



TUGAS ASINKRON 1 MINGGU 1 BULAN 2 KELAS DAI-002

Use automated machine learning in Azure Machine Learning

Oleh :

KELOMPOK 4

- Dhea Amalia Ariantoputri
- Dinda Ega Fajarwati
- Fadhil Rausyanfikir
- Dwira Kurnia Larasati
 - Renaka Agusta
 - Eka Aperinda Putri
 - Dinna Nurfadlillah

Mentor :

Noviyanti Tri Mareta Sagala, S.TI., M.Sc

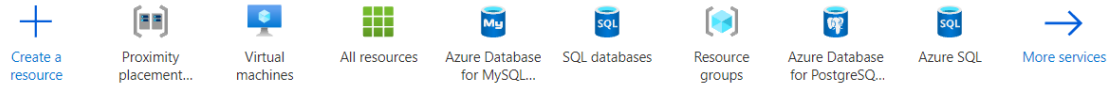
Asisten Mentor :

Ainul Zakiy

STUDI INDEPENDEN PT. MICROSOFT INDONESIA
SEPTEMBER 2021

Pada tugas asinkron ini setiap kelompok diminta untuk melatih sebuah model Machine Learning dan menggunakan model tersebut sebagai sebuah layanan mengikuti pada modul di Microsoft Learn yakni use automated machine learning in Azure Machine Learning, dengan beberapa perubahan yang diminta. Adapun langkah-langkah yang harus dilakukan diantaranya :

Azure services



1. Masuk ke Portal Azure dan buat resource baru dengan menekan tombol "Create a resource"

Machine Learning

Microsoft



Machine Learning [Add to Favorites](#)

Microsoft

★ 4.3 (113 Azure ratings)

Create

2. Dikarenakan kita akan melakukan training model maka pilihlah layanan Machine Learning dengan menekan tombol create pada bagian tersebut

Machine learning

Create a machine learning workspace

Basics Networking Advanced Tags Review + create

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * ⓘ

Azure for Students

Resource group * ⓘ

(New) cloud-shell-southasia

Create new

Workspace details

Specify the name and region for the workspace.

Workspace name * ⓘ

ml280921kel4

✓

Region * ⓘ

Southeast Asia

▼

Storage account * ⓘ

(new) ml280921kel45357519220

▼

Create new

Key vault * ⓘ

(new) ml280921kel41123808541

▼

Create new

Application insights * ⓘ

(new) ml280921kel46868462372

▼

Create new

Container registry * ⓘ

None

▼

Create new

Basics Networking Advanced Tags Review + create

Basics

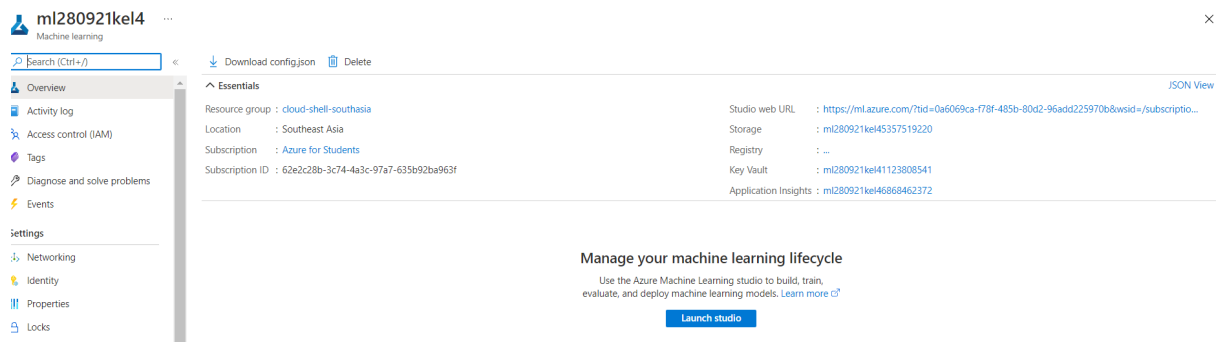
Subscription	Azure for Students
Resource group	(New) cloud-shell-southasia
Region	Southeast Asia
Workspace name	ml280921kel4
Storage account	(new) ml280921kel45357519220
Key vault	(new) ml280921kel41123808541
Application insights	(new) ml280921kel46868462372
Container registry	None

Networking

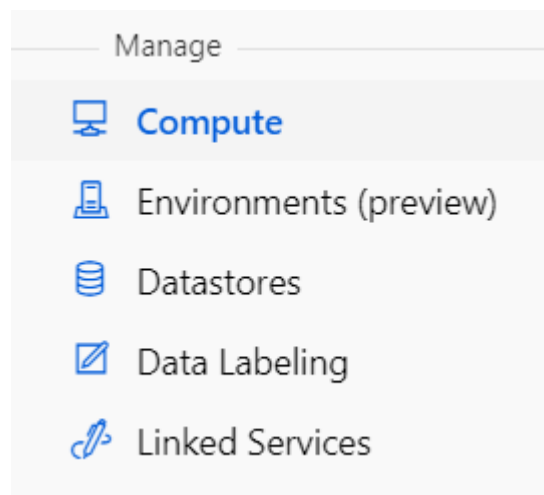
Connectivity method	Public endpoint (all networks)
---------------------	--------------------------------

Advanced

Identity type	System assigned
Encryption type	Microsoft-managed keys
Enable HBI Flag	Disabled



- Sebelum memanfaatkan layanan Azure Machine Learning, diharuskan mengisi beberapa data seperti resource group yang digunakan dan workspace. Pada workspace kali ini region yang dipilih adalah southeastasia dikarenakan merupakan region terdekat. Setelah seluruh data diisi masuk ke tab review+create dan tekan tombol create setelah seluruh data yang diisi dirasa sudah benar.



- Pada bagian pemilihan jenis instance pilihlah "Compute" dikarenakan kita membutuhkan layanan komputasi untuk mentraining model

[Compute instances](#)
[Compute clusters](#)
[Inference clusters](#)
[Attached computes](#)

Create compute instance

☒ Required Settings

☐ Advanced Settings

Configure required settings

Select the name and virtual machine size you would like to use for your compute instance. Please note that a compute instance can not be shared. It can only be used by a single assigned user. By default, it will be assigned to the creator and you can change this to a different user in the advanced settings section.

Compute name *

ci2809kel4

Location

southeastasia

Virtual machine type

☒ CPU ☐ GPU

Virtual machine size

☐ Select from recommended options ☒ Select from all options

+ Add filter

Search by VM name...

Showing 77 VM sizes | Current selection: Standard_DS11_v2

Total available quota: 4 cores

5. Pada tahap ini kita akan melakukan konfigurasi terhadap Compute instance,
- Pada penggunaan nama, kami menginput compute name dengan : “ci2809kel4”
 - Pada tipenya disini kami memilih menggunakan CPU
 - Untuk Ukuran Mesin Virtual, kami memilih semua opsi dan menggunakan Standard_DS11_v2

Create compute cluster

☒ Virtual Machine

☐ Advanced Settings

Select virtual machine

Select the virtual machine size you would like to use for your compute cluster.

Location *

Southeast Asia

Virtual machine priority

☒ Dedicated ☐ Low priority

Virtual machine type

☒ CPU ☐ GPU

Virtual machine size

☐ Select from recommended options ☒ Select from all options

+ Add filter

Search by VM name...

Showing 93 VM sizes | Current selection: Standard_DS11_v2

Total available quota: 2 cores

6. Gambar diatas menunjukkan konfigurasi pada tab Computer Cluster yang berisi
- lokasi pada Southeast Asia
 - prioritas Virtual Machine (VM) adalah dedicated
 - tipe VM adalah CPU
 - untuk ukuran Virtual machine kami memilih select from all option, lalu cari dan pilih Standard_DS11_v2

Configure Settings
Configure compute cluster settings for your selected virtual machine size.

Name	Category	Cores	Available quota	RAM	Storage	Cost/Node
Standard_DS11_v2	Memory optimized	2	2 cores	14 GB	28 GB	\$0.19/hr

Compute name * ⓘ
cc2809kel4 ⓘ

Minimum number of nodes * ⓘ
0

Maximum number of nodes * ⓘ
2

ⓘ You have insufficient subscription quota to scale up to more than 1 nodes.

Idle seconds before scale down * ⓘ
120

☐ Enable SSH access ⓘ

> Advanced settings

Back Create Download a template for automation Cancel

7. Gambar 7 merupakan lanjutan dari step 6 (masih di Computer Cluster Tab) yaitu :

- mengisi size VM nya adalah Standard_DS11_v2,
- isi computer name dengan unique name
- number of nodes minimum adalah 0 dan maximal adalah 2
- idle seconds before scale down adalah 120
- jangan enable SSH access.

Compute

Compute instances **Compute clusters** Inference clusters Attached computes

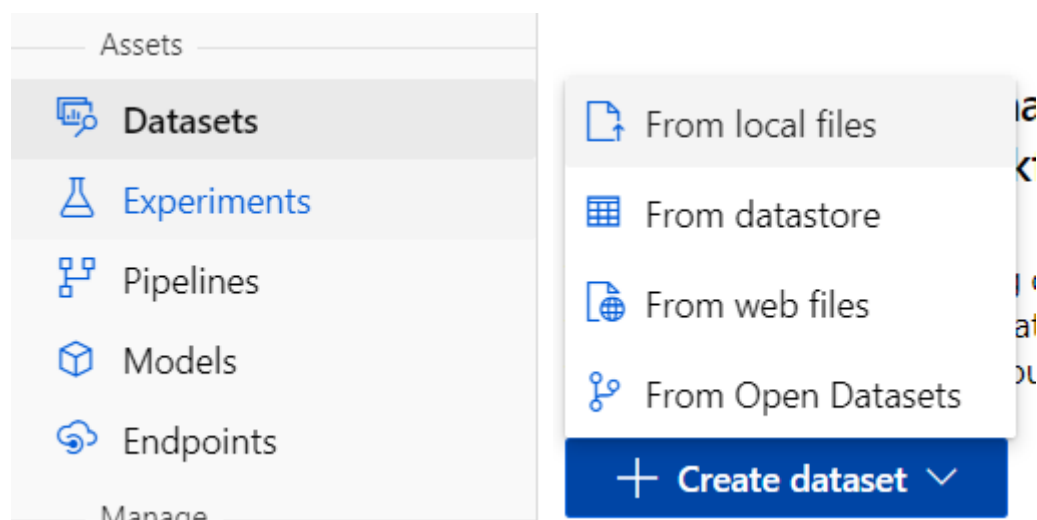
+ New Refresh Delete Edit columns Reset view View quota

Search

State Location All filters Clear all

Name	State	Size	Location	Created on ↓	Active runs	Idle nodes	Busy nodes
cc2809kel4	Creating	STANDARD_DS11_V2	southeastasia	Sep 28, 2021 8:39 AM	0	0	0

8. Setelah itu layanan compute cluster akan dibuat sesuai dengan konfigurasi yang telah ditentukan



9. Setelah Compute cluster berhasil dibuat langkah selanjutnya adalah membuat dataset. Pada bagian assets pilihlah dataset dan tekan tombol create dataset

Basic info

Name *



automobile_price *

Dataset type * ⓘ

Tabular



Description

automobile price



10. Selanjutnya setelah membuat dataset, kami mengatur penamaan file, dengan nama :
automobile_price, dengan type dataset : tabular, dan mengisi description automobile price.

Datastore and file selection

Select or create a datastore * ⓘ

workspaceblobstore



> Create new datastore

Select files for your dataset *

These files will be uploaded to your selected datastore and made available in your workspace. Supported file types include: delimited (i.e. csv, tsv), Parquet, JSON Lines, and plain text.

Upload ▾

1 files selected. Total size 0.02500 MiB. 0/1 files uploaded

File name	Size (MiB)	Upload %	Status
automobile_price.csv	0.02500		

Upload path

UI

Files will be uploaded to '\$(Upload path)/09-28-2021_014703.UTC'

☐

Skip data validation ⓘ

11. Pada bagian ini kita melakukan beberapa konfigurasi terkait dengan datastore sebagai tempat penyimpanan, disini kami menggunakan nama "workspaceblobstore". Kemudian kita mengunggah file dataset yang telah diberikan yaitu "automobile.csv"

Basic info

Datastore and file selection

Settings and preview

Schema

Confirm details

Settings and preview

These settings were automatically detected. Please verify that the selections were made correctly or update.

File format

Delimited

Delimiter Example

Comma

Field1,Field2,Field3

Encoding

UTF-8

Column headers

Only first file has headers

Skip rows

None

☐ Dataset contains multi-line data ⓘ

ⓘ Note: Processing tabular files with multi-line data is slower because multiple CPU cores cannot be used to ingest the data in parallel. Checking this option may result in slower processing times.

Id	symboling	normalized-losses	make	fuel-type
1	3	?	alfa-romero	gas

12. Pada bagian Setting & Preview kita melakukan konfigurasi pada File Format yaitu Delimited dan Delimiternya adalah Comma dikarenakan dataset yang diunggah berekstensi ".csv". Pada Column Header kita memilih Only first file headers dikarenakan hanya baris pertama pada file csv yang merupakan nama kolom. Pada bagian Skip Rows dipilih none dikarenakan kita akan menggunakan seluruh data yang ada pada dataset.

Schema

Column types are auto-detected based on the first 200 rows of the data. Please make any necessary adjustments. Values not aligning with the specified column type will fail conversion and would be either null-filled or replaced with error value.

Include	Column name	Properties ⓘ	Type
<input checked="" type="checkbox"/>	compression-ratio	Not applicable to selected type	Decimal (dot '.')
<input checked="" type="checkbox"/>	horsepower	Not applicable to selected type	String
<input checked="" type="checkbox"/>	peak-rpm	Not applicable to selected type	String
<input checked="" type="checkbox"/>	city-mpg	Not applicable to selected type	Integer
<input checked="" type="checkbox"/>	highway-mpg	Not applicable to selected type	Integer
<input checked="" type="checkbox"/>	price	Not applicable to selected type	Integer

13. Pada bagian schema, kita akan menentukan tipe data dari setiap kolom. Pada kolom "price" kita mengubahnya menjadi integer dikarenakan pada regresi tipe data target haruslah bertipe numerik

Confirm details

Basic info

Name
automobile_price

Dataset type
Tabular

Description
automobile price

Datastore and file selection

Datastore
workspaceblobstore

Selected files (1)
automobile_price.csv

Path
UI/09-28-2021_014703.UTC/automobile_price.csv

File settings

File format
Delimited

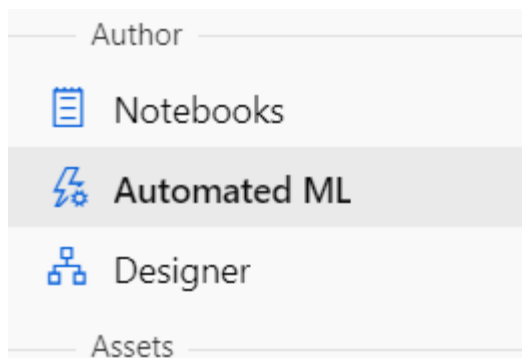
Delimiter
Comma

Encoding
UTF-8

Column headers
Only first file has headers

Back Create Cancel

14. kemudian masuk ke konfirmasi pada detail. setelah dicek dan benar, klik create.



15. Pada Azure Machine Learning Studio, klik Automated ML yang berada di bawah 'Author'.

Create a new Automated ML run

Select dataset

Select an input dataset from the list below, or create a new dataset. Automated ML currently only supports tabular data for authoring runs.

+ Create dataset Refresh Show supported datasets only

Search

Showing 1-1 of 1 datasets

Page size: 25

Dataset name	Dataset type	Created on ↓	Modified on
automobile-price	Tabular	Sep 28, 2021 9:02 AM	Sep 28, 2021 9:02 AM

16. kemudian pilih dataset yang tersedia, yaitu automobile-price, lalu pilih next

Create a new Automated ML run

- ✓ Select dataset
- **Configure run**
- Select task and settings

Configure run

Select from existing experiments or create a new experiment, then select the target column and training compute. [Learn more on how to configure the experiment.](#)

Dataset

automobile-price ([View dataset](#))

Experiment name *

☒ Create new

New experiment name

mslearn-automobile-price

Target column * ⓘ

price (Integer)

Select compute cluster * ⓘ

cc2809kel4

[Create a new compute](#) [Refresh compute](#)

17. masuk ke bagian konfigurasi (Configure Run) masukkan pengaturan sebagai berikut :

- new experiment name : mslearn-automobile-price
- target column : price (integer)
- select compute cluster : computer cluster yang telah dibuat sebelumnya

Create a new Automated ML run

- ✓ Select dataset
- ✓ Configure run
- **Select task and settings**

Select task type

Select the machine learning task type for the experiment. To fine tune the experiment, choose additional configuration or featurization settings.



Classification

To predict one of several categories in the target column. yes/no, blue, red, green.



Regression

To predict continuous numeric values



Time series forecasting

To predict values based on time

18. Pada tab select and settings, pilih regression untuk task type nya

[View additional configuration settings](#) [View featurization settings](#)

Additional configurations

Primary metric ⓘ

Normalized root mean squared error

☒ Explain best model ⓘ

Blocked algorithms ⓘ

FastLinearRegressor × ExtremeRandomTrees × ElasticNet ×
GradientBoosting × DecisionTree × KNN × LassoLars × SGD ×
RandomForest ×

A list of algorithms that Automated ML will not use during training.

Exit criterion

Training job time (hours) ⓘ

0.5

Metric score threshold ⓘ

0.05

Validation

Validation type ⓘ

Auto

Concurrency

Max concurrent iterations ⓘ

2

19. Pada additional configuration settings, setting

primary metric : normalized root mean squared error

blocked algorithms : seperti pada gambar

training job : 0.5 (ini menyebabkan percobaan berakhir setelah maksimal 30 menit) ->

ketika hendak mencoba 15 menit (0.25 hour) tetapi error karena requirementnya adalah minimum 0.5 hours (30 menit)

metric score threshold : 0.05 (ini menyebabkan percobaan berakhir jika model mencapai skor metrik kesalahan kuadrat rata-rata akar yang dinormalisasi 0,05 atau kurang)

Featurization



Feature selection identifies the actions performed on the dataset to prepare the data for training. This will not impact the input data needed for inferencing i.e., if columns are excluded from training, the excluded columns will still be required as input for inferencing on the model. [Learn more about Automated ML's featurization.](#)

☒ Enable featurization ⓘ

Column name	Included	Feature type	Impute with	Data example
symboling	<input checked="" type="checkbox"/>	<div>Auto</div>	<div>Auto</div>	3, 3, 1
normalized-losses	<input checked="" type="checkbox"/>	<div>Auto</div>	<div>Auto</div>	?, ?, ?
make	<input checked="" type="checkbox"/>	<div>Auto</div>	<div>Auto</div>	alfa-romero, alfa-romero, alfa-romero
fuel-type	<input checked="" type="checkbox"/>	<div>Auto</div>	<div>Auto</div>	gas, gas, gas
aspiration	<input checked="" type="checkbox"/>	<div>Auto</div>	<div>Auto</div>	std, std, std
num-of-doors	<input checked="" type="checkbox"/>	<div>Auto</div>	<div>Auto</div>	two, two, two

20. Pada featurization pastikan selected semua

helpful_basil_9g16mzsp (Run 3) ✎

Refresh Cancel Delete

Details Data guardrails Models Outputs + logs Child runs Snapshot

✓ Completed

Created

Sep 28, 2021 10:33 AM

Started

Sep 28, 2021 10:33 AM

Duration

9m 44.83s

Compute duration

9m 44.83s

Compute target

cc2809kel4

Run ID

AutoML_7477c097-d828-4f8a-ba20-d014f4a3b920

Script name

--

Created by

Dhea Ariantoputri

Input datasets

Input name: training_data, Dataset: automobile_price: Version 1

Output datasets

None

StandardScalerWrapper, XGBoostRegressor

Hyperparameters

[View hyperparameters](#)

Normalized root mean squared error

0.04991 [View all other metrics](#)

Sampling

100.00 % ⓘ

Registered models

No registration yet

Deploy status

No deployment yet

Run summary

Task type

Regression [View configuration settings](#)

Featurization

Auto

Primary metric

Normalized root mean squared error

Experiment name

apkel4

21. setelah sukses, maka akan tampil seperti ini

Best model summary

Algorithm name

StandardScalerWrapper, XGBoostRegressor

22. Untuk melihat review secara detail, klik pada algorithm name

green_corn_xgk3ywpd (Run 9) [🔗](#)

[Refresh](#) [Deploy](#) [Download](#) [Explain model](#) [Cancel](#) [Delete](#)

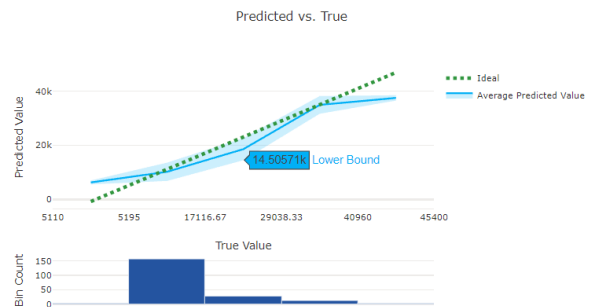
[Details](#) [Model](#) [Explanations \(preview\)](#) **Metrics** [Data transformation \(preview\)](#) [Outputs + logs](#) [Images](#) [Child runs](#) [Snapshot](#) [Monitoring \(preview\)](#)

Select a metric to see a visualization or table of the data.

[Search](#)

- ☒ mean_absolute_error
- ☒ mean_absolute_percentage_error
- ☒ median_absolute_error
- ☒ normalized_mean_absolute_error
- ☐ normalized_median_absolute_er...
- ☐ normalized_root_mean_squared...
- ☐ normalized_root_mean_squared...
- ☒ predicted_true
- ☐ r2_score
- ☒ residuals
- ☐ root_mean_squared_error
- ☐ root_mean_squared_log_error
- ☐ spearman_correlation

View as: ☒ Chart ☐ Table



23. Klik Metrics dan pilih checklist predicted_true, maka akan tampil grafik seperti diatas. Dari hasil grafik diatas menunjukkan diagonal trend antara nilai yang diprediksi dan nilai yang benar (true). nilai yang diprediksi ditunjukkan pada garis titik-titik, sedangkan nilai yang true ditunjukkan pada garis lurus. semakin dekat nilai true dengan nilai yang diprediksi, maka semakin bagus modelnya. terlihat pada grafik ada sedikit penyimpangan pada nilai true, ini karena model tidak bisa 100% sempurna memprediksi, tetapi ada error. sedangkan histogram dibawah diagonal trend menunjukkan distribusi dari nilai true nya. semakin tinggi nilai histogramnya, maka semakin dekat nilai true dengan nilai prediksi, sehingga menunjukkan semakin akurat model tersebut.

green_corn_xgk3ywpd (Run 9) [🔗](#)

[Refresh](#) [Deploy](#) [Download](#) [Explain model](#) [Cancel](#) [Delete](#)

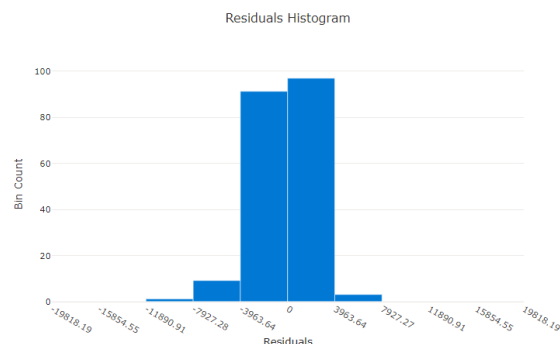
[Details](#) [Model](#) [Explanations \(preview\)](#) **Metrics** [Data transformation \(preview\)](#) [Outputs + logs](#) [Images](#) [Child runs](#) [Snapshot](#) [Monitoring \(preview\)](#)

Select a metric to see a visualization or table of the data.

[Search](#)

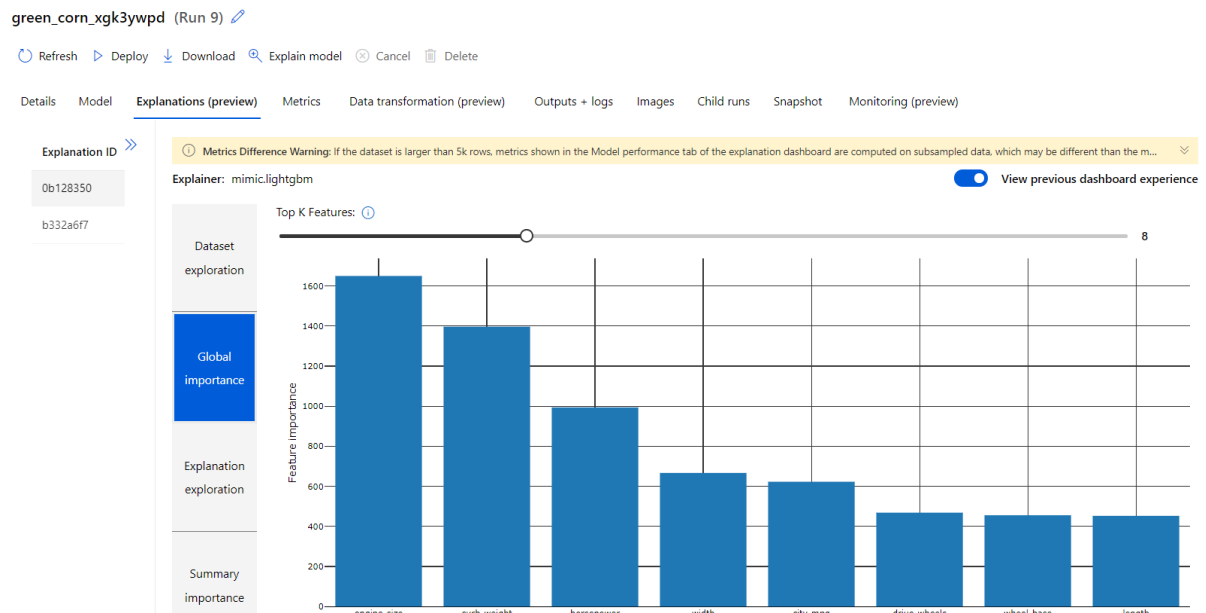
- ☒ mean_absolute_error
- ☒ mean_absolute_percentage_error
- ☒ median_absolute_error
- ☒ normalized_mean_absolute_error
- ☐ normalized_median_absolute_er...
- ☐ normalized_root_mean_squared...
- ☐ normalized_root_mean_squared...
- ☒ predicted_true
- ☐ r2_score
- ☒ residuals
- ☐ root_mean_squared_error
- ☐ root_mean_squared_log_error
- ☐ spearman_correlation

View as: ☒ Chart ☐ Table








24. Jika residuals dipilih maka akan tampil charts seperti ini. residual adalah selisih antara nilai sebenarnya dan nilai prediksi, ini juga menunjukkan jumlah error dari model tersebut. residual histogram menunjukkan frekuensi dari jangkauan nilai residual ditunjukkan oleh varians antara nilai true dan nilai prediksi (disebut juga dengan error). Dari chart ini

menunjukkan bahwa histogram paling tinggi berada di sekitar 0, serta makin ke ujung, histogramnya makin rendah. hal ini berarti bahwa frekuensi terjadinya nilai residual paling banyak adalah di nilai 0, alias kebanyakan error nilainya kecil.



25. Chart ini menunjukkan seberapa banyak setiap fitur dalam dataset memengaruhi prediksi label

green_corn_xgk3ywpd (Run 9) 

 Refresh  Deploy  Download 

26. setelah itu, lanjutkan mendeploy service yang sudah dibuat.

Deploy a model

Name * ⓘ

am-price-kel4

Description ⓘ

predict automobile price

Compute type * ⓘ

Azure Container Instance

Models: AutoML7477c097d2

Enable authentication

☒

Keys can be found on the endpoint details page.

This model supports no-code deployment. You may optionally override the default environment and driver file.

Use custom deployment assets

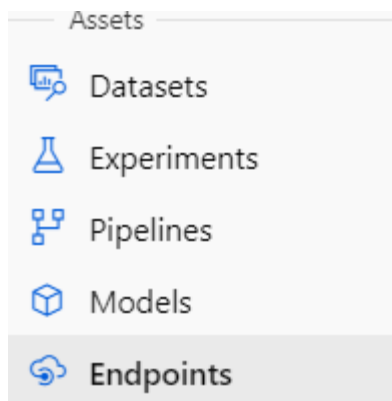
☐ Use custom deployment assets

> Advanced

Deploy

Cancel

27. konfigurasi yang perlu diisi, seperti name, untuk deskripsi optional, pilih compute typenya azure container instance dan select enable authentication. Setelah itu klik deploy.






28. setelah berhasil di deploy, klik tab endpoints

am-price-kel4

Details Test Consume Deployment logs


Basic consumption info


REST endpoint
 


Authentication
Primary key
6h2oyJSFSu7cg3Xz48uSP8pPrWS9wRDN  [Regenerate](#)
Secondary key
Aez2YXET9UtTJYBfn36zcfaEmGbXZMwf  [Regenerate](#)


pada tab consume akan tampil rest endpoint beserta primary key


Author


 Notebooks


 Automated ML


 Designer


 Users


 dhea.amalia.ariantoputri-2019...


 Create new file

 Create new folder

 Upload files


 Upload folder

 Copy folder path

 Open terminal


29. Lalu klik tab notebooks dan buat file baru pada file yang telah ada

Create new file

File location
Users/dhea.amalia.ariantoputri-2019  [Edit location](#)

File name *

File type

Notebook (*.ipynb) 

☒ **Overwrite if already exists**

This will replace any existing file with the same name

CreateCancel

30. beri nama file dan tipe filenya. Disini menggunakan .ipynb

```
ci2809kel4 · Kernel idle Last saved a few seconds ago
We would love to hear your feedback on the notebooks experience! Please take a few minutes to complete our survey.

1 endpoint = 'http://835c1ab8-f013-4425-ab52-4420b2b1e6bc.southeastasia.azurecontainer.io/score' #Replace with your endpoint
2 key = '6h2oy3SF5u7cg3Xz48uSP8pPrW59wRDN' #Replace with your key
3
4 import urllib.request
5 import json
6 import os
7
8 # Prepare the input data
9 data = {
10     "Inputs": {
11         "WebServiceInput0":
12             [
13                 {
14                     'symboling': 3,
15                     'normalized-losses': None,
16                     'make': "alfa-romero",
17                     'fuel-type': "gas",
18                     'aspiration': "std",
19                     'num-of-doors': "two",
20                     'body-style': "convertible",
21                     'drive-wheels': "rwd",
22                     'engine-location': "front",
23                     'wheel-base': 88.6,
24                     'length': 168.8,
25                     'width': 64.1,
26                     'height': 48.8,
27                     'curb-weight': 2548,
28                     'engine-type': "dohc",
29                     'num-of-cylinders': "four",
30                     'engine-size': 130,
31                     'fuel-system': "mpfi",
32                     'bore': 3.47,
33                     'stroke': 2.68,
34                     'compression-ratio': 9,
35                     'horsepower': 111,
36                     'peak-rpm': 5000,
37                     'city-mpg': 21,
38                     'highway-mpg': 27,
39                 },
40             ],
41     },
42     "GlobalParameters": {
43     }
44 }
45 body = str.encode(json.dumps(data))
```

31. masukkan kode ke dalam box lalu klik run

```
02809ker4 - Kerenidie Last saved a few seconds ago
① We would love to hear your feedback on the notebooks experience! Please take a few minutes to complete our survey.

42     "GlobalParameters": {
43     }
44 }
45 body = str.encode(json.dumps(data))
46 headers = {'Content-Type': 'application/json', 'Authorization': ('Bearer ' + key)}
47 req = urllib.request.Request(endpoint, body, headers)
48
49 try:
50     response = urllib.request.urlopen(req)
51     result = response.read()
52     json_result = json.loads(result)
53     y = json_result["Results"]["WebServiceOutput0"][0]
54     print(y)
55
56 except urllib.error.HTTPError as error:
57     print("The request failed with status code: " + str(error.code))
58
59     # Print the headers to help debug the error
60     print(error.info())
61     print(json.loads(error.read().decode("utf8", 'ignore'))))

[15] ✖ <1 sec - JSONDecodeError: Expecting value: line 1 column 1 (char 0)

... The request failed with status code: 502
Content-Length: 23
Content-Type: text/html; charset=utf-8
Date: Tue, 28 Sep 2021 06:00:08 GMT
Server: nginx/1.14.0 (Ubuntu)
X-Ms-Request-Id: 07dbb028-acc5-4c91-9c7a-f19b7fdeeeef3
X-Ms-Run-Function-Failed: True
Connection: close

-----
JSONDecodeError                                Traceback (most recent call last)
<ipython-input-15-48cbc9e22dc7> in <module>
    59     # Print the headers to help debug the error
    60     print(error.info())
--> 61     print(json.loads(error.read().decode("utf8", 'ignore')))
```

32. Hasil run yang diperoleh oleh kelompok kami adalah error. Hal ini mungkin terjadi karena expecting value : line 1 column 1 (char 0) error karena kemungkinan encoding karakter yang incompatible. Ini dikarenakan, pada dataset, terdapat beberapa nilai yang memiliki karakter incompatible seperti: ?

Karakter incompatible tersebut menurut kami sangat erat kaitannya dengan tahapan pre-processing yakni dalam tahap preparation atau secara lebih jelasnya adalah menyiapkan sumber data yang akan di-train atau dilatih untuk membangun model ML. Data yang baik akan dapat menghasilkan model yang baik. Oleh karenanya, ini juga bisa berarti dipengaruhi oleh proses normalisasi dalam tahapan preparation yang belum dikerjakan secara maksimal. Banyak faktornya, utamanya adalah kondisi data-set, atau pun human error sendiri dalam membangun model ML yang berkaitan.

Hal ini juga berkaitan dengan proses regresi. Pada proses regresi untuk normalisasi data tidak dipilih dan terdapat pembersihan data yang hilang. Pada proses automated ML tidak terdapat layanan seperti itu. Ada baiknya, untuk automated ML ini menggunakan dataset yang full (tidak ada data yang missing), karena pada automated ML tidak ada pipeline seperti menambahkan step 'clean missing data', sehingga semua column diproses dan tentunya ada column yang memiliki incompatible karakter.