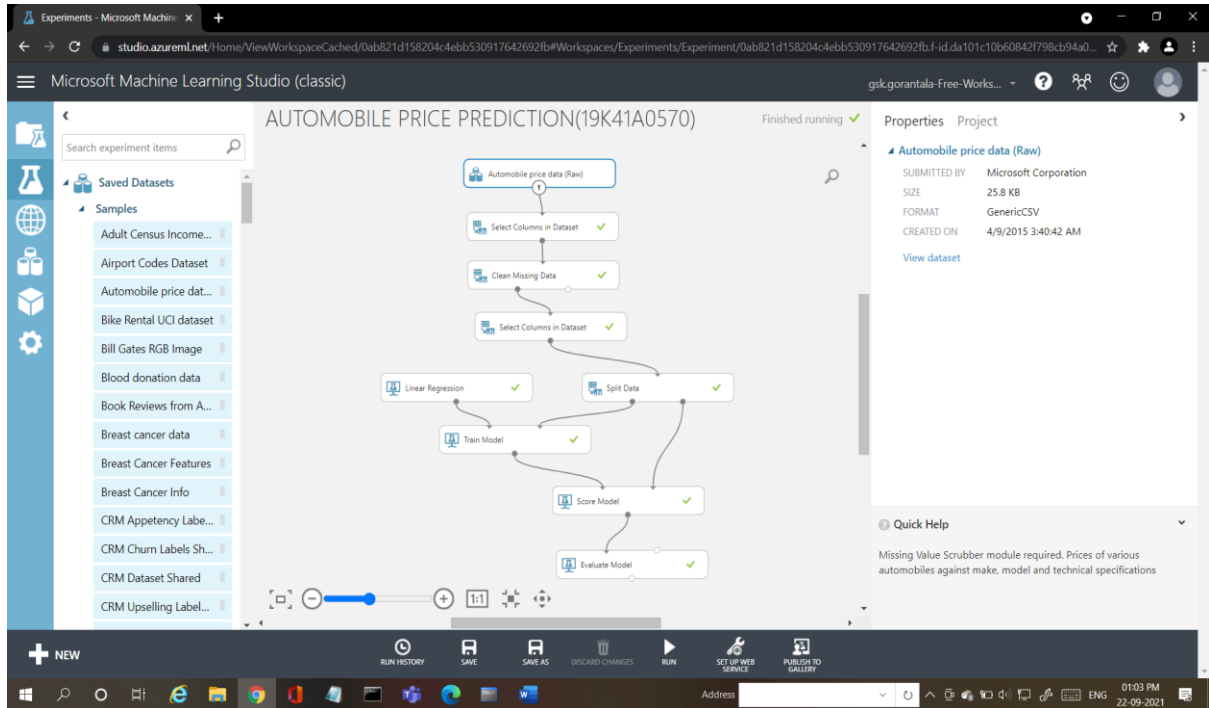


AI ASSIGNMENT (19K41A0570)

MACHINE LEARNING PROJECT WORKFLOW:-

WORKFLOW:-



EXPLORE DATA:-

The screenshot displays the Microsoft Machine Learning Studio (classic) interface. The main workspace shows an experiment titled "AUTOMOBILE PRICE PREDICTION(19K41A0570)" which has finished running. The workflow consists of the following steps:

- Automobile price data (Raw)
- Select Columns in Dataset (1)
- Clean Missing Data
- Select Columns in Dataset
- Linear Regression
- Split Data
- Train Model
- Score Model
- Evaluate Model

The "Properties" pane on the right shows the "Select Columns in Dataset" task configuration:

- Select columns:** Selected columns: All columns; Exclude column names: normalized-losses
- START TIME:** 9/22/2021 1:00:19 PM
- END TIME:** 9/22/2021 1:00:19 PM
- ELAPSED TIME:** 0:00:00.000
- STATUS CODE:** Finished
- STATUS DETAILS:** Task output was present in output cache

The left sidebar lists various datasets and samples, including "Adult Census Income...", "Airport Codes Dataset", "Automobile price dat...", "Bike Rental UCI dataset", "Bill Gates RGB Image", "Blood donation data", "Book Reviews from A...", "Breast cancer data", "Breast Cancer Features", "Breast Cancer Info", "CRM Appetency Labe...", "CRM Churn Labels Sh...", "CRM Dataset Shared", and "CRM Upselling Label...".

This screenshot shows the "Select columns" dialog box in Microsoft Machine Learning Studio (classic). The dialog is titled "Select columns" and has two tabs: "BY NAME" and "WITH RULES". The "WITH RULES" tab is selected.

The "Begin With" section shows "ALL COLUMNS" selected. Below it, the "Exclude" dropdown is set to "column names", and the text "normalized-losses" is entered in the adjacent field. A checkbox labeled "Allow duplicates and preserve column order in selection" is present and unchecked.

The background shows the same experiment workflow as the first screenshot, with the "Evaluate Model" task highlighted.

Experiments - Microsoft Machine Learning Studio (classic)

studio.azureml.net/Home/ViewWorkspaceCached/0ab821d158204c4ebb530917642692fb#Workspaces/Experiments/Experiment/0ab821d158204c4ebb530917642692fb.f-id.da101c10b60842f798cb94a0...

gsk.gorantala-Free-Works...

AUTOMOBILE PRICE PREDICTION(19K41A0570)

Finished running ✓

Properties Project

Clean Missing Data

Columns to be cleaned

Selected columns: All columns

Launch column selector

Minimum missing value ratio: 0

Maximum missing value ratio: 1

Cleaning mode: Remove entire row

START TIME: 9/22/2021 1:00:20 PM

END TIME: 9/22/2021 1:00:20 PM

ELAPSED TIME: 0:00:00.000

STATUS CODE: Finished

STATUS DETAILS: Task output was present in output cache

Quick Help

Specifies how to handle the values missing from a dataset (more help...)

Workflow:

- Automobile price data (Raw)
- Select Columns in Dataset ✓
- Clean Missing Data ✓
- Select Columns in Dataset ✓
- Linear Regression ✓
- Split Data ✓
- Train Model ✓
- Score Model ✓
- Evaluate Model ✓

Left Panel:

- Search experiment items
- Saved Datasets
- Samples
 - Adult Census Income...
 - Airport Codes Dataset
 - Automobile price dat...
 - Bike Rental UCI dataset
 - Bill Gates RGB Image
 - Blood donation data
 - Book Reviews from A...
 - Breast cancer data
 - Breast Cancer Features
 - Breast Cancer Info
 - CRM Appetency Labe...
 - CRM Churn Labels Sh...
 - CRM Dataset Shared
 - CRM Upselling Label...

Bottom Bar:

- NEW
- RUN HISTORY
- SAVE
- SAVE AS
- DISCARD CHANGES
- RUN
- SET UP WEB SERVICE
- PUBLISH TO GALLERY

Address: 01:04 PM 22-09-2021

Experiments - Microsoft Machine Learning Studio (classic)

studio.azureml.net/Home/ViewWorkspaceCached/0ab821d158204c4ebb530917642692fb#Workspaces/Experiments/Experiment/0ab821d158204c4ebb530917642692fb.f-id.da101c10b60842f798cb94a0...

gsk.gorantala-Free-Works...

AUTOMOBILE PRICE PREDICTION(19K41A0570)

Finished running ✓

Properties Project

Clean Missing Data

Columns to be cleaned

Selected columns: All columns

Launch column selector

Minimum missing value ratio: 0

Maximum missing value ratio: 1

Cleaning mode: Remove entire row

START TIME: 9/22/2021 1:00:20 PM

END TIME: 9/22/2021 1:00:20 PM

ELAPSED TIME: 0:00:00.000

STATUS CODE: Finished

STATUS DETAILS: Task output was present in output cache

Quick Help

Specifies how to handle the values missing from a dataset (more help...)

Workflow:

- Automobile price data (Raw)
- Select Columns in Dataset ✓
- Clean Missing Data ✓
- Select Columns in Dataset ✓
- Linear Regression ✓
- Split Data ✓
- Train Model ✓
- Score Model ✓
- Evaluate Model ✓

Left Panel:

- Search experiment items
- Saved Datasets
- Samples
 - Adult Census Income...
 - Airport Codes Dataset
 - Automobile price dat...
 - Bike Rental UCI dataset
 - Bill Gates RGB Image
 - Blood donation data
 - Book Reviews from A...
 - Breast cancer data
 - Breast Cancer Features
 - Breast Cancer Info
 - CRM Appetency Labe...
 - CRM Churn Labels Sh...
 - CRM Dataset Shared
 - CRM Upselling Label...

Bottom Bar:

- NEW
- RUN HISTORY
- SAVE
- SAVE AS
- DISCARD CHANGES
- RUN
- SET UP WEB SERVICE
- PUBLISH TO GALLERY

Address: 01:05 PM 22-09-2021

Select columns dialog:

BY NAME

WITH RULES

Allow duplicates and preserve column order in selection

Begin With

ALL COLUMNS NO COLUMNS

Experiments - Microsoft Machine Learning Studio (classic)

studio.azureml.net/Home/ViewWorkspaceCached/0ab821d158204c4ebb530917642692fb#Workspaces/Experiments/Experiment/0ab821d158204c4ebb530917642692fb.f-id.da101c10b60842f798cb94a0...

Microsoft Machine Learning Studio (classic)

gsk.gorantala-Free-Works...

AUTOMOBILE PRICE PREDICTION(19K41A0570)

Finished running ✓

Properties Project

Select Columns in Dataset

Select columns

Selected columns:
All columns
Column names: price

Launch column selector

START TIME 9/22/2021 1:00:20 PM
END TIME 9/22/2021 1:00:20 PM
ELAPSED TIME 0:00:00.000
STATUS CODE Finished
STATUS DETAILS Task output was present in output cache

Quick Help

Selects columns to include or exclude from a dataset in an operation. Formerly known as Project Columns. (more help...)

Workflow:

- Automobile price data (Raw)
- Select Columns in Dataset ✓
- Clean Missing Data ✓
- Select Columns in Dataset ✓
- Linear Regression ✓
- Split Data ✓
- Train Model ✓
- Score Model ✓
- Evaluate Model ✓

Left Panel:

- Search experiment items
- Saved Datasets
- Samples
 - Adult Census Income...
 - Airport Codes Dataset
 - Automobile price dat...
 - Bike Rental UCI dataset
 - Bill Gates RGB Image
 - Blood donation data
 - Book Reviews from A...
 - Breast cancer data
 - Breast Cancer Features
 - Breast Cancer Info
 - CRM Appetency Labe...
 - CRM Churn Labels Sh...
 - CRM Dataset Shared
 - CRM Upselling Label...

Bottom Bar:

- NEW
- RUN HISTORY
- SAVE
- SAVE AS
- DISCARD CHANGES
- RUN
- SET UP WEB SERVICE
- PUBLISH TO GALLERY

Address: 01:05 PM 22-09-2021

Experiments - Microsoft Machine Learning Studio (classic)

studio.azureml.net/Home/ViewWorkspaceCached/0ab821d158204c4ebb530917642692fb#Workspaces/Experiments/Experiment/0ab821d158204c4ebb530917642692fb.f-id.da101c10b60842f798cb94a0...

Microsoft Machine Learning Studio (classic)

gsk.gorantala-Free-Works...

AUTOMOBILE PRICE PREDICTION(19K41A0570)

Finished running ✓

Properties Project

Select Columns in Dataset

Select columns

Selected columns:
All columns
Column names: price

Launch column selector

START TIME 9/22/2021 1:00:20 PM
END TIME 9/22/2021 1:00:20 PM
ELAPSED TIME 0:00:00.000
STATUS CODE Finished
STATUS DETAILS Task output was present in output cache

Quick Help

Selects columns to include or exclude from a dataset in an operation. Formerly known as Project Columns. (more help...)

Workflow:

- Automobile price data (Raw)
- Select Columns in Dataset ✓
- Clean Missing Data ✓
- Select Columns in Dataset ✓
- Linear Regression ✓
- Split Data ✓
- Train Model ✓
- Score Model ✓
- Evaluate Model ✓

Left Panel:

- Search experiment items
- Saved Datasets
- Samples
 - Adult Census Income...
 - Airport Codes Dataset
 - Automobile price dat...
 - Bike Rental UCI dataset
 - Bill Gates RGB Image
 - Blood donation data
 - Book Reviews from A...
 - Breast cancer data
 - Breast Cancer Features
 - Breast Cancer Info
 - CRM Appetency Labe...
 - CRM Churn Labels Sh...
 - CRM Dataset Shared
 - CRM Upselling Label...

Bottom Bar:

- NEW
- RUN HISTORY
- SAVE
- SAVE AS
- DISCARD CHANGES
- RUN
- SET UP WEB SERVICE
- PUBLISH TO GALLERY

Address: 01:05 PM 22-09-2021

SPLIT DATA:-

The screenshot displays the Microsoft Machine Learning Studio interface for an experiment titled "AUTOMOBILE PRICE PREDICTION(19K41A0570)". The workflow consists of the following steps:

- Automobile price data (Raw)
- Select Columns in Dataset
- Clean Missing Data
- Select Columns in Dataset
- Linear Regression
- Split Data
- Train Model
- Score Model
- Evaluate Model

The "Split Data" step is highlighted, and its properties are shown on the right:

- Splitting mode:** Split Rows
- Fraction of rows in the first output dataset:** 0.70
- Randomized split:** ☒
- Random seed:** 0
- Stratified split:** False
- START TIME:** 9/22/2021 1:00:20 PM
- END TIME:** 9/22/2021 1:00:20 PM
- ELAPSED TIME:** 0:00:00.000
- STATUS CODE:** Finished
- STATUS DETAILS:** Task output was present in output cache

The bottom of the interface shows a toolbar with options like "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SET UP WEB SERVICE", and "PUBLISH TO GALLERY". The Windows taskbar at the very bottom shows the time as 01:06 PM on 22-09-2021.

USING LINEAR REGRESSION TO TRAIN THE MODEL:-

The screenshot displays the Microsoft Machine Learning Studio (classic) interface. The main workspace is titled "AUTOMOBILE PRICE PREDICTION(19K41A0570)" and shows a workflow diagram. The workflow starts with the "Automobile price data (Raw)" dataset, followed by "Select Columns in Dataset", "Clean Missing Data", and another "Select Columns in Dataset". The data is then split into two paths: one leading to "Linear Regression" and another to "Split Data". The "Linear Regression" module is configured with the "Ordinary Least Squares" solution method, an L2 regularization weight of 0.001, and the "Include intercept term" checkbox is checked. The "Split Data" module is also configured. The workflow continues with "Train Model", "Score Model", and "Evaluate Model". The "Properties" pane on the right shows the configuration for the "Linear Regression" module, including the solution method, regularization weight, and intercept term. The "Quick Help" section provides a brief description of the module.

Microsoft Machine Learning Studio (classic) interface showing an experiment workflow for "AUTOMOBILE PRICE PREDICTION(19K41A0570)". The workflow includes modules for data selection, cleaning, splitting, training a Linear Regression model, scoring, and evaluation. The "Linear Regression" module is configured with the "Ordinary Least Squares" solution method, an L2 regularization weight of 0.001, and the "Include intercept term" checkbox is checked. The "Properties" pane on the right shows the configuration for the "Linear Regression" module, including the solution method, regularization weight, and intercept term. The "Quick Help" section provides a brief description of the module.

Workflow steps:

- Automobile price data (Raw)
- Select Columns in Dataset
- Clean Missing Data
- Select Columns in Dataset
- Linear Regression
- Split Data
- Train Model
- Score Model
- Evaluate Model

Properties for Linear Regression:

- Solution method: Ordinary Least Squares
- L2 regularization weight: 0.001
- Include intercept term: ☒
- Random number seed:
- Allow unknown categorical levels: ☒

Quick Help:

Creates a linear regression model (more help...)

TRAIN MODELLING AND ALGORITHM:-

The screenshot displays the Microsoft Machine Learning Studio interface for an experiment titled "AUTOMOBILE PRICE PREDICTION(19K41A0570)". The workflow is as follows:

- Automobile price data (Raw)** (Input)
- Select Columns in Dataset** (Green checkmark)
- Clean Missing Data** (Green checkmark)
- Select Columns in Dataset** (Green checkmark)
- Linear Regression** (Green checkmark)
- Split Data** (Green checkmark)
- Train Model** (Green checkmark)
- Score Model** (Green checkmark)
- Evaluate Model** (Green checkmark)

The **Properties** pane on the right shows the **Train Model** task details:

- Label column:** price
- Selected columns:** price
- Column names:** price
- Launch column selector:** (button)
- START TIME:** 9/22/2021 1:00:20 PM
- END TIME:** 9/22/2021 1:00:20 PM
- ELAPSED TIME:** 0:00:00.000
- STATUS CODE:** Finished
- STATUS DETAILS:** Task output was present in output cache

The bottom toolbar includes buttons for **NEW**, **RUN HISTORY**, **SAVE**, **SAVE AS**, **DISCARD CHANGES**, **RUN**, **SET UP WEB SERVICE**, and **PUBLISH TO GALLERY**.

The screenshot shows the same Microsoft Machine Learning Studio interface, but with the **Select a single column** dialog box open. The dialog box has the following fields:

- BY NAME:** (empty)
- WITH RULES:** (checked)
- Include:** (dropdown menu)
- column names:** (dropdown menu)
- price** (text input)

The **Properties** pane on the right shows the **Train Model** task details, which are the same as in the first screenshot.

SCORE MODEL AND EVALUATION MODEL:-

Microsoft Machine Learning Studio (classic) interface showing the workflow for 'AUTOMOBILE PRICE PREDICTION(19K41A0570)'. The workflow includes steps: 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 'Score Model' step is highlighted, and the 'Properties' pane on the right shows the 'Score Model' properties.

Score Model Properties:

- Append score columns to output: ☒
- START TIME: 9/22/2021 1:00:20 PM
- END TIME: 9/22/2021 1:00:20 PM
- ELAPSED TIME: 0:00:00.000
- STATUS CODE: Finished
- STATUS DETAILS: Task output was present in output cache

Microsoft Machine Learning Studio (classic) interface showing the workflow for 'AUTOMOBILE PRICE PREDICTION(19K41A0570)'. The 'Score Model' step is highlighted, and the 'Scored dataset' is displayed in a table view. The table shows columns: symboling, make, fuel-type, aspiration, num-of-doors, body-style, drive-wheels, engine-location, and wheel-base. The 'Statistics' and 'Visualizations' panes are also visible.

Scored dataset:

symboling	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location	wheel-base
2	saab	gas	std	four	sedan	fwd	front	99.1
1	mitsubishi	gas	turbo	two	hatchback	fwd	front	93
2	volkswagen	gas	std	four	sedan	fwd	front	97.3
1	chevrolet	gas	std	two	hatchback	fwd	front	94.5
2	isuzu	gas	std	two	hatchback	rwd	front	96
-1	mercedes-benz	gas	std	four	sedan	rwd	front	115.6
1	nissan	gas	std	two	sedan	fwd	front	94.5
1	dodge	gas	turbo	two	hatchback	fwd	front	93.7
0	bmw	gas	std	two	sedan	rwd	front	103.5
-1	volvo	gas	turbo	four	sedan	rwd	front	109.1

EVALUATION RESULTS:-

