



Azure ML Classic Studio

Predicting Automobile prices using Regression Model in Azure ML Classic Studio.

This model (Pipeline) trains a linear regressor to predict a car's price based on technical features such as make, model, horsepower, and size. Because you're trying to answer the question "How much?" this is called a regression problem.

However, you can apply the same fundamental steps in this example to tackle any type of machine learning problem whether it be regression, classification, clustering, and so on.

Gallery Link:

19K41A0555 Automobile Price Prediction

Automobile Price Prediction using pre-available dataset and training the model using Linear Regression. Tags: Linear Regression, Automobile, Azure ML, Microsoft

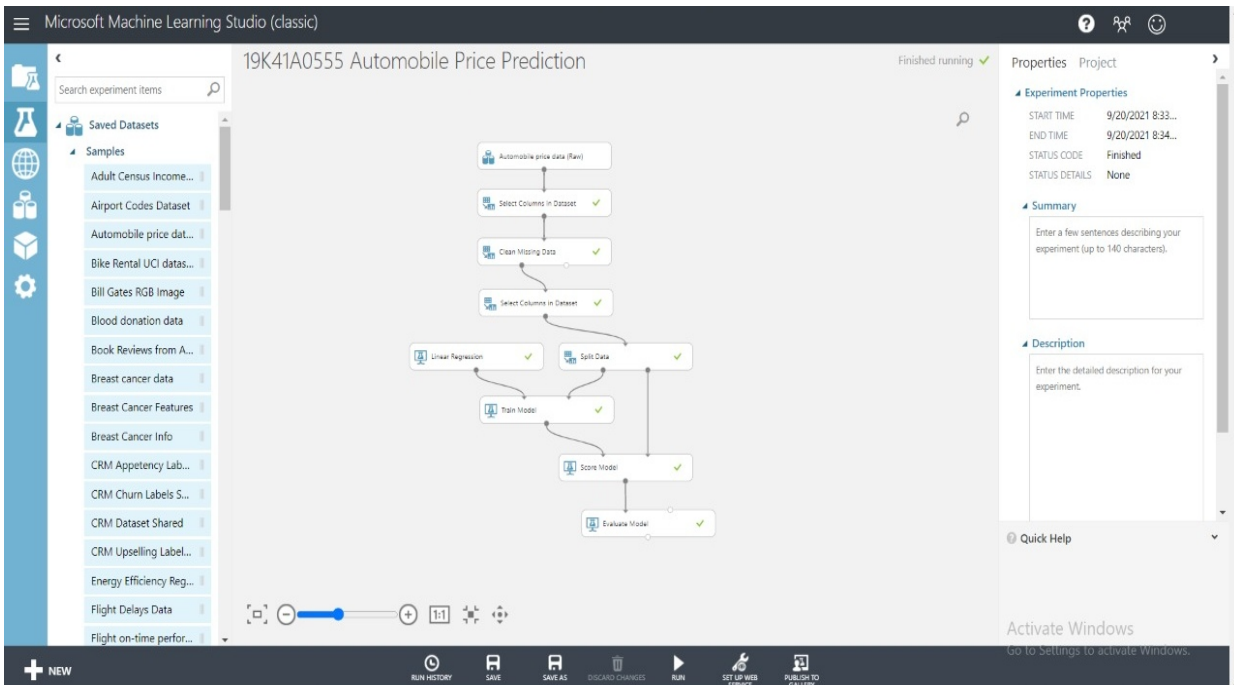
<https://gallery.cortanaintelligence.com/Experiment/19K41A0555-Automobile-Price-Prediction>

navigate to the link to see the Workflow and you can download the project as well.

Machine Learning Project Workflow

- 1.ImportData
- 2.Explore Data (Missing values,outliers)
- 3.Preprocess data (Missing value imputation, outlier treatment, normalization)
- 4.Model Selection
- 5.Model Training
- 6.Model Testing
- 7.ModelDeployment

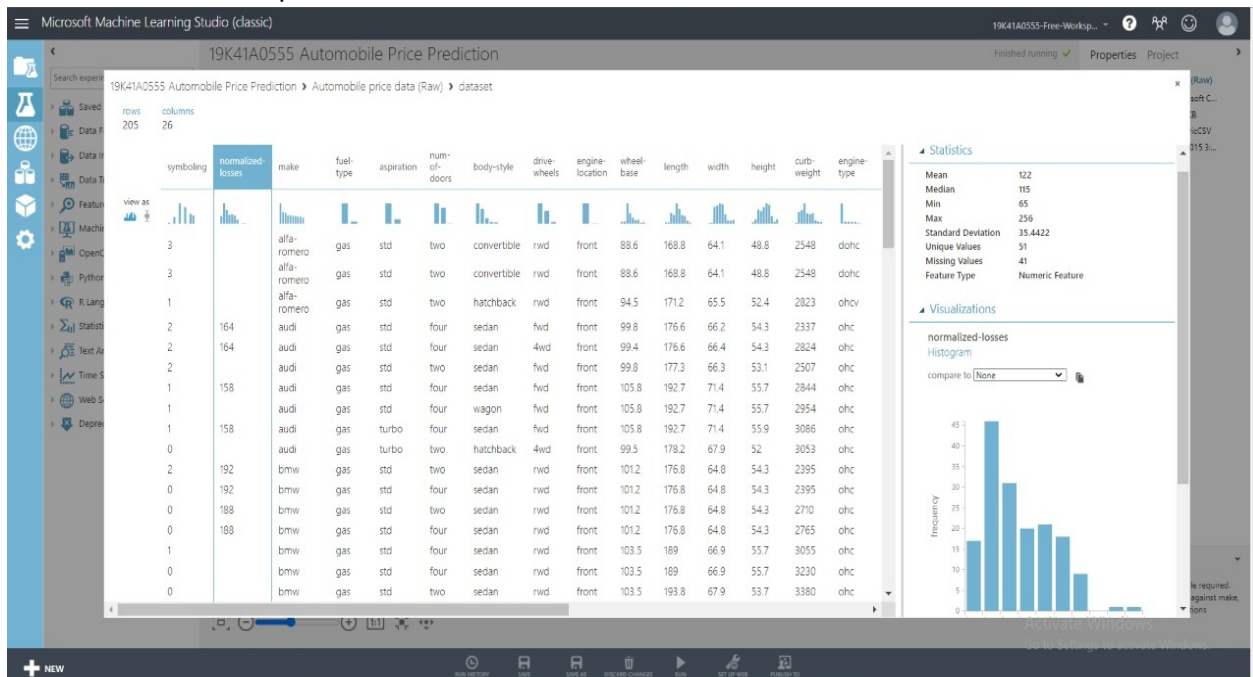
Workflow



Project Workflow

Import Data:

- importing the RAW dataset which is in CSV format.
- the dataset is pre-available in the Azure ML Classic Studio.



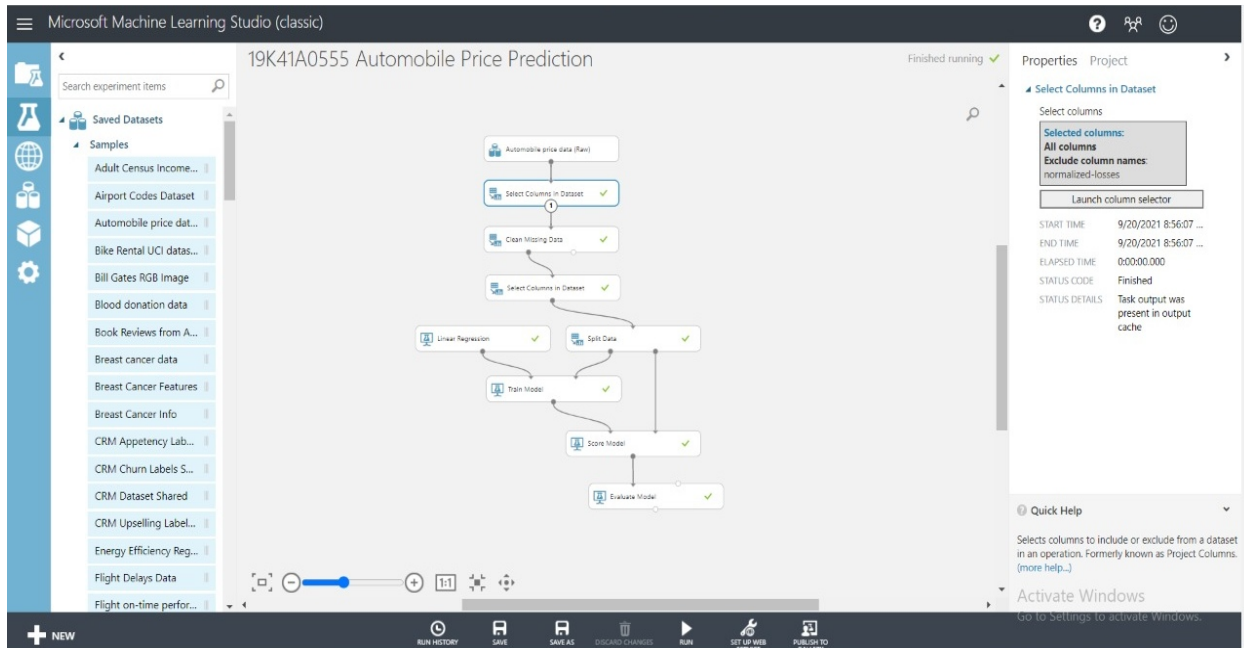
Automobile Price RAW dataset (CSV format)

Azure ML Classic Studio

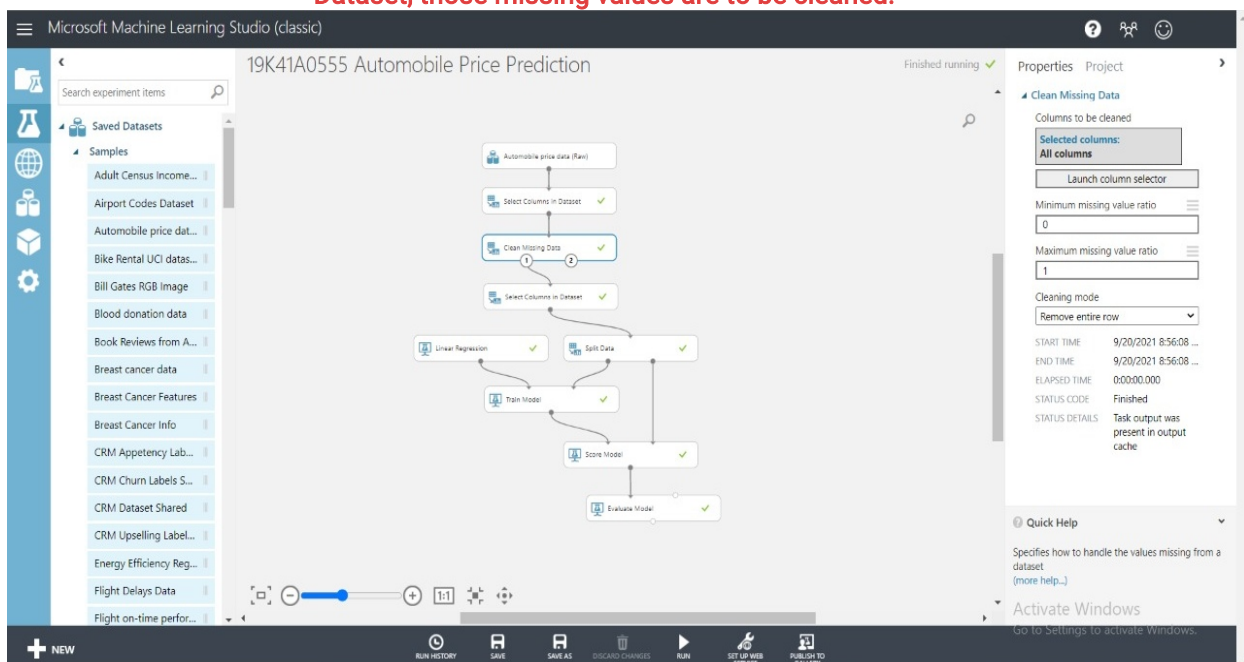
19K41A0555

Explore Data

- This basically includes data visualization to search for any missing values in the Dataset.
- If any missing values are found, then they need to be cleaned.
- Selecting the required columns and clean the data using the Clean Missing Value module (Just Drag n' Drop)



as the normalized loss has 41 missing values in the Dataset, those missing values are to be cleaned.



Data Cleaning

Split Data

- Use the Split Data module to randomly divide the input data so that the training dataset contains 70% of the original data and the testing dataset contains 30% of the original data.

The screenshot displays the Microsoft Machine Learning Studio (classic) interface for a project titled "19K41A0555 Automobile Price Prediction". The central workspace shows a workflow diagram with modules: "Automobile price data (Raw)", "Select Columns in Dataset", "Clean Missing Data", "Select Columns in Dataset", "Linear Regression", "Split Data", "Train Model", "Score Model", and "Evaluate Model". The "Split Data" module is highlighted, and its properties are shown on the right. The "Splitting mode" is set to "Split Rows", the "Fraction of rows in the first output" is 0.70, and "Randomized split" is checked. The "Random seed" is 0. The "Stratified split" is set to "False". The "Properties" pane on the right also shows the "START TIME", "END TIME", "ELAPSED TIME", "STATUS CODE", and "STATUS DETAILS".

Data Splitting

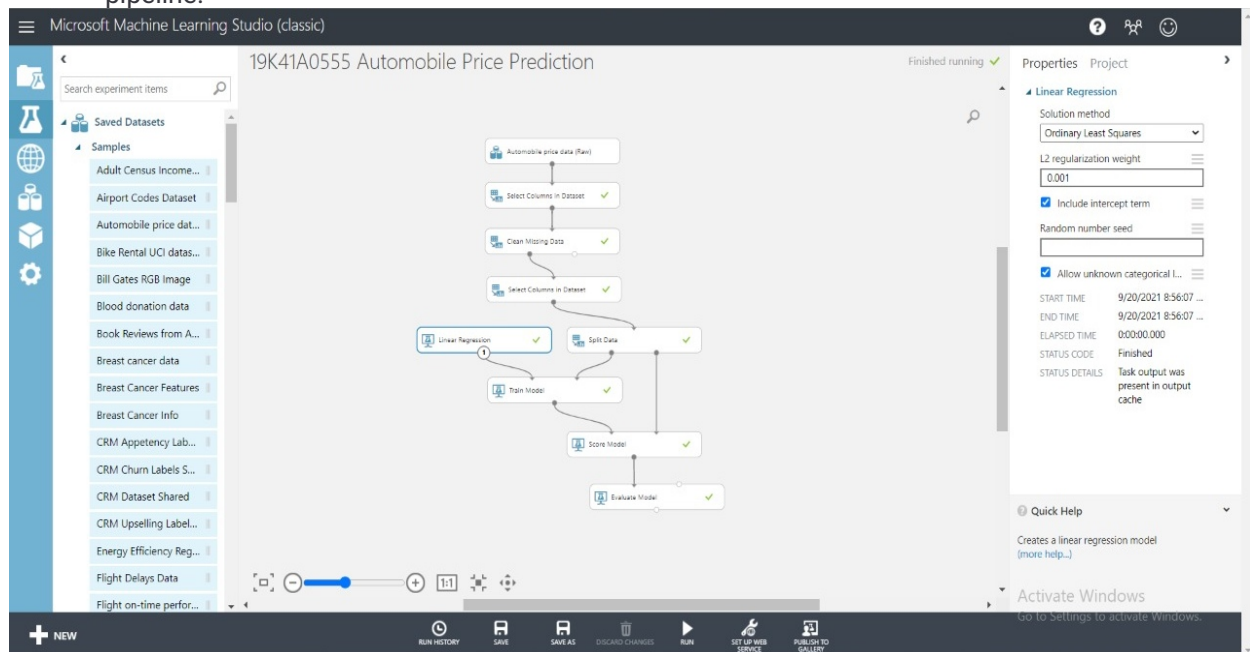
Model Training and Algorithm

The screenshot displays the Microsoft Machine Learning Studio (classic) interface for the same project. The workflow diagram is identical to the previous screenshot, but the "Train Model" module is now highlighted. The "Properties" pane on the right shows the "Train Model" configuration. The "Label column" is set to "price", and the "Column names" are "price". The "Launch column selector" button is visible. The "Properties" pane also shows the "START TIME", "END TIME", "ELAPSED TIME", "STATUS CODE", and "STATUS DETAILS".

Model Training

using Linear regression to train the model

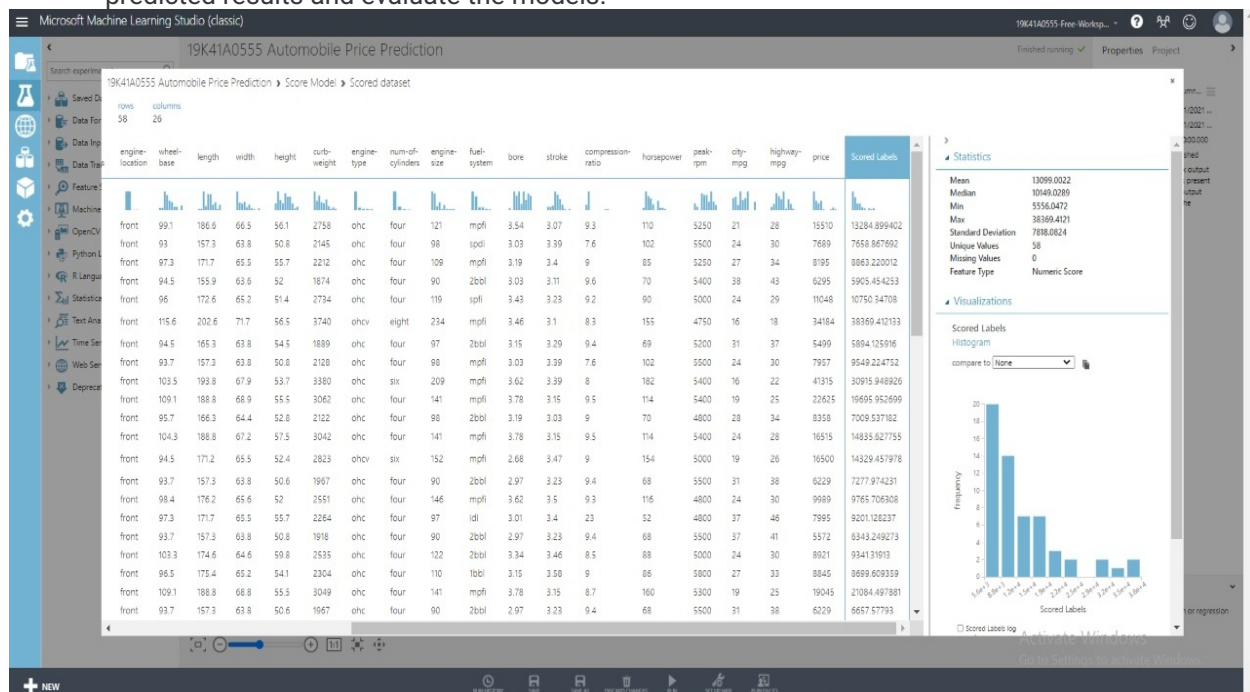
- Since the goal of this sample is to predict automobile prices, and because the label column (price) is continuous data, a regression model can be a good choice. We use Linear Regression for this pipeline.



Linear Regression

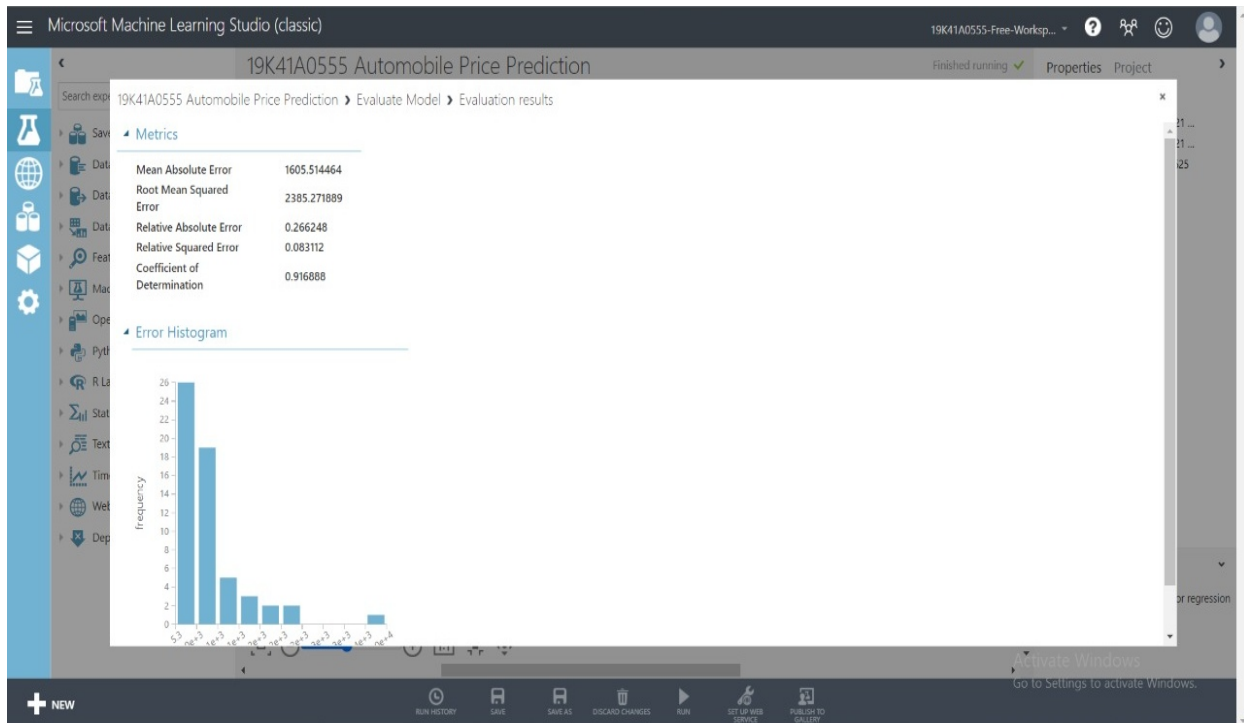
Score Model and Evaluate Model

- After the model is trained, we can use the Score Model and Evaluate Model modules to generate predicted results and evaluate the models.



Score Labels

Evaluation Results



Model Evaluation Results