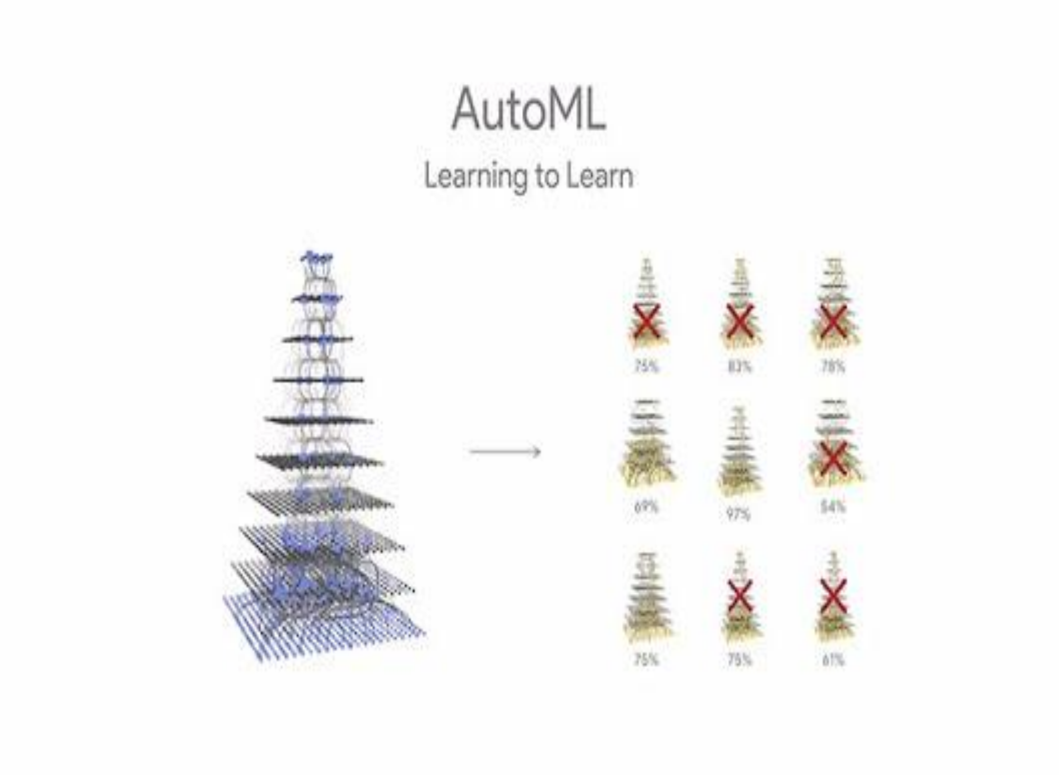
Group 6

Yinyin, Xinglong, Himanshu



AutoML Introduction

1. Machine Learning
2. What is AutoML
3. Types of Frameworks
4. Frameworks

Machine Learning

1. Preprocess and clean the data
2. Feature engineering
    a. Select and construct appropriate features
3. Model building
    a. Select an appropriate model family
4. Hyperparameter optimization
    a. Optimize model hyperparameters
5. Postprocess machine learning models
6. Analyze the results obtained

Reference: http://www.ml4aad.org/automl/

Machine Learning

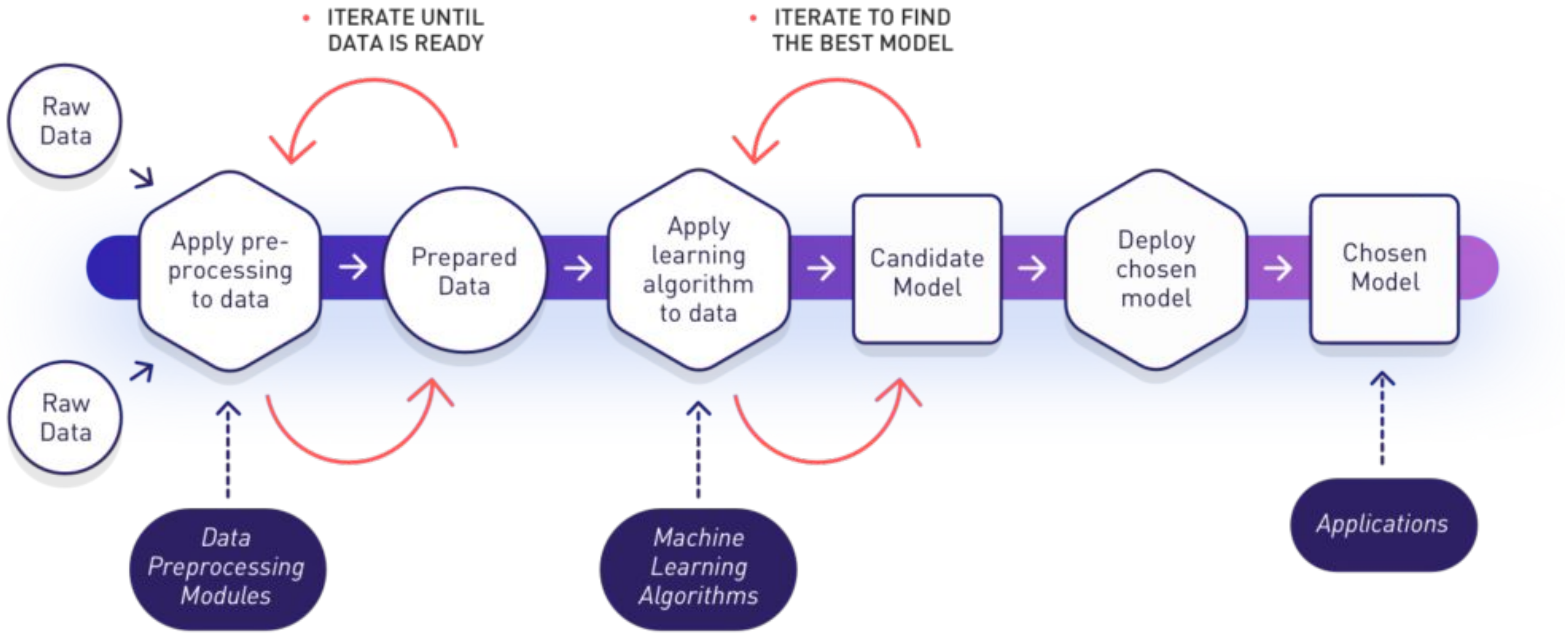


Image from: https://www.datasciencecentral.com/profiles/blogs/soccer-and-machine-learning-2-hot-topics-for-2018

What is AutoML

Automated Machine Learning provides methods and processes to make Machine Learning available for non-Machine Learning experts, to improve efficiency of Machine Learning and to accelerate research on Machine Learning.

Automation of machine learning

Reference: http://www.ml4aad.org/automl/

Types of frameworks

- Automated feature engineering
    - feature selection
    - feature extraction
    - meta learning and transfer learning
    - Detection and handling of skewed data/missing values
- Hyperparameter optimization
- Model Selection

Reference: https://en.wikipedia.org/wiki/Automated_machine_learning

Well-known frameworks

- Full pipeline automation
  - Auto-WEKA
  - Auto-sklearn
- Hyperparameter optimization and Model Selection
  - H2O AutoML
- Deep Neural Network Architecture search
  - Google Cloud AutoML

Reference: http://www.ml4aad.org/automl/

Auto-WEKA

Auto-WEKA is a tool that performs combined algorithm selection and hyper-parameter optimisation over the classification and regression algorithms implements in WEKA

Auto-WEKA explores hyperparameter settings for many algorithms and recommends to a user which method will likely have good generalisation performance, using model based optimisation techniques.

Reference: http://www.cs.ubc.ca/labs/beta/Projects/autoweka/manual.pdf

Weka is a collection of machine learning algorithms for data mining tasks. It contains tools for data preparation, classification, regression, clustering, association rules mining, and visualization.

SMAC (sequential model-based algorithm configuration) is a versatile tool for optimizing algorithm parameters (or the parameters of some other process we can run automatically, or a function we can evaluate, such as a simulation).

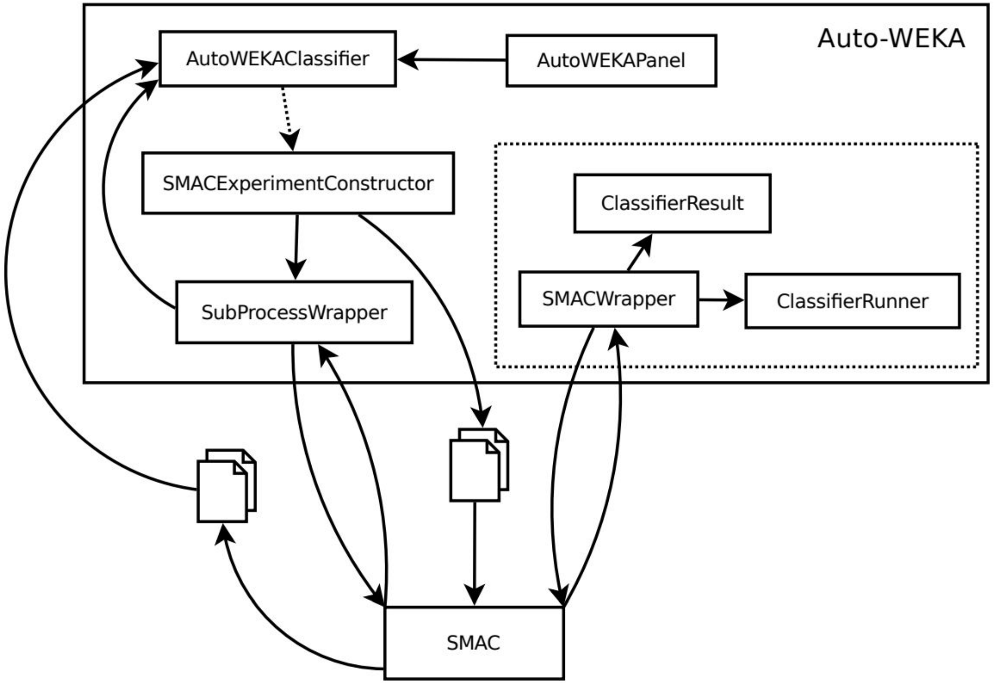
Reference:http://www.cs.ubc.ca/labs/beta/Projects/SMAC/#software

High-level overview of Auto-WEKA internal structure

User interface: AutoWEKAClassifier, AutoWEKAPanlel

SMAC optimization tool:
SMACExperimentConstructor, SubProcessWrapper

Optimization process:
SMACWrapper, ClassifierRunner, ClassifierResult



Reference: http://www.cs.ubc.ca/labs/beta/Projects/autoweka/manual.pdf

Hyper-parameter optimization and Model Selection

- ## H2O AutoML - demo

- Although H2O has made it easy for non-experts to experiment with machine learning, there is still a fair bit of knowledge and background in data science that is required to produce high-performing machine learning models.
- H2O's AutoML can also be a helpful tool for the advanced user, by providing a simple wrapper function that performs a large number of modeling-related tasks that would typically require many lines of code, and by freeing up their time to focus on other aspects of the data science pipeline tasks such as data-preprocessing, feature engineering and model deployment.
- The current version of AutoML trains and cross-validates the following algorithms (in the following order): A default Random Forest (DRF), an Extremely Randomized Forest (XRT), three pre-specified XGBoost GBM (Gradient Boosting Machine) models, five pre-specified H2O GBMs, a near-default Deep Neural Net, a random grid of XGBoost GBMs, a random grid of H2O GBMs, and lastly if there is time, a random grid of Deep Neural Nets.

Deep Neural Network Architecture search

- Google cloud autoML - demo on natural

Cloud AutoML Products

AutoML Natural Language

AutoML Natural Language enables you to train your own, custom machine learning models to classify documents according to labels that you define.

AutoML Translation

AutoML Translation enables you to create your own, custom translation models so that translation queries return results specific to your domain.

AutoML Vision

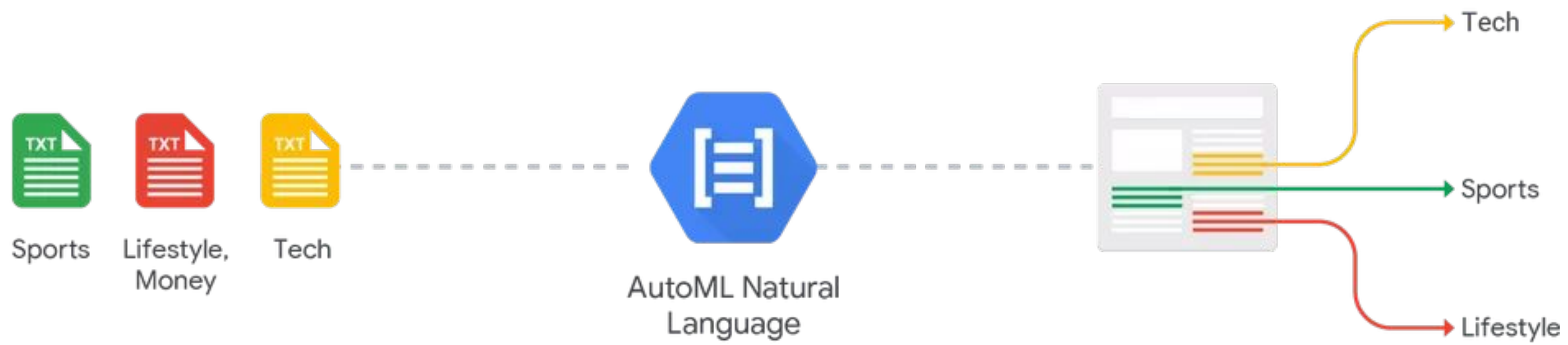
AutoML Vision enables you to train your own, custom machine learning models to classify your images according to labels that you define.

AutoML Natural Language

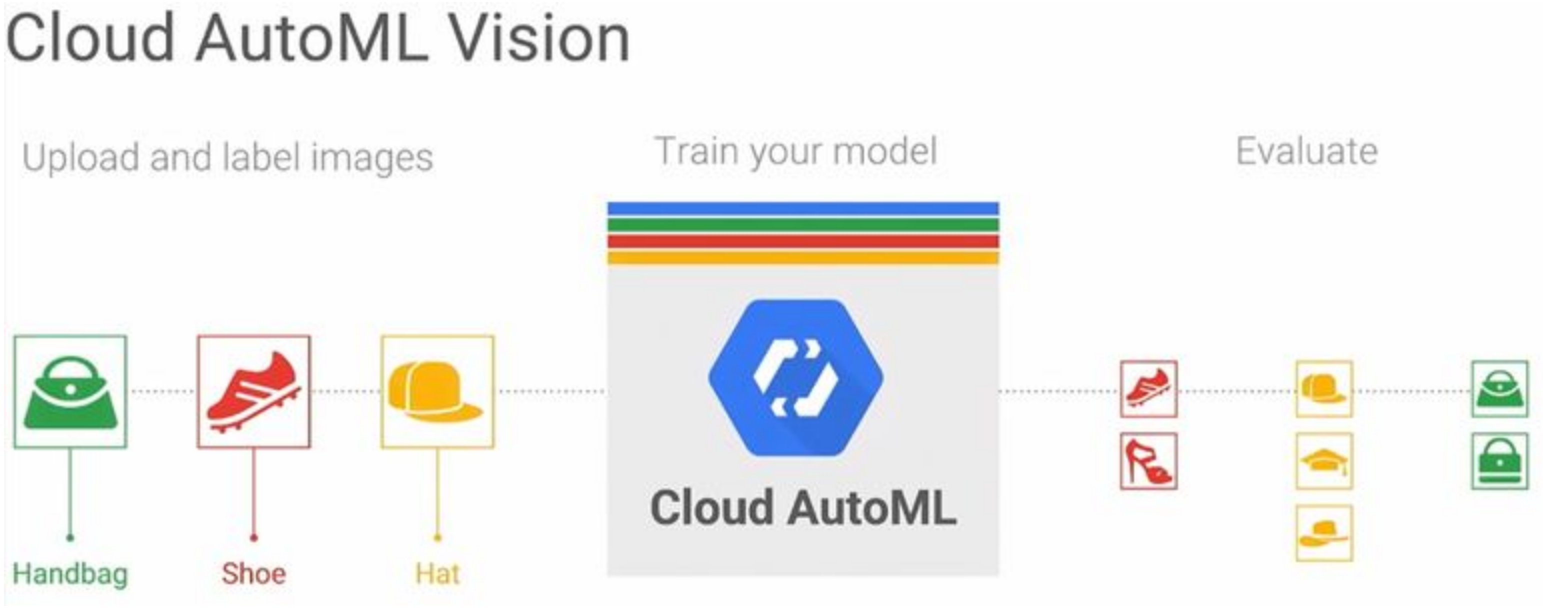
1 Upload and label text

2 Train your model

3 Evaluate

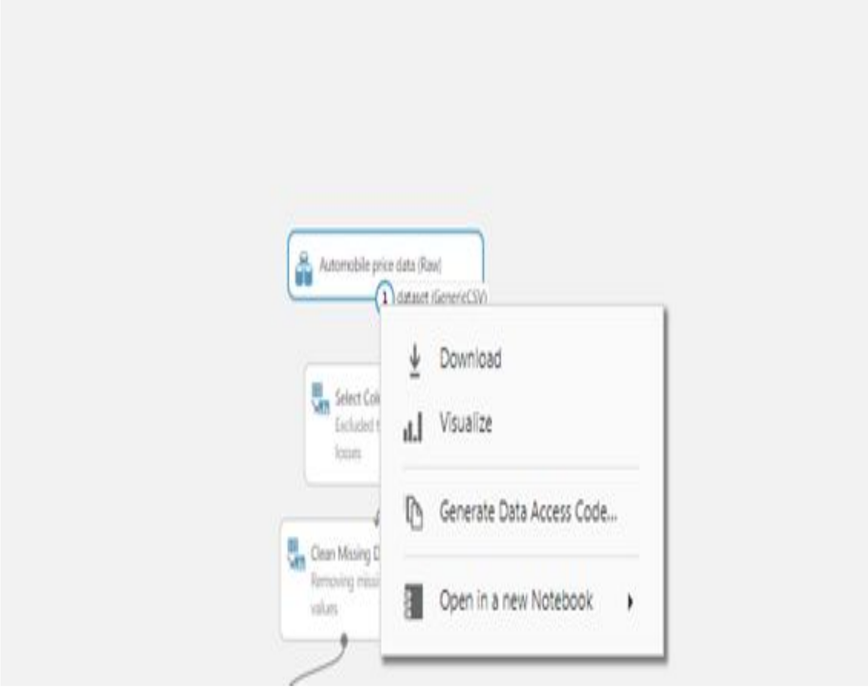


AutoML Vision



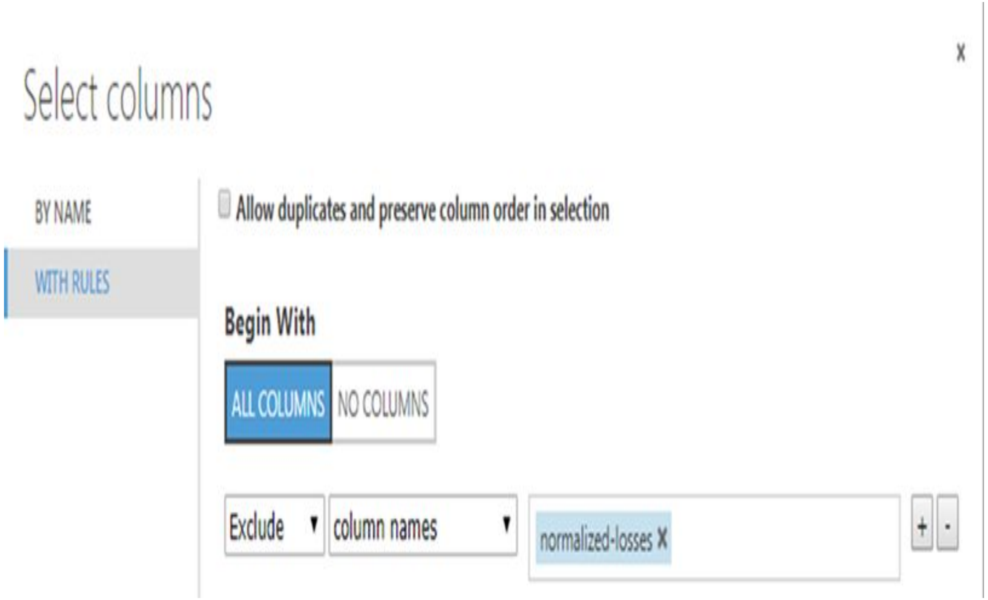
Azure Machine Learning

Step1: Load the Data, see thevisualization at a glance:



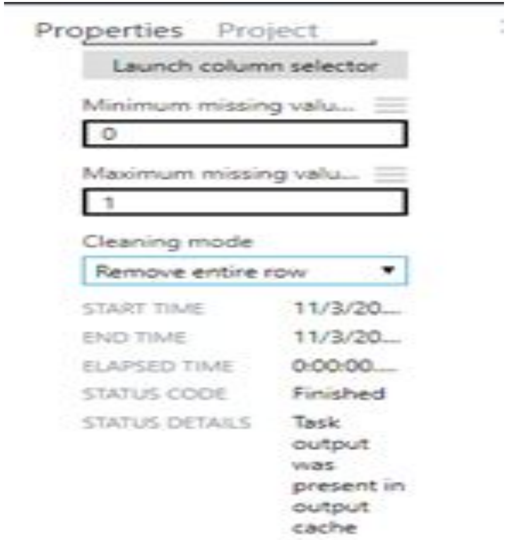
Step2: Prepare the Data

Remove the columns by column names by using Select column module and removing Losses

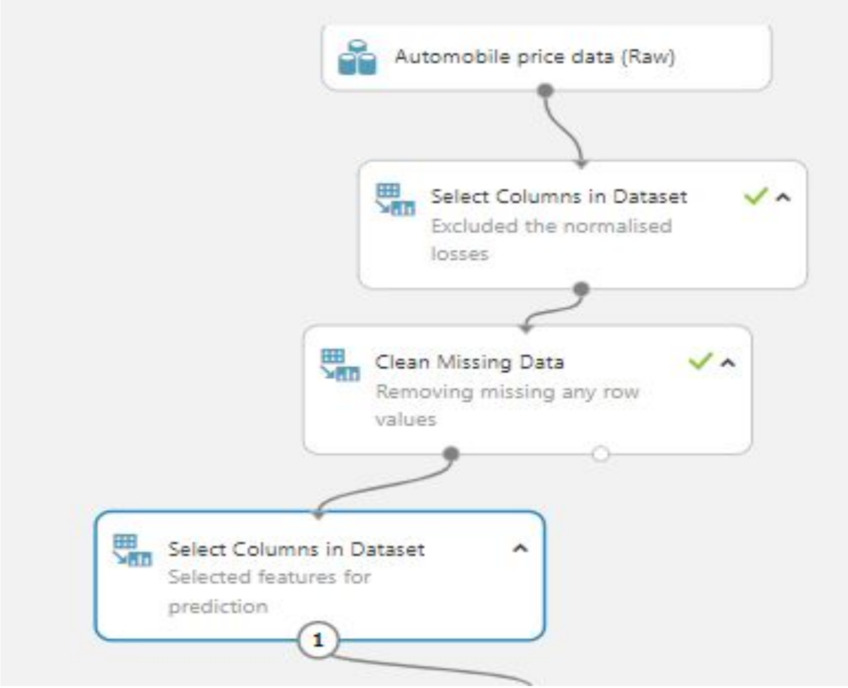


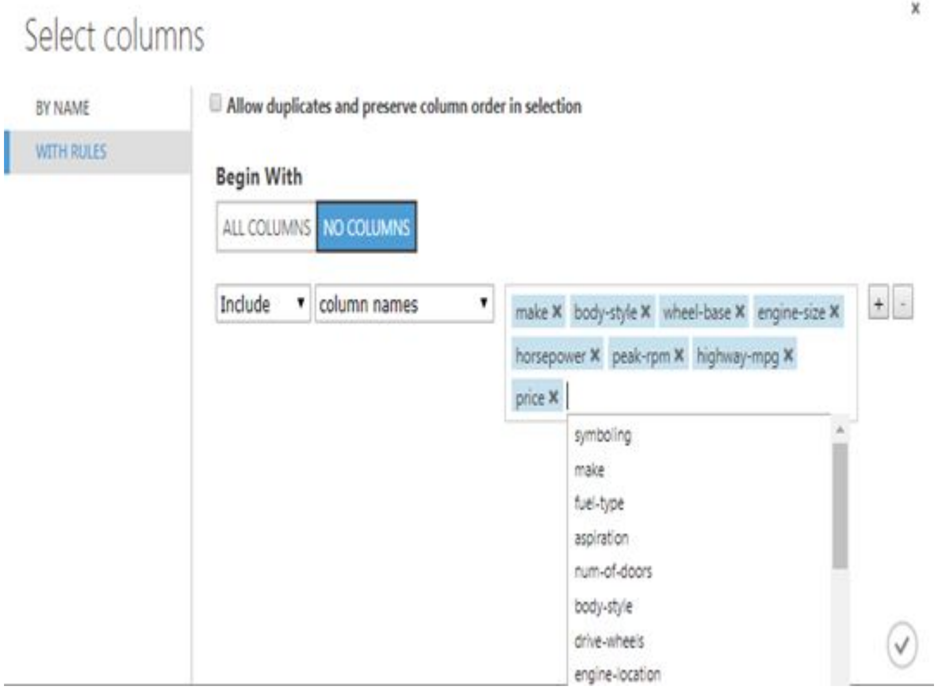
Additional Cleaning

- Remove the rows with missing values by using Clean Missing Dataset module
- Visualize and see the clean dataset from the output port

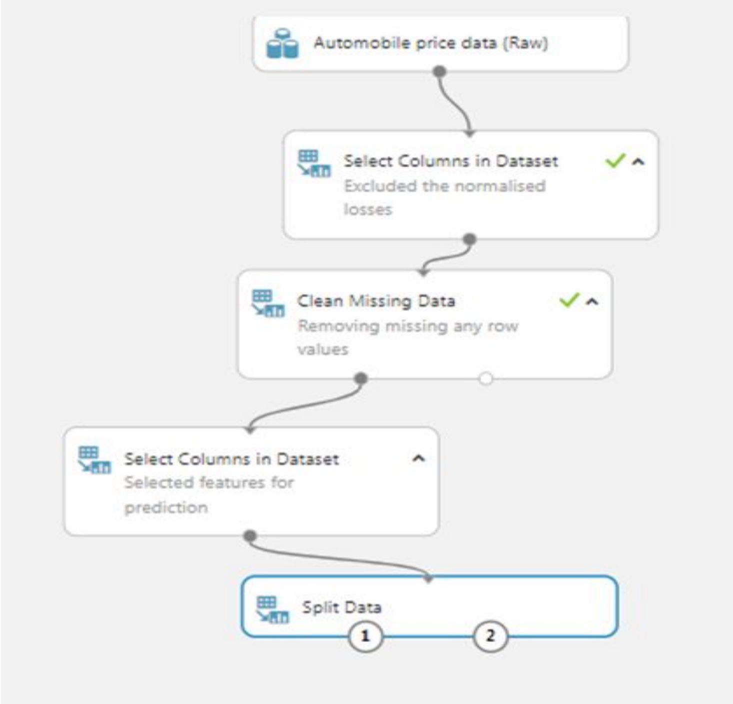


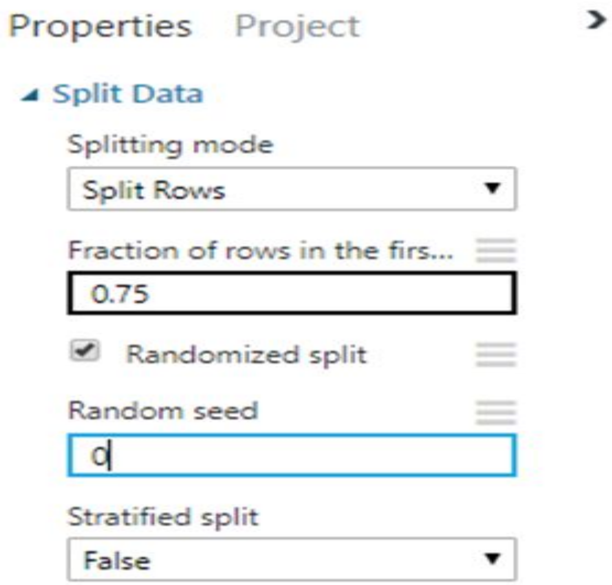
Step 3: Define Features



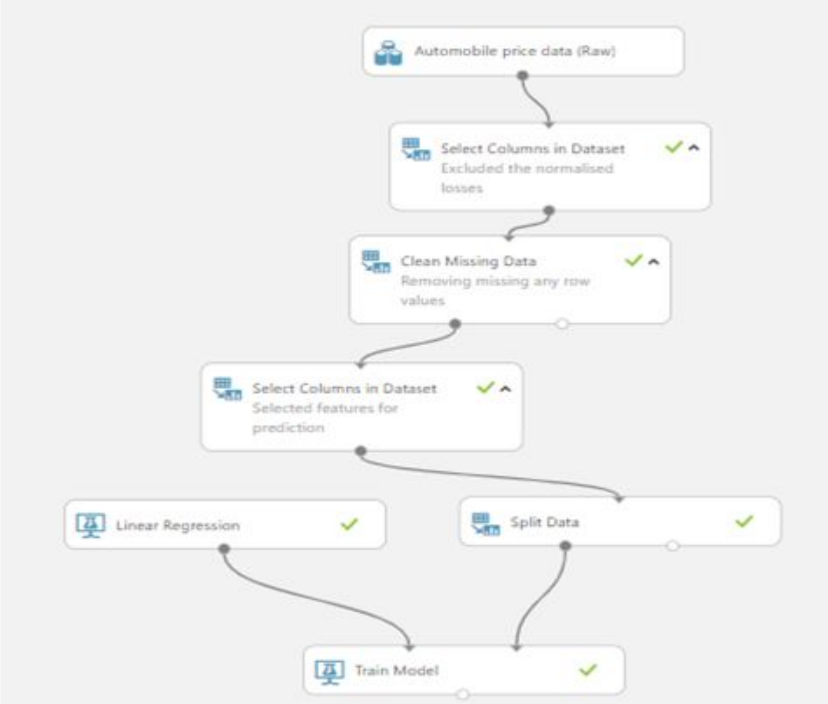


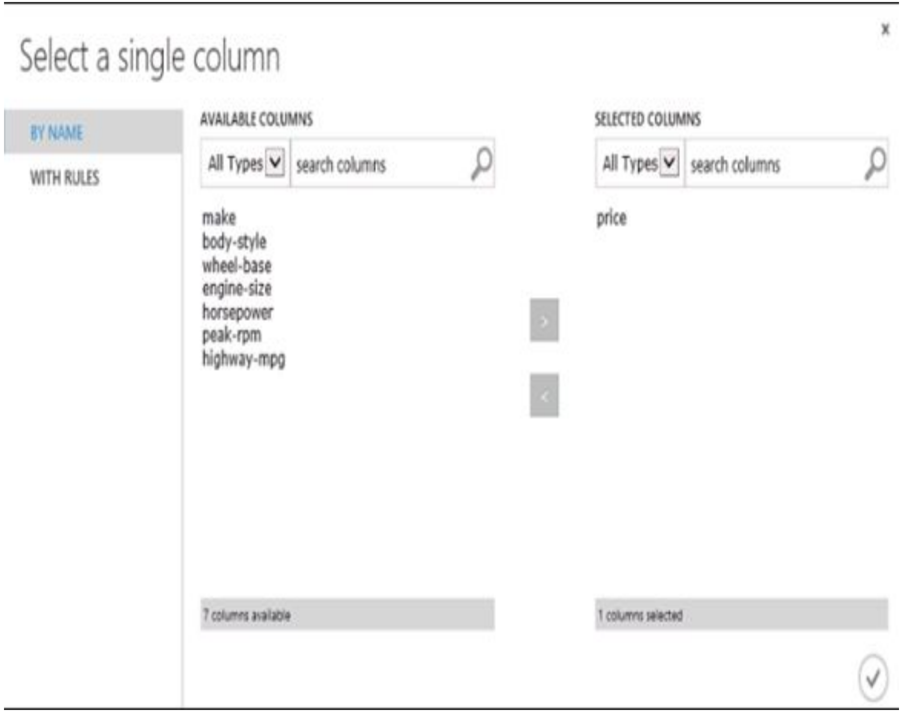
Step 4: Choose and apply a learning algorithm





Add the Linear Regression module and Train model module into the experiment canvas





Step5: Predict new automobile prices

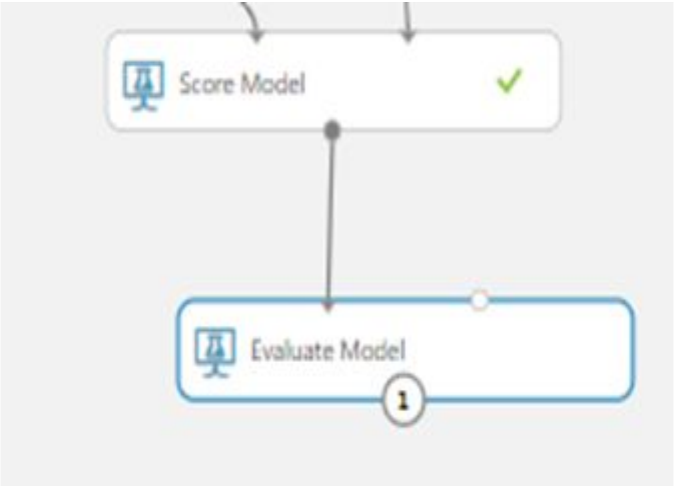
Backorder › Score Model › Scored dataset

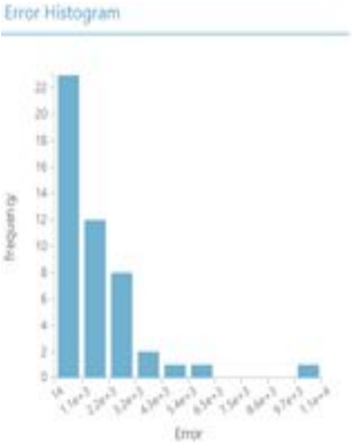
rows 48

columns 9

make	body-style	wheel-base	engine-size	horsepower	peak-rpm	highway-mpg	price	Scored Labels
subaru	sedan	97	108	111	4800	29	11259	10286.204819
mitsubishi	hatchback	93.7	92	68	5500	38	6669	5446.847864
dodge	hatchback	93.7	90	68	5500	38	6229	6344.800711
honda	hatchback	86.6	92	76	6000	38	6855	5528.302953
alfa-romero	convertible	88.6	130	111	5000	27	16500	13498.476233
volvo	wagon	104.3	141	114	5400	28	16515	16097.608038
isuzu	hatchback	96	119	90	5000	29	11048	8315.257218
dodge	hatchback	93.7	90	68	5500	41	5572	6630.154608
bmw	sedan	101.2	108	101	5800	29	16430	19913.408695

Evaluate Models





Metrics

Mean Absolute Error	1656.147651
Root Mean Squared Error	2456.983209
Relative Absolute Error	0.276606
Relative Squared Error	0.089608
Coefficient of Determination	0.910392