**DIGITALTALENTSCHOLARSHIP–KOMINFO(BPPTIKCIKARANG) VOCATIONAL SCHOOL GRADUATE ACADEMY**

**ASSOCIATEDATASCIENCE**

**Gelombang- 12**

**Pengajar:Mona Arif Muda Batubara**

**Tugas Akhir/ Project ADS**

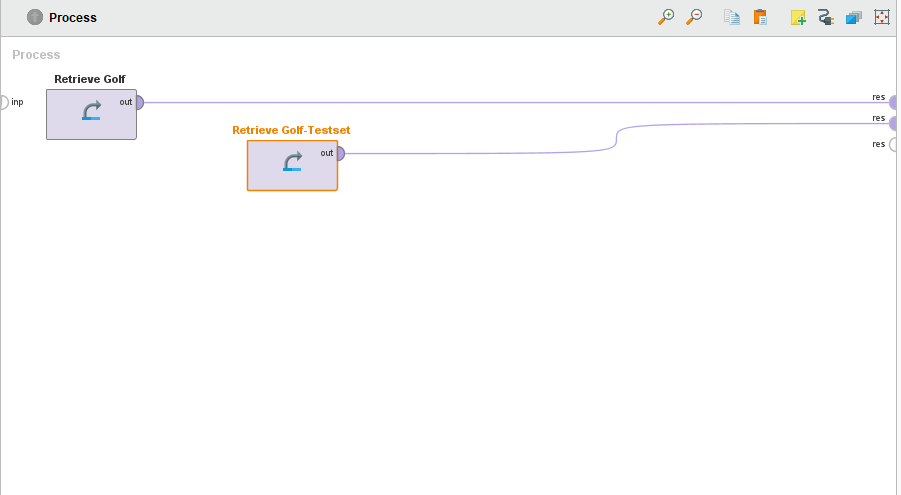
**Nama Peserta : Hendry Imam Sanjaya**

**Soal:**

1. Dengan menggunakan dataset latih Golf dan dataset tes Golf-Testset yang ada di data samples di RapidMiner, lakukan:
   1. telaah data tersebut dengan menggunakan RapidMiner
   2. validasi, pembersihan, dan/atau konstruksi data bila diperlukan, dengan RapidMiner
   3. bangun Model Classification dengan algoritma K-NN dan evaluasi hasil pemodelannya
   4. buat laporan tertulisnya
2. Dengan menggunakan dataset Facebook-Live-Sellers-in-Thailand\_20210128.csv terlampir, lakukan:
   1. telaah data tersebut dengan menggunakan RapidMiner
   2. validasi, pembersihan, dan/atau konstruksi data bila diperlukan, dengan RapidMiner
   3. bangun Model Clustering dengan algoritma K-Means dan evaluasi hasil pemodelannya
   4. buat laporan tertulisnya.
3. Dengan menggunakan dataset ToyotaCorolla.csv terlampir, lakukan:
   1. telaah data tersebut dengan menggunakan RapidMiner
   2. validasi, pembersihan, dan/atau konstruksi data bila diperlukan, dengan RapidMiner
   3. Bangun Model Regresi Linier dengan 3 variable bebas dan evaluasi hasil pemodelannya
   4. buat laporan tertulisnya.
4. **LAPORAN ANALISIS GOLF dan GOLF –TEST (CLASSIFICATION)**

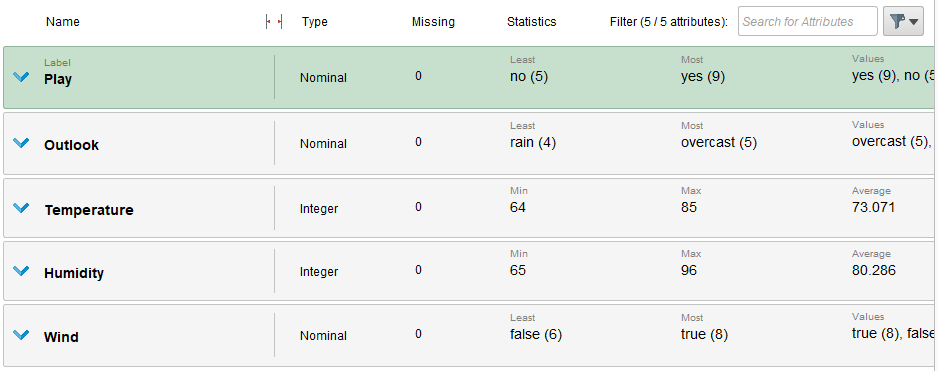
Dataset ini berisi data mengenai kondisi cuaca dan keputusan untuk bermain golf. Tujuan dari analisis ini adalah untuk membangun model klasifikasi menggunakan algoritma K-NN dan mengevaluasi performanya.

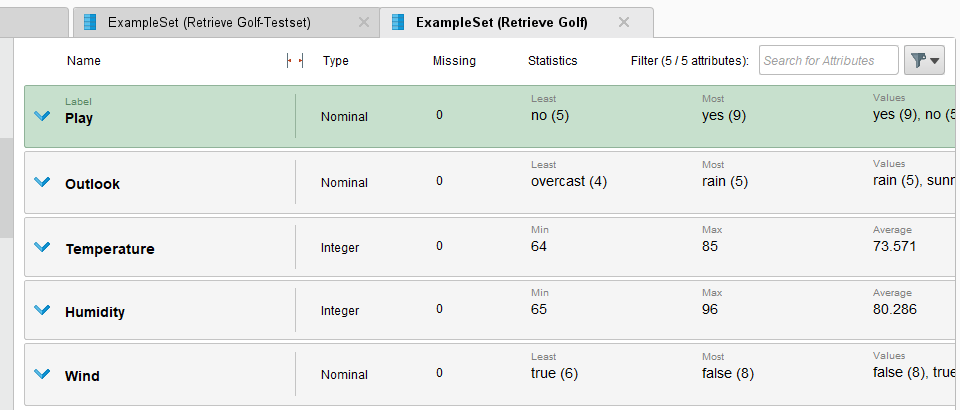
* 1. Ini merupaka telaahan data menggunakan RapidMiner yang mana kita
* Mengimpor dataset Golf dan dataset tes Golf-Testset dari data samples di RapidMiner.
* Pilih “Samples” pada repository, dan cari dataset yang dimaksud.
* Drag dan drop dataset ke dalam proses baru lalu run
* Gambar di bawah ini merupakan hasil data yang di dapat



|  |  |
| --- | --- |
| **Golf** | **Golf-Testset** |
|  |  |

* 1. validasi, pembersihan, dan/atau konstruksi data bila diperlukan, dengan RapidMiner. Setelah kita lihat datanya di missing values dan outliers tidak ada yang harus di bersihkan.

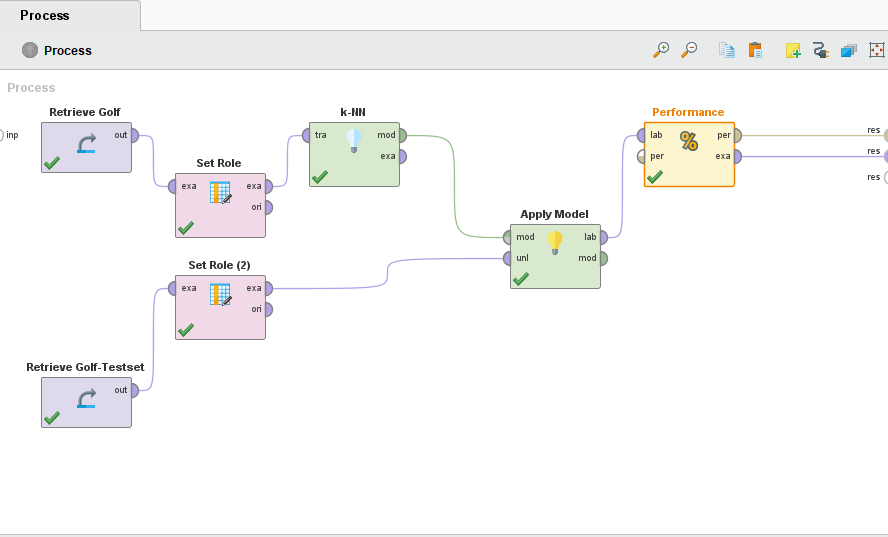


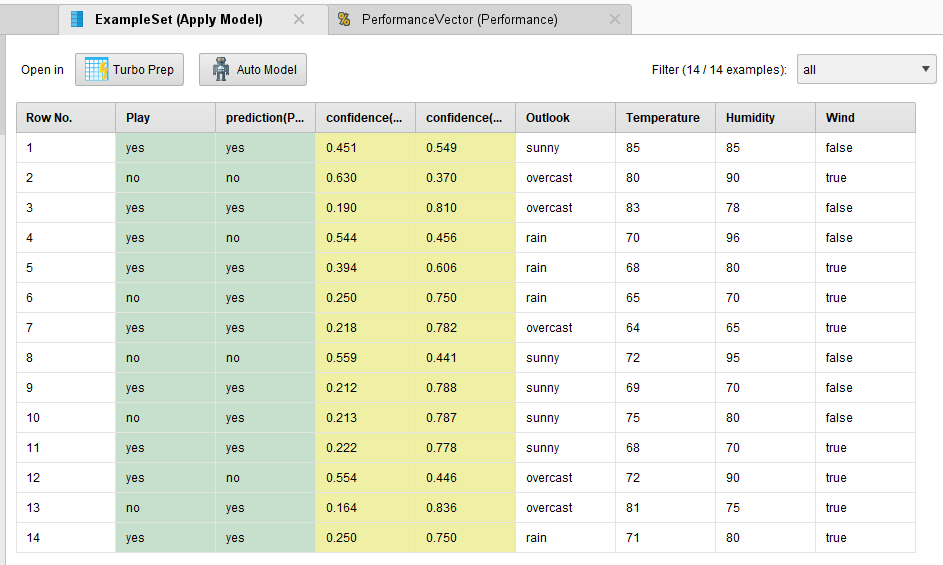


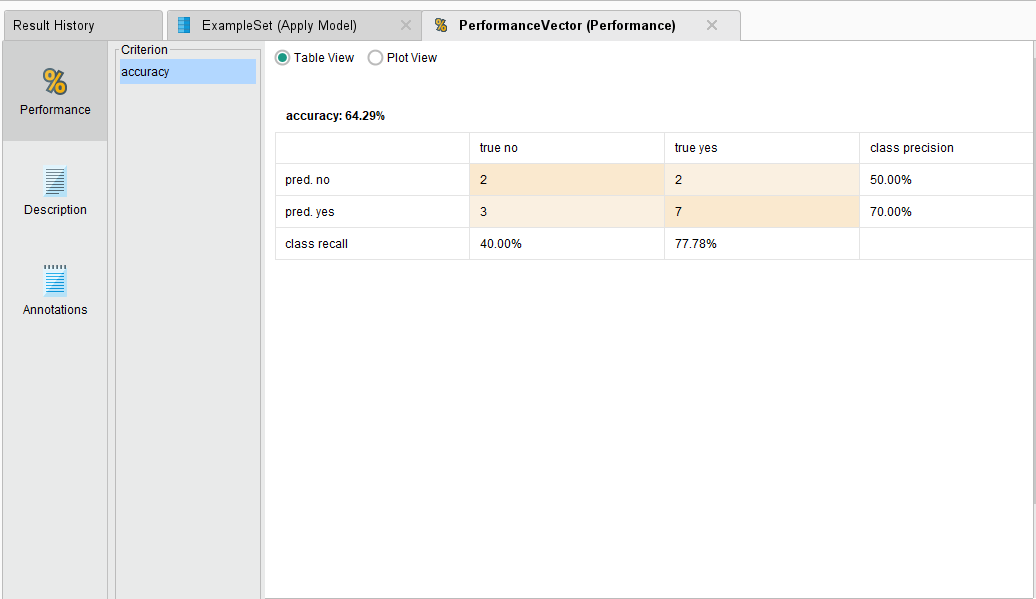
* 1. bangun Model Classification dengan algoritma K-NN dan evaluasi hasil pemodelannya
* Setup Proses dan Evaluasi Hasil:
  + Hubungkan Data set ke Set Role untuk menetapkan atribut target (disini saya memberikan lable pada artibut Play).
  + Tambahkan “K-NN” untuk membangun model klasifikasi dengan parameter K yang sesuai (misalnya, K=3).
  + Sambungkan “Apply Model” untuk menerapkan model pada dataset tes.
  + Tambahkan “Performance (Classification)” untuk mengevaluasi model.
  + Lalu Run.

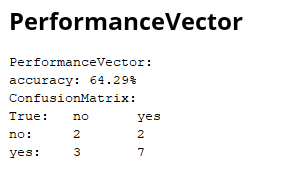
Algoritma K-NN dengan parameter K=3 digunakan untuk membangun model. Berikut adalah struktur proses yang digunakan:

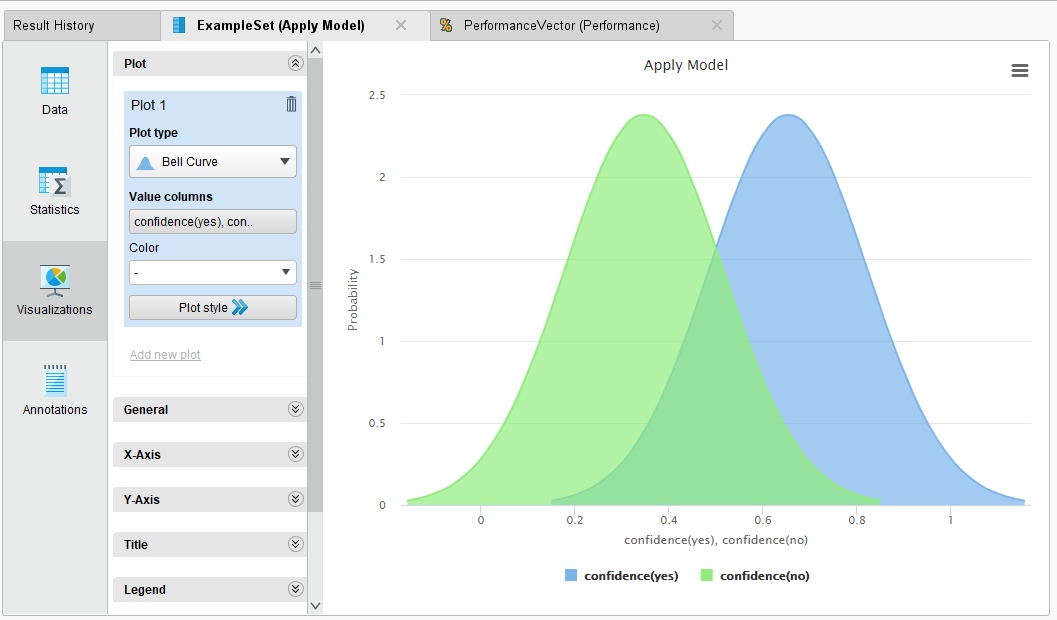
* Data Set-> Set Role -> K-NN -> Apply Model -> Performance (Classification)









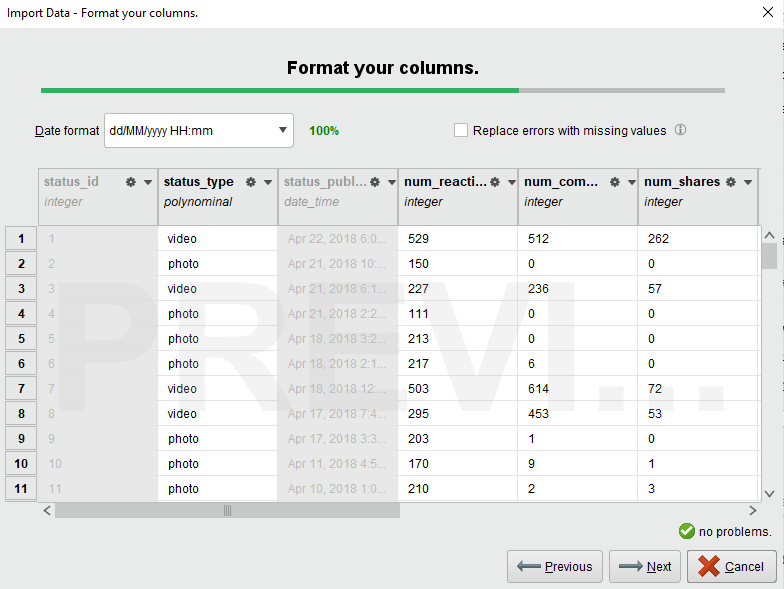


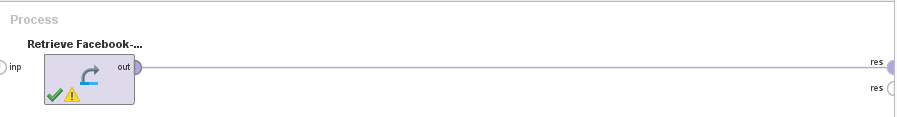
Model K-NN menunjukkan performa yang cukup baik dengan akurasi sebesar 64.29%.

1. **LAPORAN ANALISIS FACEBOOK LIVE SELLERS IN THAILAND (CULUSTERING)**

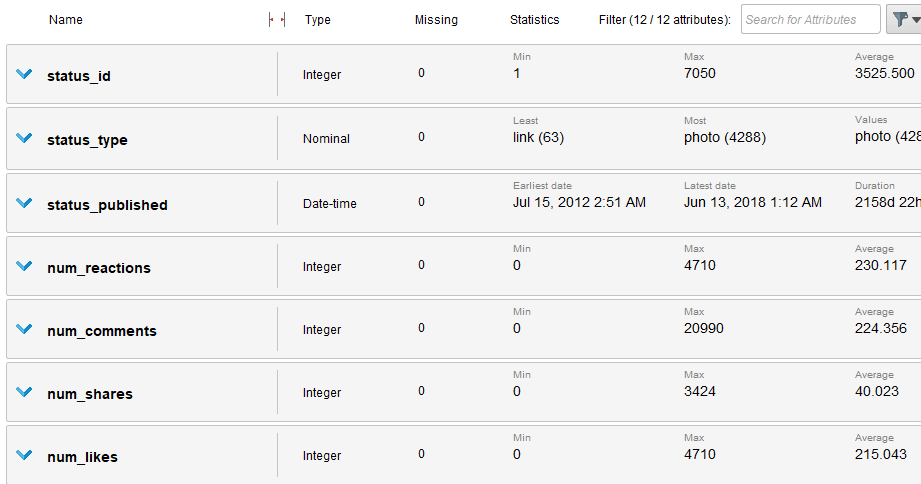
Dataset ini berisi data mengenai penjual live di Facebook di Thailand. Tujuan dari analisis ini adalah untuk membangun model clustering menggunakan algoritma K-Means dan mengevaluasi performanya.

* 1. telaah data tersebut dengan menggunakan RapidMiner
* Buka RapidMiner Studio dan buat proyek baru. Impor dataset Facebook-Live-Sellers-in-Thailand\_20210128.csv ke dalam RapidMiner.
* Pilih “Import Data” dan arahkan ke file Facebook-Live-Sellers-in-Thailand\_20210128.csv.
* Drag dan drop dataset ke dalam proses baru.
* Lihat distribusi data untuk memahami jenis atribut (categorical atau numerical). Disini saya mengexclud data yag menurut saya tidak di gunakan
* Gunakan operator “Statistics” untuk melihat statistik dasar dari setiap atribut.
* Visualisasikan data dengan “Charts” untuk memahami hubungan antara variabel.

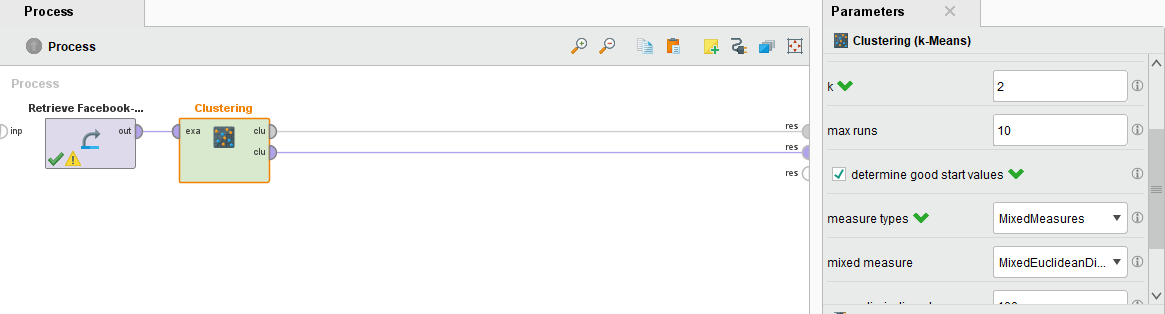


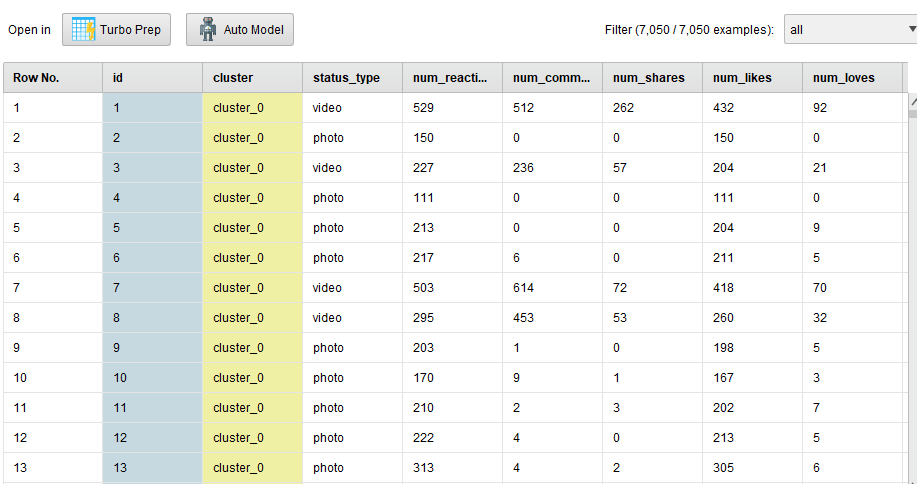


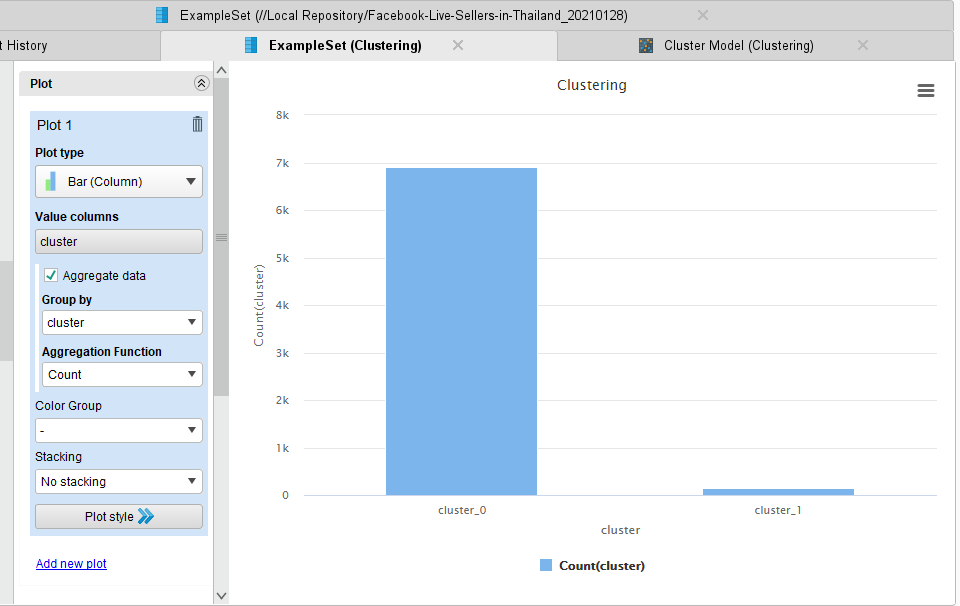
* 1. validasi, pembersihan, dan/atau konstruksi data bila diperlukan, dengan RapidMiner. Setelah kita lihat datanya di missing values dan outliers tidak ada yang harus di bersihkan.

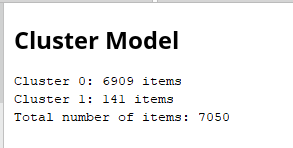


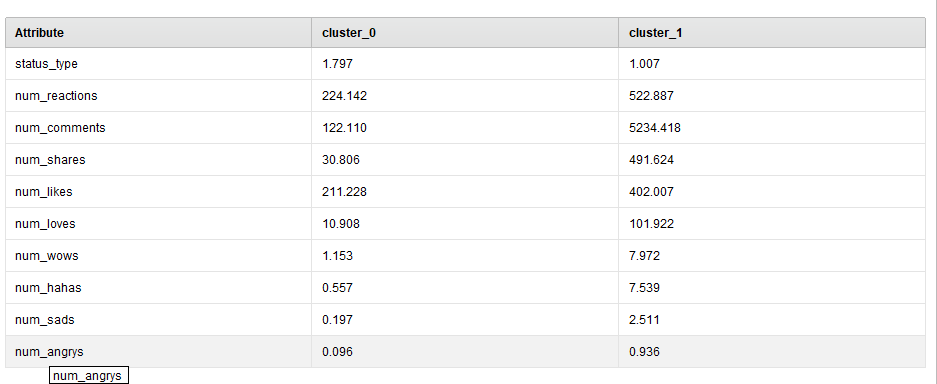
* 1. Model Clustering dengan algoritma K-Means dan evaluasi hasil pemodelannya.
* **Setup Proses**:
  + Drag operator “Read CSV” untuk membaca dataset.
  + Hubungkan “Read CSV” ke “Set Role” untuk menetapkan atribut yang akan digunakan dalam clustering (pastikan atribut non-numerik diabaikan atau diubah ke bentuk numerik jika perlu).
  + Tambahkan “K-Means” untuk membangun model clustering dengan jumlah cluster yang sesuai (misalnya, k=2).
  + Sambungkan “Apply Model” untuk menerapkan model pada dataset.
  + Tambahkan “Performance (Clustering)” untuk mengevaluasi model.
  + Lihat metrik seperti Silhouette Coefficient, Within-Cluster Sum of Squares (WCSS), dan Elbow Method untuk menentukan jumlah cluster yang optimal.

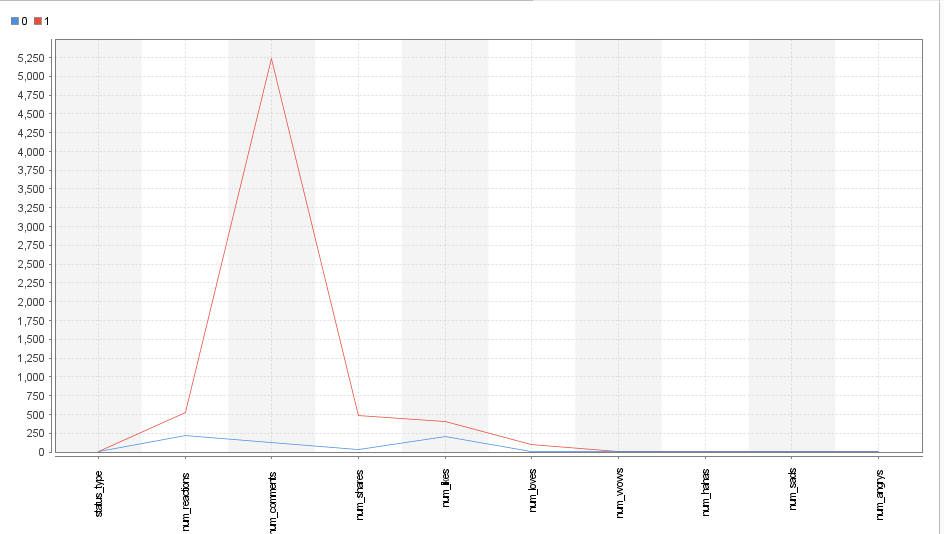




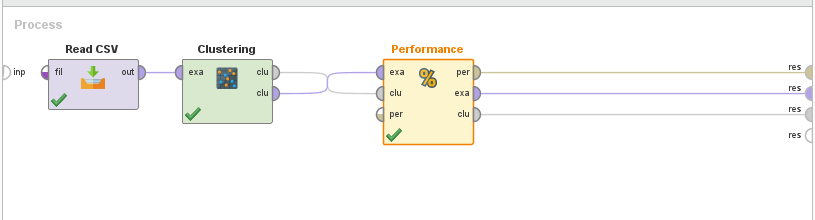




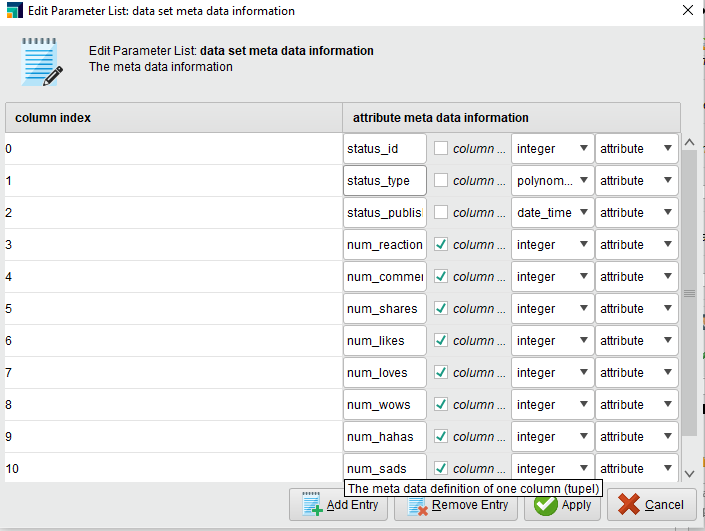




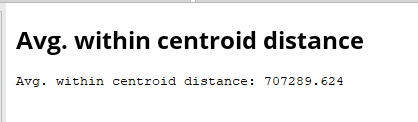
Penambahan Performance



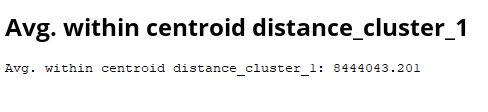
Karena ada data polynominal kita hapus check list status\_type



Ini merupakan hasil performace vector yang mana kita dapat mengujinya di excel atau aplikasi lainnya





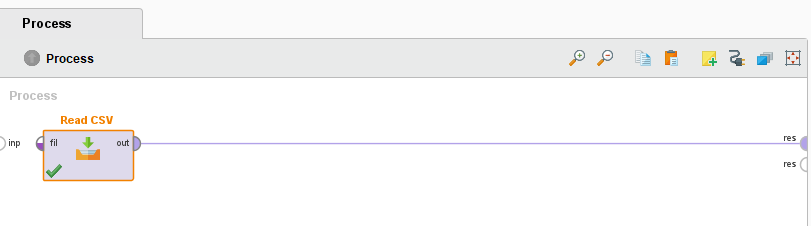


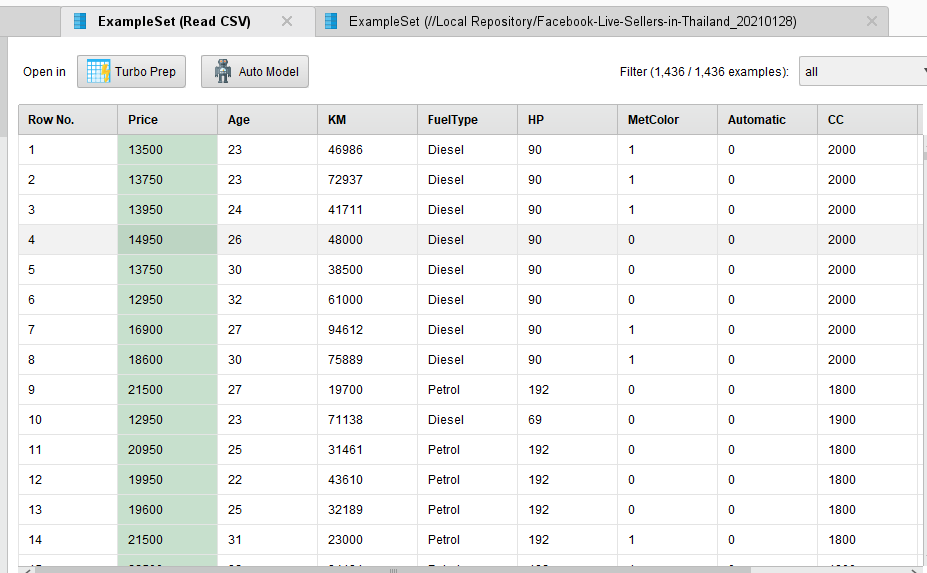
1. **LAPORAN ANALISIS TOYOTA COROLLA (REGRESI LINIER)**

Dataset ini berisi data mengenai berbagai atribut mobil Toyota Corolla, termasuk usia, jarak tempuh, tenaga kuda, dan harga. Tujuan dari analisis ini adalah untuk membangun model regresi linier untuk memprediksi harga mobil berdasarkan tiga variabel bebas.

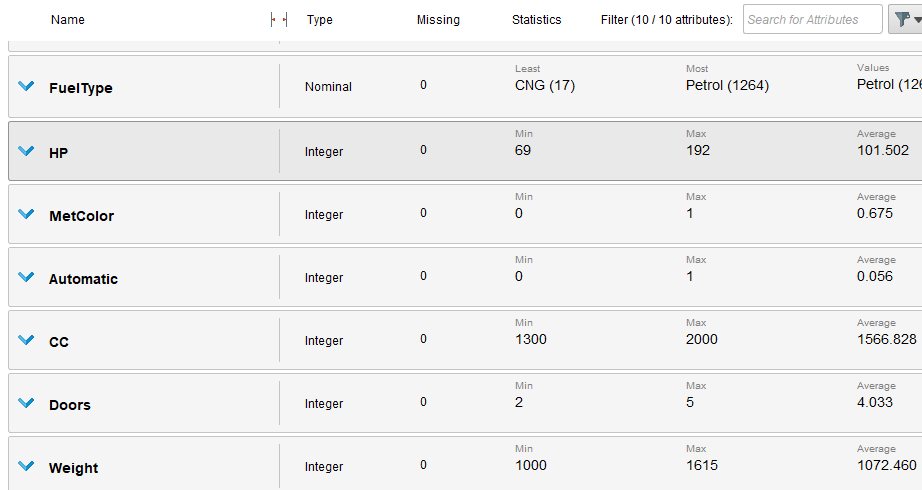
1. telaah data tersebut dengan menggunakan RapidMiner

* **Buka RapidMiner Studio** dan buat proyek baru.
* **Impor dataset ToyotaCorolla.csv** ke dalam RapidMiner.
* Pilih “Import Data” dan arahkan ke file ToyotaCorolla.csv.
* Drag dan drop dataset ke dalam proses baru.
* **Buka dataset ToyotaCorolla** di RapidMiner.
* Lihat distribusi data untuk memahami jenis atribut (categorical atau numerical).
* Gunakan operator “Statistics” untuk melihat statistik dasar dari setiap atribut.
* Visualisasikan data dengan “Charts” untuk memahami hubungan antara variabel.



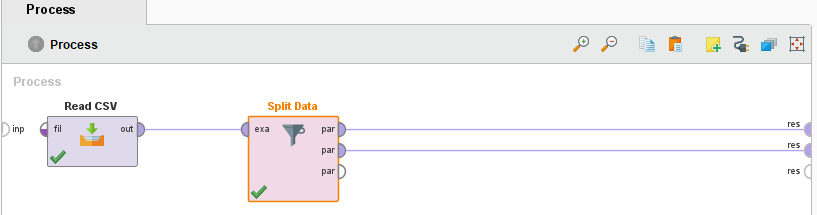


1. validasi, pembersihan, dan/atau konstruksi data bila diperlukan, dengan RapidMiner. Setelah kita lihat datanya di missing values dan outliers tidak ada yang harus di bersihkan.

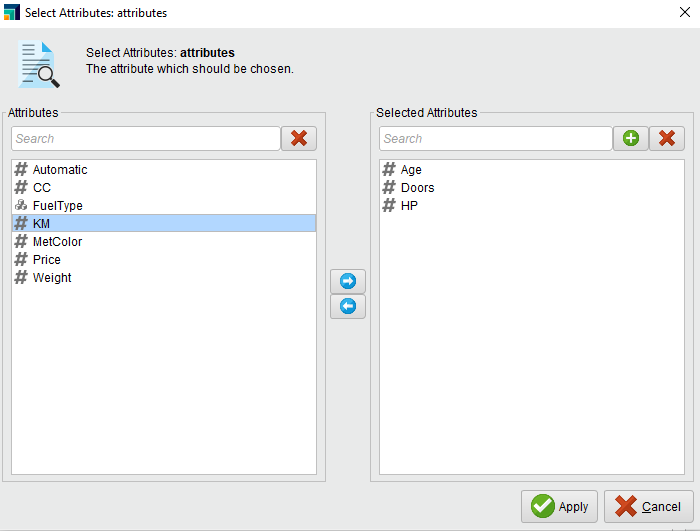


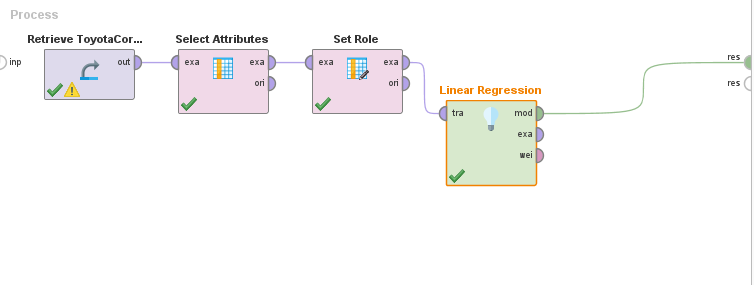
1. telaah data tersebut dengan menggunakan RapidMiner dan Model Regresi Linier

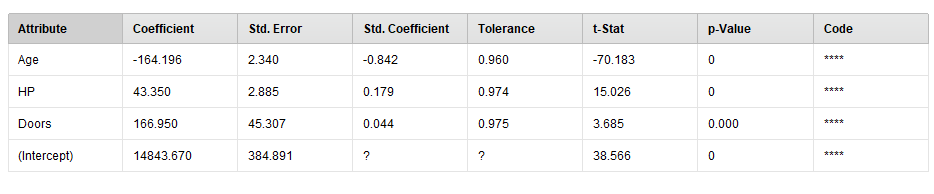
* Setup Proses:
  + Drag operator “Toyota Corolla” untuk membaca dataset.
  + Hubungkan “Toyota Corolla” ke “Set Role” untuk menetapkan atribut target (misalnya, Price).
  + Pilih tiga variabel bebas yang akan digunakan dalam model (misalnya, Age, KM, HP).
  + Tambahkan “Linear Regression” untuk membangun model regresi linier.
  + Tambahkan “Apply Model” untuk menerapkan model pada dataset.
  + Tambahkan “Performance (Regression)” untuk mengevaluasi model.

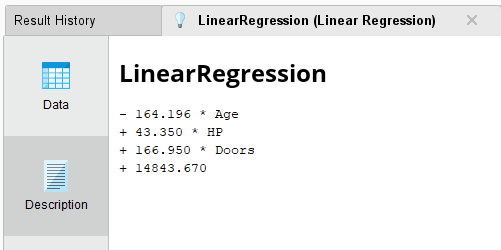


|  |  |
| --- | --- |
| Data Traning | Data Testing |
|  |  |









Hasil setelah di performance

