	824.0	868.7	28.0	10.54	{10.710317214544457}	10.710317214544457

only showing top 20 rows

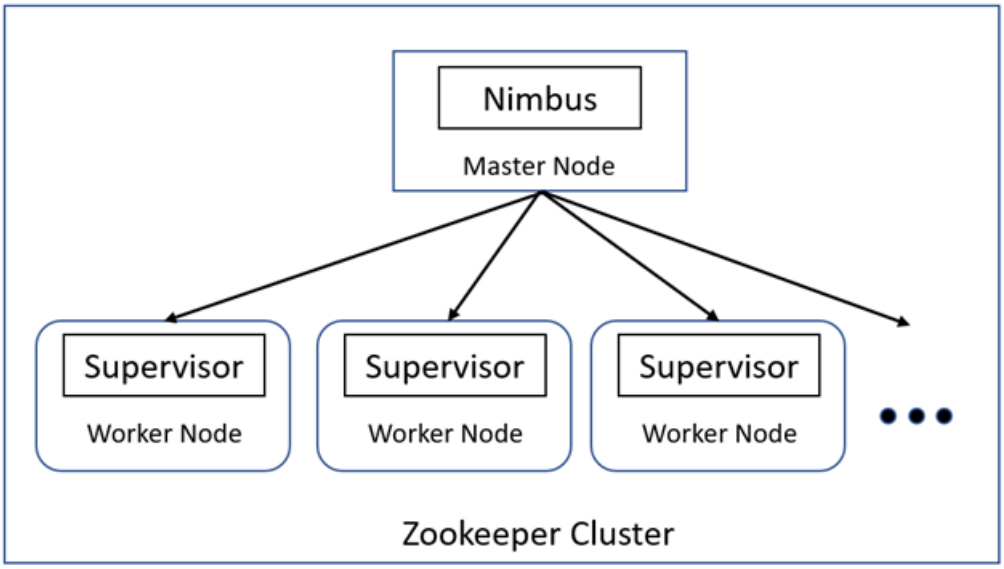
Chapter 13: Using H2O AutoML with Other Technologies

fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality	color
6.8	0.18	0.37	1.6	0.055	47	154	0.9934	3.08	0.45	9.1	5	white
7.2	0.27	0.74	12.5	0.037	47	156	0.9981	3.04	0.44	8.7	5	white
7	0.39	0.31	5.3	0.169	32	162	0.9965	3.2	0.48	9.4	5	white
9.2	0.25	0.34	1.2	0.026	31	93	0.9916	2.93	0.37	11.3	7	white
7.4	0.35	0.24	6	0.042	28	123	0.99304	3.14	0.44	11.3	5	white
6.5	0.3	0.39	7.8	0.038	61	219	0.9959	3.19	0.5	9.4	5	white
7.8	0.76	0.04	2.3	0.092	15	54	0.997	3.26	0.65	9.8	5	red
...	...	...	...	...	...	...	...	...	...	...	...	...

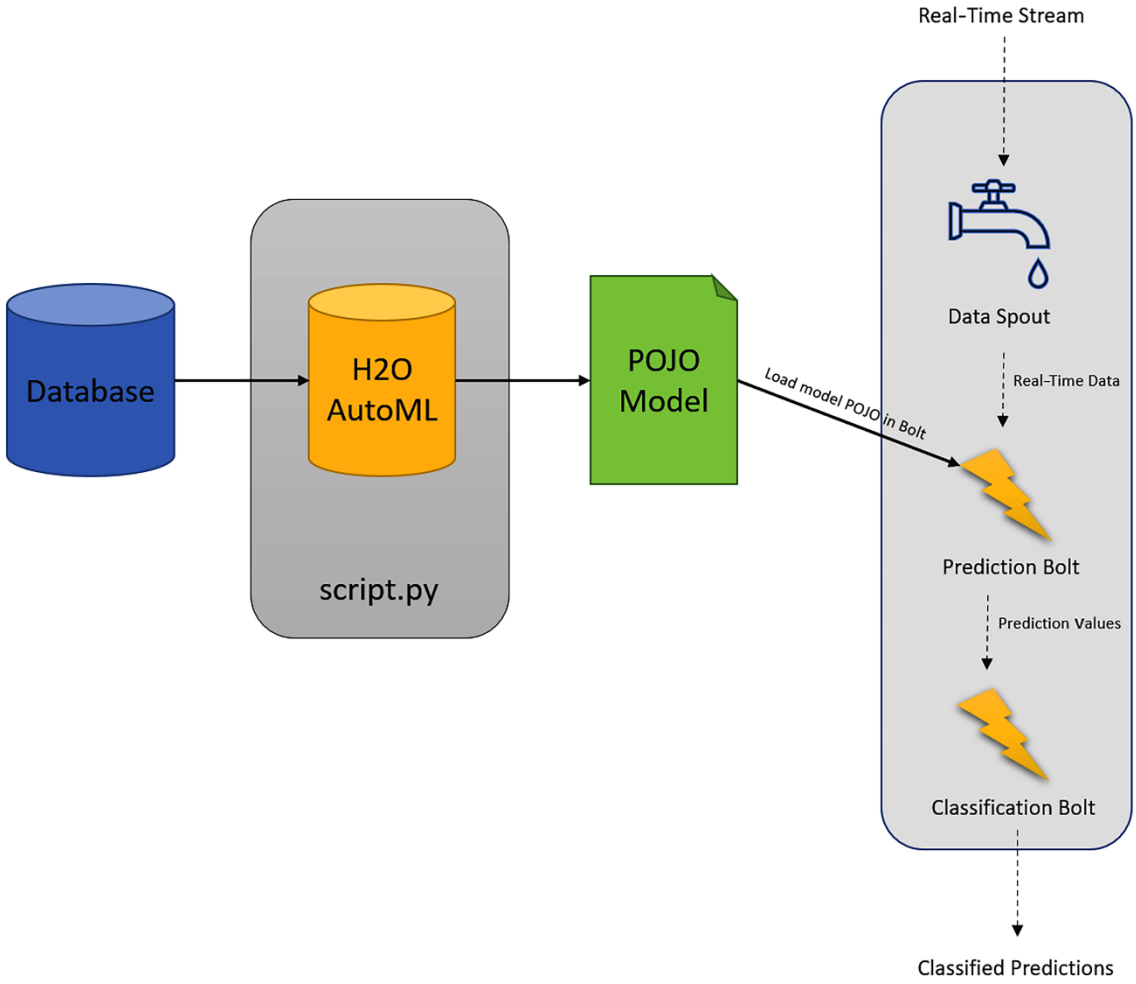


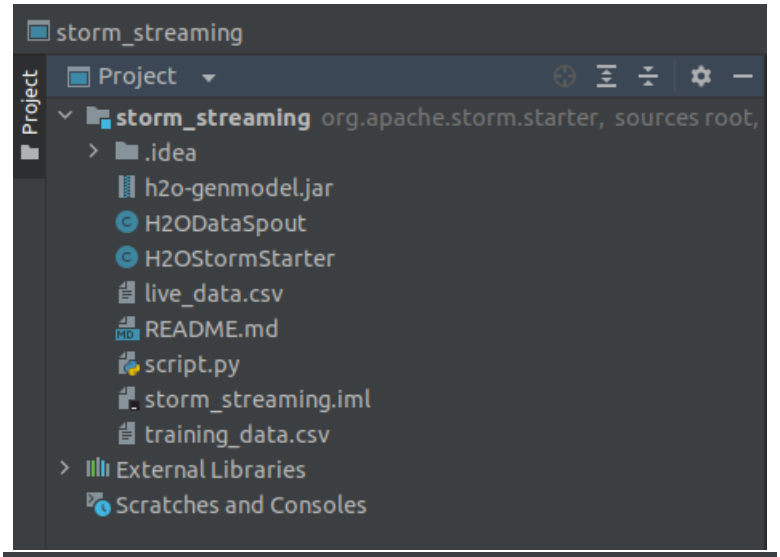
```
salil@salil-VirtualBox:~/Desktop/Chapter_13/Final/h2o_wine_predictor/demo$ mvn clean spring-boot:run -e
[INFO] Error stacktraces are turned on.
[INFO] Scanning for projects...
[INFO]
[INFO] --------------------< com.h2o_wine_predictor:demo >---------------------
[INFO] Building demo 0.0.1-SNAPSHOT
[INFO] --------------------------------[ jar ]---------------------------------
[INFO]
[INFO] --- maven-clean-plugin:3.2.0:clean (default-clean) @ demo ---
[INFO] Deleting /home/salil/Desktop/Chapter_13/Final/h2o_wine_predictor/demo/target
[INFO]
[INFO] >>> spring-boot-maven-plugin:2.7.2:run (default-cli) > test-compile @ demo >>>
[INFO]
[INFO] --- maven-resources-plugin:3.2.0:resources (default-resources) @ demo ---
[INFO] Using 'UTF-8' encoding to copy filtered resources.
[INFO] Using 'UTF-8' encoding to copy filtered properties files.
[INFO] Copying 1 resource
[INFO] Copying 2 resources
[INFO]
[INFO] --- maven-compiler-plugin:3.10.1:compile (default-compile) @ demo ---
[INFO] Changes detected - recompiling the module!
[INFO] Compiling 5 source files to /home/salil/Desktop/Chapter_13/Final/h2o_wine_predictor/demo/target/classes
[INFO]
[INFO] --- maven-resources-plugin:3.2.0:testResources (default-testResources) @ demo ---
[INFO] Using 'UTF-8' encoding to copy filtered resources.
[INFO] Using 'UTF-8' encoding to copy filtered properties files.
[INFO] skip non existing resourceDirectory /home/salil/Desktop/Chapter_13/Final/h2o_wine_predictor/demo/src/test/resources
[INFO]
[INFO] --- maven-compiler-plugin:3.10.1:testCompile (default-testCompile) @ demo ---
[INFO] Changes detected - recompiling the module!
[INFO] Compiling 1 source file to /home/salil/Desktop/Chapter_13/Final/h2o_wine_predictor/demo/target/test-classes
[INFO]
[INFO] <<< spring-boot-maven-plugin:2.7.2:run (default-cli) < test-compile @ demo <<<
[INFO]
[INFO]
[INFO] --- spring-boot-maven-plugin:2.7.2:run (default-cli) @ demo ---
[INFO] Attaching agents: []

  .   ____          _            __ _ _
 /\\ / ___'_ __ _ _(_)_ __  __ _ \ \ \ \
( ( )\___ | '_ | '_| | '_ \/ _` | \ \ \ \
 \\/  ___)| |_)| | | | | || (_| |  ) ) ) )
  '  |____| .__|_| |_|_| |_\__, | / / / /
 =========|_|==============|___/=/_/_/_/
 :: Spring Boot ::                (v2.7.2)

2022-07-24 18:29:03.261  INFO 4701 --- [           main] c.h.demo.DemoApplication                 : Starting DemoApplication using Java 11.0.15 on salil-VirtualBox with PID 4701 (/home/salil/Desktop/Chapter_13/Final/h2o_wine_predictor/demo/target/classes started by salil in /home/salil/Desktop/Chapter_13/Final/h2o_wine_predictor/demo)
2022-07-24 18:29:03.266  INFO 4701 --- [           main] c.h.demo.DemoApplication                 : No active profile set, falling back to 1 default profile: "default"
2022-07-24 18:29:04.422  INFO 4701 --- [           main] o.s.b.w.embedded.tomcat.TomcatWebServer  : Tomcat initialized with port(s): 8082 (http)
2022-07-24 18:29:04.428  INFO 4701 --- [           main] o.apache.catalina.core.StandardService   : Starting service [Tomcat]
2022-07-24 18:29:04.429  INFO 4701 --- [           main] org.apache.catalina.core.StandardEngine  : Starting Servlet engine: [Apache Tomcat/9.0.65]
2022-07-24 18:29:04.592  INFO 4701 --- [           main] o.a.c.c.C.[Tomcat].[localhost].[/]       : Initializing Spring embedded WebApplicationContext
2022-07-24 18:29:04.592  INFO 4701 --- [           main] w.s.c.ServletWebServerApplicationContext : Root WebApplicationContext: initialization completed in 1221 ms
2022-07-24 18:29:05.126  INFO 4701 --- [           main] o.s.b.w.embedded.tomcat.TomcatWebServer  : Tomcat started on port(s): 8082 (http) with context path ''
2022-07-24 18:29:05.156  INFO 4701 --- [           main] c.h.demo.DemoApplication                 : Started DemoApplication in 2.452 seconds (JVM running for 2.943)
```

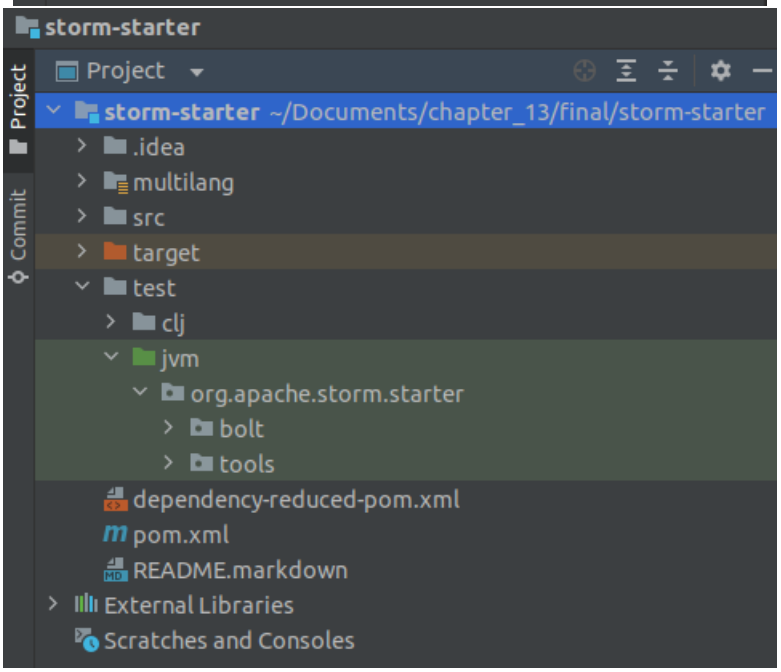

```
{
  "labelIndex" : 1,
  "label" : "white",
  "classProbabilities" : [
    0.0,
    1.0
  ],

  "quality" : 5.31746032124474
}
```

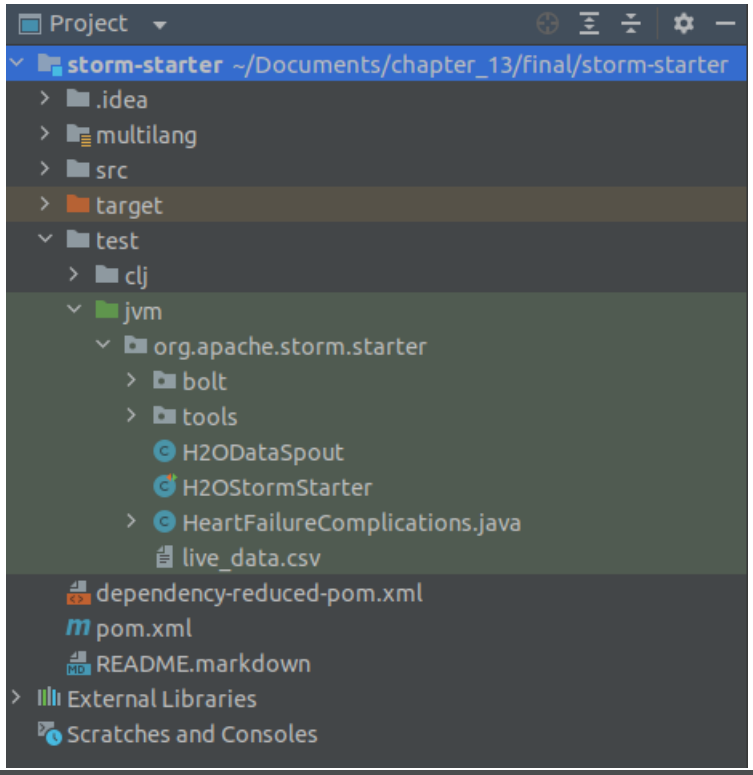



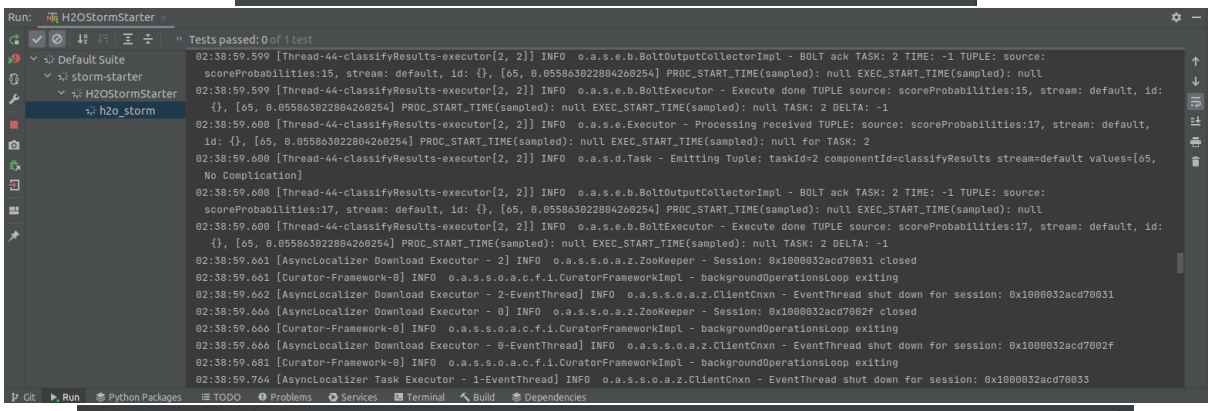
age	anemia	creatinine_ phosphokinase	high_ blood_ pressure	diabetes	ejection_ fraction	platelets	sex	serum_ creatinine	serum_ sodium	smoking	time	complications
75	0	582	1	0	20	265000	1	1.9	130	0	4	1
55	0	7861	0	0	38	263358	1	1.1	136	0	6	1
65	0	146	0	0	20	162000	1	1.3	129	1	7	1
50	1	111	0	0	20	210000	1	1.9	137	0	7	1
65	1	160	0	1	20	327000	0	2.7	116	0	8	1
90	1	47	1	0	40	204000	1	2.1	132	1	8	1
...	...	...	...	...	...	...	...	...	...	...	...	...

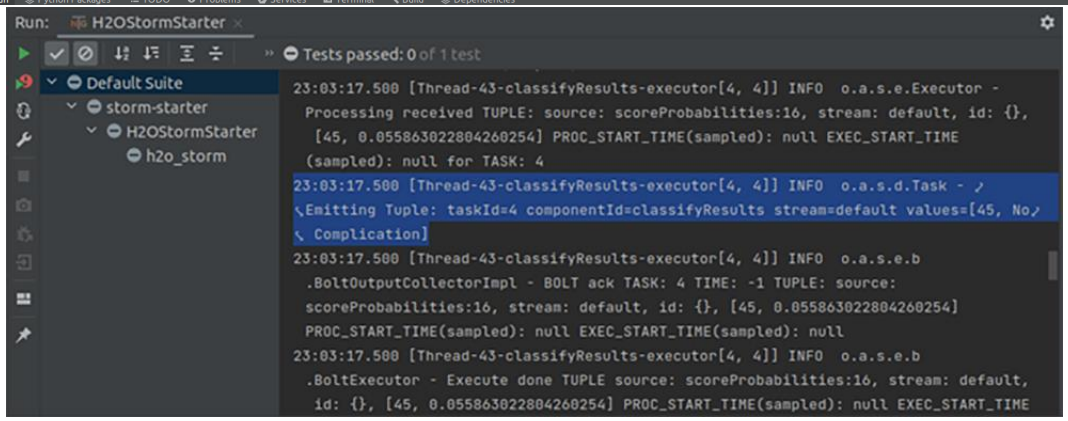



storm_streaming
Project
storm_streaming org.apache.storm.starter, sources root,
.idea
h2o-genmodel.jar
H2ODataSpout
H2OStormStarter
live_data.csv
README.md
script.py
storm_streaming.iml
training_data.csv
External Libraries
Scratches and Consoles


storm-starter
Project
Commit
storm-starter ~/Documents/chapter_13/final/storm-starter
.idea
multilang
src
target
test
clj
jvm
org.apache.storm.starter
bolt
tools
dependency-reduced-pom.xml
pom.xml
README.markdown
External Libraries
Scratches and Consoles


Project
storm-starter ~/Documents/chapter_13/final/storm-starter
.idea
multilang
src
target
test
clj
jvm
org.apache.storm.starter
bolt
tools
H2ODataSpout
H2OStormStarter
HeartFailureComplications.java
live_data.csv
dependency-reduced-pom.xml
pom.xml
README.markdown
External Libraries
Scratches and Consoles


Run: H2OStormStarter
Tests passed: 0 of 1 test
Default Suite
storm-starter
H2OStormStarter
h2o_storm
02:38:59.599 [Thread-44-classifyResults-executor[2, 2]] INFO o.a.s.e.b.BoltOutputCollectorImpl - BOLT ack TASK: 2 TIME: -1 TUPLE: source: scoreProbabilities:15, stream: default, id: {}, [65, 0.05586302280426025 4] PROC_START_TIME(sampled): null EXEC_START_TIME(sampled): null
02:38:59.599 [Thread-44-classifyResults-executor[2, 2]] INFO o.a.s.e.b.BoltExecutor - Execute done TUPLE source: scoreProbabilities:15, stream: default, id: {}, [65, 0.05586302280426025 4] PROC_START_TIME(sampled): null EXEC_START_TIME(sampled): null TASK: 2 DELTA: -1
02:38:59.600 [Thread-44-classifyResults-executor[2, 2]] INFO o.a.s.e.Executor - Processing received TUPLE: source: scoreProbabilities:17, stream: default, id: {}, [65, 0.05586302280426025 4] PROC_START_TIME(sampled): null EXEC_START_TIME(sampled): null for TASK: 2
02:38:59.600 [Thread-44-classifyResults-executor[2, 2]] INFO o.a.s.d.Task - Emitting Tuple: taskId=2 componentId=classifyResults stream=default values=[65, No Complication]
02:38:59.600 [Thread-44-classifyResults-executor[2, 2]] INFO o.a.s.e.b.BoltOutputCollectorImpl - BOLT ack TASK: 2 TIME: -1 TUPLE: source: scoreProbabilities:17, stream: default, id: {}, [65, 0.05586302280426025 4] PROC_START_TIME(sampled): null EXEC_START_TIME(sampled): null
02:38:59.600 [Thread-44-classifyResults-executor[2, 2]] INFO o.a.s.e.b.BoltExecutor - Execute done TUPLE source: scoreProbabilities:17, stream: default, id: {}, [65, 0.05586302280426025 4] PROC_START_TIME(sampled): null EXEC_START_TIME(sampled): null TASK: 2 DELTA: -1
02:38:59.661 [AsyncLocalizer Download Executor - 2] INFO o.a.s.s.o.a.z.ZooKeeper - Session: 0x1000032acd70031 closed
02:38:59.661 [Curator-Framework-0] INFO o.a.s.s.o.a.c.f.i.CuratorFrameworkImpl - backgroundOperationsLoop exiting
02:38:59.662 [AsyncLocalizer Download Executor - 2-EventThread] INFO o.a.s.s.o.a.z.ClientCnxn - EventThread shut down for session: 0x1000032acd70031
02:38:59.666 [AsyncLocalizer Download Executor - 0] INFO o.a.s.s.o.a.z.ZooKeeper - Session: 0x1000032acd7002f closed
02:38:59.666 [Curator-Framework-0] INFO o.a.s.s.o.a.c.f.i.CuratorFrameworkImpl - backgroundOperationsLoop exiting
02:38:59.666 [AsyncLocalizer Download Executor - 0-EventThread] INFO o.a.s.s.o.a.z.ClientCnxn - EventThread shut down for session: 0x1000032acd7002f
02:38:59.681 [Curator-Framework-0] INFO o.a.s.s.o.a.c.f.i.CuratorFrameworkImpl - backgroundOperationsLoop exiting
02:38:59.764 [AsyncLocalizer Task Executor - 1-EventThread] INFO o.a.s.s.o.a.z.ClientCnxn - EventThread shut down for session: 0x1000032acd70033
Git
Run
Python Packages
TODO
Problems
Services
Terminal
Build
Dependencies


Run: H2OStormStarter
Tests passed: 0 of 1 test
Default Suite
storm-starter
H2OStormStarter
h2o_storm
23:03:17.500 [Thread-43-classifyResults-executor[4, 4]] INFO o.a.s.e.Executor - Processing received TUPLE: source: scoreProbabilities:16, stream: default, id: {}, [45, 0.055863022804260254] PROC_START_TIME(sampled): null EXEC_START_TIME (sampled): null for TASK: 4
23:03:17.500 [Thread-43-classifyResults-executor[4, 4]] INFO o.a.s.d.Task - Emitting Tuple: taskId=4 componentId=classifyResults stream=default values=[45, No Complication]
23:03:17.500 [Thread-43-classifyResults-executor[4, 4]] INFO o.a.s.e.b.BoltOutputCollectorImpl - BOLT ack TASK: 4 TIME: -1 TUPLE: source: scoreProbabilities:16, stream: default, id: {}, [45, 0.055863022804260254] PROC_START_TIME(sampled): null EXEC_START_TIME(sampled): null
23:03:17.500 [Thread-43-classifyResults-executor[4, 4]] INFO o.a.s.e.b.BoltExecutor - Execute done TUPLE source: scoreProbabilities:16, stream: default, id: {}, [45, 0.055863022804260254] PROC_START_TIME(sampled): null EXEC_START_TIME