

Dokumen Teknis Capstone Project Bidang MariBisnis

Spesifikasi Laptop yang digunakan: Asus A456U dengan spesifikasi Processor Intel core i5 7200U dengan RAM DDR4 sebesar 12GB, storage harddisk 1TB, dengan sistem operasi Windows 11.

Software yang digunakan: Microsoft Azure digunakan untuk memodelkan dataset MariBisnis menggunakan permodelan linear regression, Microsoft Power BI Service dan Desktop digunakan untuk memvisualisasikan data hasil machine learning dari dataset maribisnis kedalam bentuk chart atau grafik.

Langkah-langkah pengerjaan:

1. Membuat Workspace di Microsoft Azure

- Masuk ke dalam Portal Azure, menggunakan kredensial Microsoft yang dimiliki.
- Membuat resource Machine Learning dengan memilih + Create a Resource pada homepage Azure portal
- Cari Machine Learning kemudian masukkan pengaturan sesuai arahan berikut:
Subscription : Azure for Students
Resource group : capstonehari
Workspace name : capstoneproject
Region : Southeast Asia
Storage account : capstoneproject1028042322
Key vault : capstoneproject4249825250
Application insights : capstoneproject5209469679
Container registry : None
- Selanjutnya klik create

2. Membuat Compute Resources

- Untuk membuat compute resources ada 2 opsi yakni compute instance dan compute clusters
- Compute instances dengan pengaturan sebagai berikut:

Compute name : instance1

Location : Southeast Asia

Virtual Machine type : CPU

Virtual Machine size : Standard_DS11_v2

- Compute clusters dengan pengaturan sebagai berikut:

Location : Southeast Asia

Virtual Machine tier : Low Priority

Virtual Machine type : CPU

Virtual Machine size : Standard_DS11_v2
Compute name : clusters1
Min number of nodes : 0
Max number of nodes : 2
Idle before scale down : 120 seconds
Enable SSH access : Unselected

3. Membuat Dataset MariBisnis

- Pada tab Dataset, pilih + Create dataset dan pilih import from local files dan pilih maribisnis.csv
- Pada Basic info atur dengan pengaturan sebagai berikut:
 Name : MariBisnis
 Dataset type : Tabular
 Description : Maribisnis dataset for the capstone project
- Pada Datastore and file selection:
 Select or create a datastore : workspaceblobstore
 Select files for your dataset : Upload files of maribisnis.csv.
 Skip data validation : Unchecked
- Pada Settings and preview:
 - File format : Delimited
 - Delimiter : Comma
 - Encoding : UTF-8
 - Column header : Only first file has headers
 - Skip rows : None
- Pada Schema, tidak ada perubahan, sehingga dapat langsung memilih tombol Next.
- Confirm details, setelah semuanya aman, maka bisa dilanjutkan untuk Create.

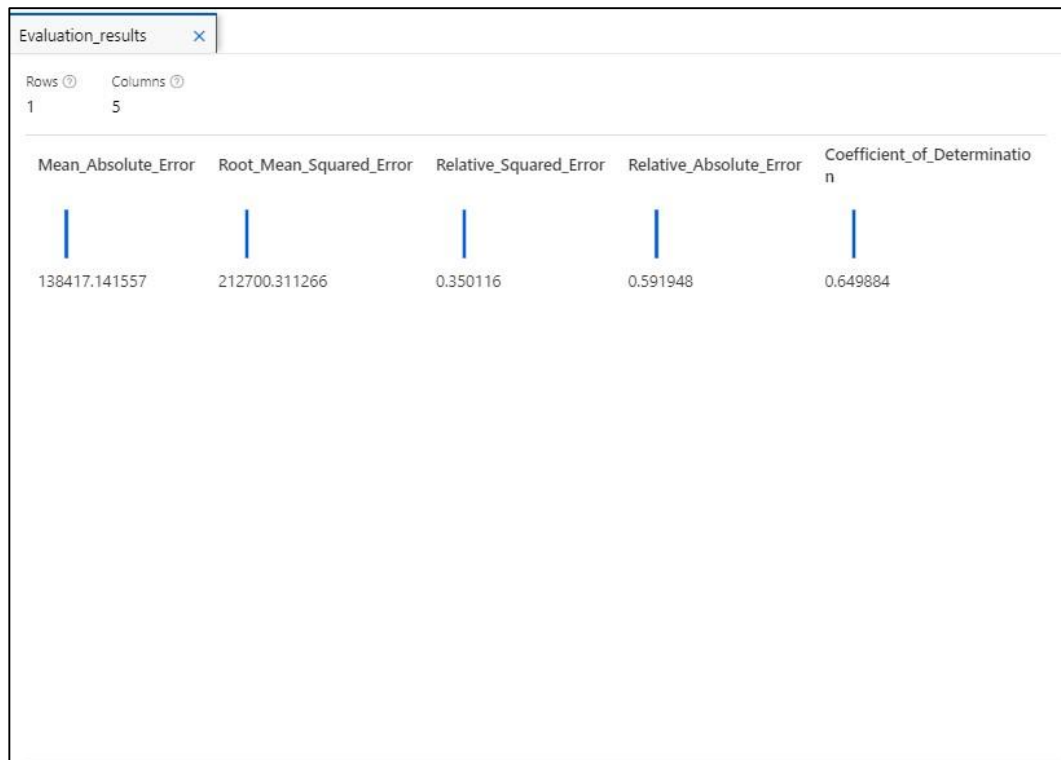
Membuat Pipeline

1. Pada Azure Machine Learning Studio, pada laman Designer, pilih + untuk membuat pipeline baru.
2. Pada sisi kiri atas layar, klik nama pipeline default (Pipeline-Created-on- date) dan ubah namanya ke MariBisnis Price Training.
3. Perlu compute target untuk menjalankan pipeline. Pilih compute cluster yang telah dibuat sebelumnya.
4. Pada panel asset library, pilih **Dataset** > **MariBisnis**, kemudian drag and drop ke dalam kanvas.
5. Selanjutnya pilih **Data Transformation** > **Select Columns in Dataset**, kemudian drag and drop ke dalam kanvas.

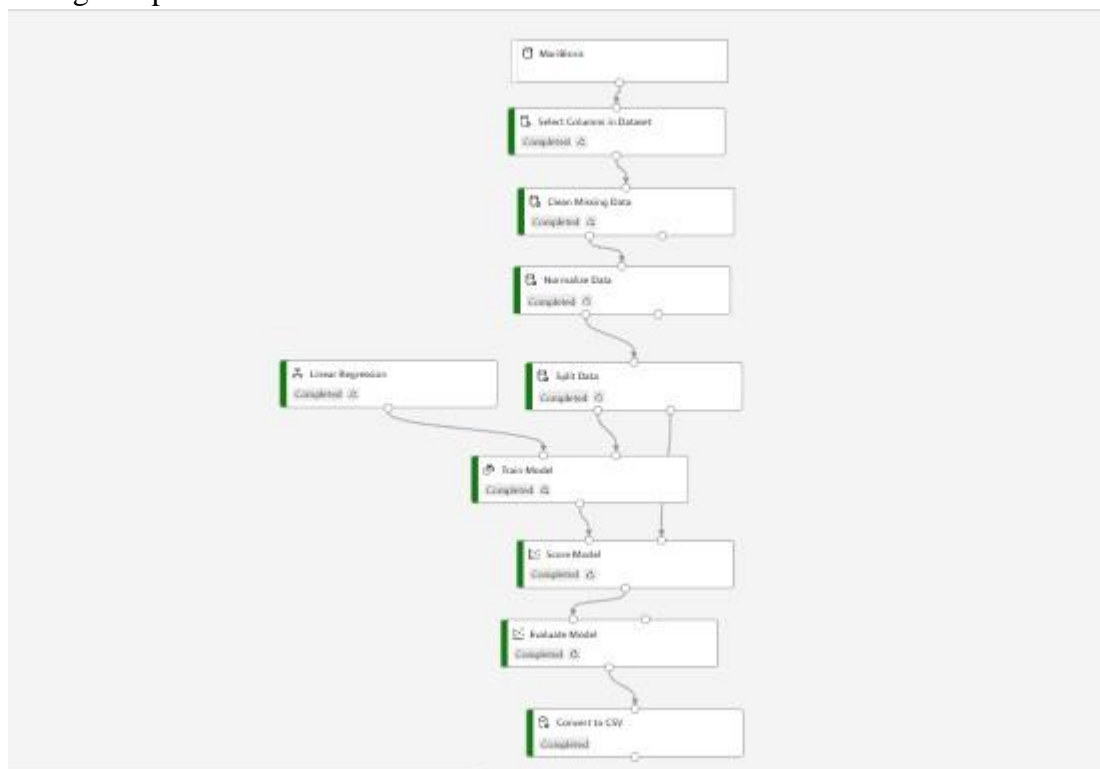
- a. Pada bagian ini, hubungkan dataset MariBisnis dengan nodes yang ada di Select Columns in Dataset.
- b. Pada bagan **Select Columns**, isikan kolom-kolom sebagai berikut:
- Price
 - Bedroom
 - Bathroom
 - sqft_living
 - sqft_lot
 - floors
 - waterfront
 - view
 - grade
 - condition
 - sqft_above
 - sqft_basement
 - yr_built
 - yr_renovated
 - sqft_living15
 - sqft_lot15
6. Selanjutnya pilih Data Transformation > Clean Missing Data, kemudian drag and drop ke dalam kanvas.
- a. Pada bagian ini, hubungkan Select Coloumn in dataset dengan nodes yang ada di Clean Missing Data.
- b. **Columns to be cleaned** , isikan dengan All columns.
- c. **Minimum missing value ratio**: 0.0
- d. **Maximum missing value ratio**: 1.0
- e. **Cleaning mode**: Remove entire row. Bahwa setiap baris yang memiliki nilai kosong atau NaN akan dihapus secara otomatis pada step Clean Missing Data.
7. Selanjutnya pilih **Data Transformation > Normalize Data**, kemudian drag and drop ke dalam kanvas.
- a. Pada bagian ini, hubungkan Clean Missing Data dengan nodes yang ada di Normalize Data.
- b. Pada bagan Columns to transform, isikan kolom-kolom sebagai berikut:
- Bedroom
 - Bathroom
 - sqft_living
 - sqft_lot
 - floors
 - waterfront
 - view
 - grade
 - condition
 - sqft_above
 - sqft_basement

- yr_built
 - yr_renovated
 - sqft_living15
 - sqft_lot15
8. Selanjutnya pilih **Data Transformation > Split Data**, kemudian drag and drop ke dalam kanvas.
 - a. Pada bagian ini, hubungkan Clean Missing Data dengan nodes yang ada di Split Data.
 - b. Pada **Splitting mode**, pilih Split Rows, karena akan membagi barisnya menjadi training dan testing dataset.
 - c. Pada **Fraction**, masukkan 0.7. Hal ini diartikan sebagai 70% dari dataset MariBisnis akan dialokasikan sebagai training data, dan sisanya, sebanyak 30% akan dialokasikan sebagai testing data.
 - d. Pada **Randomized split**, pilih True. Hal ini diartikan bahwa pemilihan dataset training maupun testing semuanya dilakukan secara teracak.
 - e. Pada **Random seed**, masukkan 123 sebagai seed acaknya.
 - f. Pada **Stratified split**, isikan False. Hal ini guna menghindari pemisahan baris secara berkelompok menggunakan kolom bertingkat.
 9. Selanjutnya pilih **Machine Learning Algorithms > Linear Regression**, kemudian drag and drop ke dalam kanvas.
 - a. Pada bagian ini, hubungkan Linear Regression dengan nodes kiri yang ada di Train Model.
 10. Selanjutnya pilih **Model Scoring and Evaluation > Score Model**, kemudian drag and drop ke dalam kanvas.
 - a. Resampling method : Bagging Resamplinga. Pada bagian ini, hubungkan Split Data dengan nodes kanan yang ada di Score.
 - b. Pada Append score, pilih True.
 11. Selanjutnya pilih **Model Scoring and Evaluation > Evaluate Model**, kemudian drag and drop ke dalam kanvas.
 - a. Pada bagian ini, hubungkan Score Model dengan nodes kiri yang ada di Evaluate Model.
 - b. Pada Evaluate Model, tidak ada parameter yang perlu dimasukkan.
 12. Setelah semua modul sudah berada dan terhubung di dalam kanvas, maka dapat dilakukan **Submit** dan menunggu pipeline selesai running.

13. Setelah pipeline telah usai dibuat, pada **Evaluate Model** akan nampak lima matriks evaluasi sebagai berikut:



14. Selanjutnya pilih **Data Transformation > Convert to CSV**, kemudian drag and drop ke dalam kanvas. Convert to CSV berguna untuk mengkonversi data kedalam format csv agar dapat didownload untuk divisualisasikan ke Power BI.



15. Setelah training pipeline selesai, klik **submit** kemudian klik **deploy** pilih real-time inference

