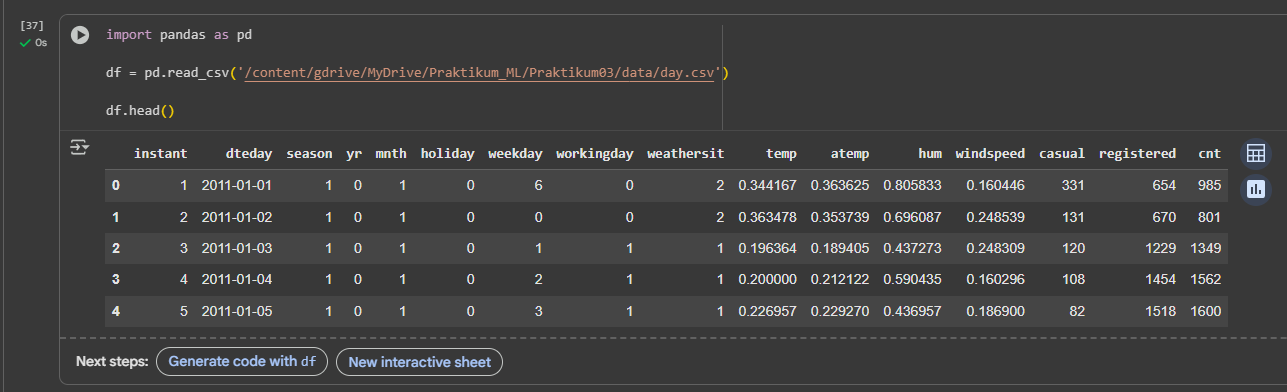
**Tugas 3: Praktikum Mandiri 3 Machine Learning**

**Aria Kristallinacht Sundanis - 0110222076**

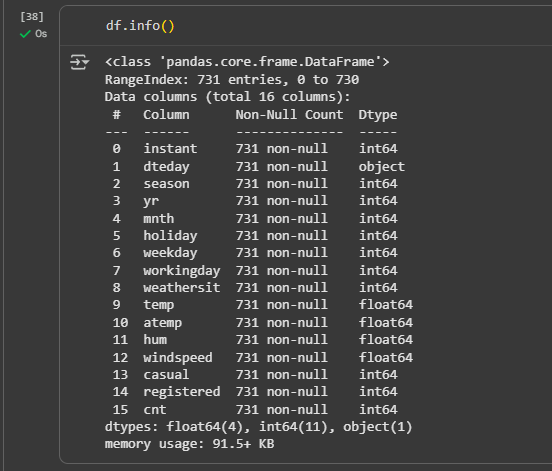
Teknik Informatika, STT Terpadu Nurul Fikri, Depok

E-mail: [aria22076ti@student.nurulfikri.ac.id](mailto:aria22076ti@student.nurulfikri.ac.id)

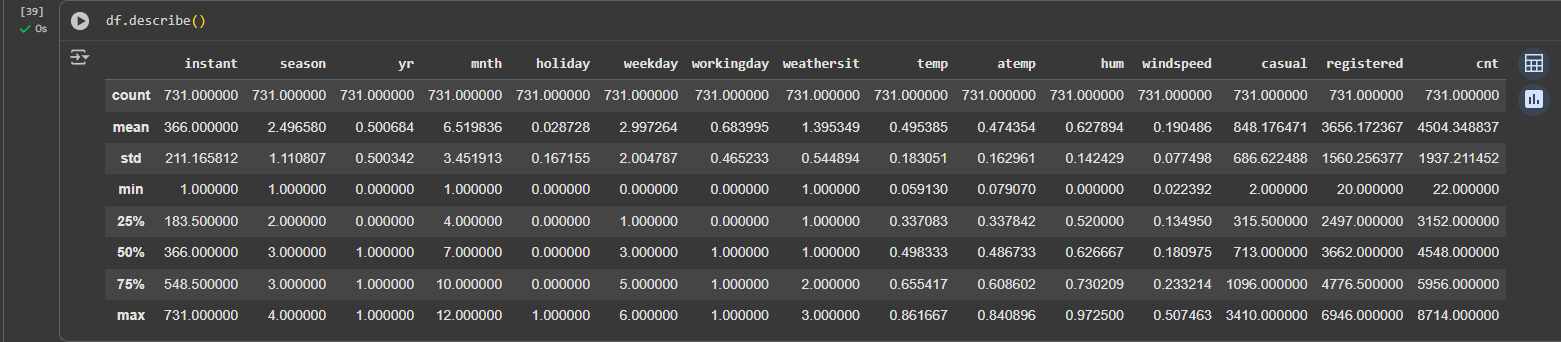
**Praktikum Mandiri 3**



Pada tahap ini menggunakan library pandas untuk membaca dataset. Lokasi day.csv yang sudah disimpan di Google Drive dipanggil melalui sebuah variabel path. Setelah itu, file tersebut dibaca dengan fungsi **pd.read\_csv()** dan hasilnya disimpan ke dalam sebuah DataFrame dengan nama **df**, untuk memastikan data berhasil dimuat, digunakan perintah df.head() yang menampilkan lima baris pertama dari dataset. Secara keseluruhan, dataset ini berisi 731 baris dan 16 kolom yang merepresentasikan data penyewaan sepeda.

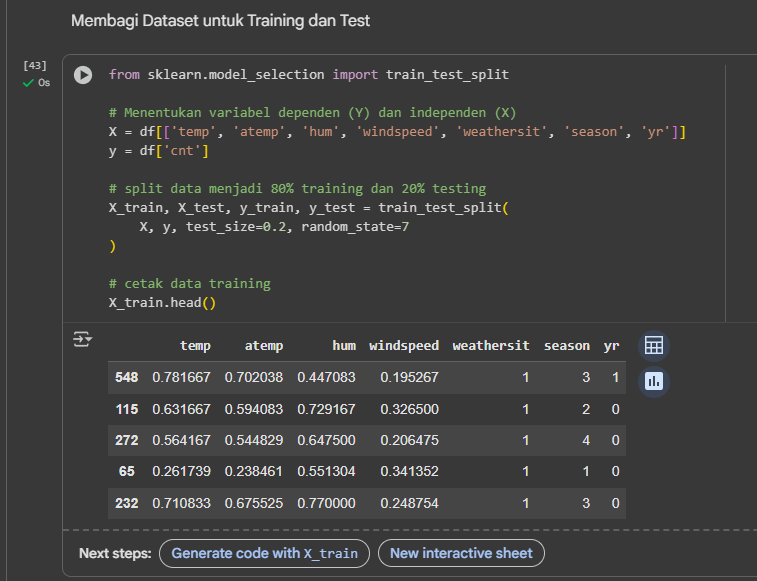


Perintah df.info() digunakan untuk menampilkan informasi umum mengenai struktur dataset. Dari hasil tersebut terlihat bahwa dataset memiliki 731 baris dan 16 kolom, dengan semua kolom berisi data lengkap, tidak ada missing value. Tipe data yang digunakan terdiri dari int64, float64, dan object (untuk kolom tanggal).

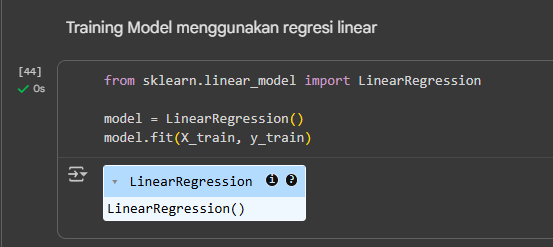


Perintah df.describe() digunakan untuk menampilkan statistik deskriptif dari setiap kolom numerik dalam dataset. Dari hasilnya terlihat bahwa:

* Dataset memiliki 731 data pada tiap kolom.
* Nilai rata-rata penyewaan sepeda (cnt) adalah sekitar 4504 dengan nilai maksimum 8714.
* Kolom temp, atemp, hum, dan windspeed menunjukkan nilai rata-rata kondisi cuaca harian.
* Statistik seperti mean, min, max, dan quartile (25%, 50%, 75%) membantu memahami sebaran serta rentang nilai dari masing-masing variabel sebelum dilakukan analisis lebih lanjut.



Pada tahap ini digunakan fungsi train\_test\_split dari library sklearn.model\_selection untuk membagi dataset menjadi dua bagian, yaitu data training dan data testing. Variabel X diisi dengan beberapa kolom fitur seperti temp, atemp, hum, windspeed, weathersit, season, dan yr, sedangkan variabel y berisi kolom cnt yang merupakan jumlah penyewaan sepeda dan menjadi target yang ingin diprediksi. Proses pembagian data dilakukan dengan proporsi 80% untuk data training dan 20% untuk data testing menggunakan random\_state=7 agar hasil pembagian tetap konsisten. Setelah itu, ditampilkan beberapa baris awal dari data training menggunakan X\_train.head() untuk memastikan data sudah terbagi dengan benar.



Selanjutnya melatih model dengan menggunakan regresi linear. Untuk menerapkan regresi linear kita perlu menggunakan LinearRegression milik scikit-learn. Lalu mengiput data train milik variabel dependen dan independen.



Pada tahap ini dilakukan pembuatan dan evaluasi model regresi linear menggunakan scikit-learn. Model dilatih dengan data training untuk mempelajari hubungan antara variabel seperti suhu, kelembapan, dan kecepatan angin terhadap jumlah penyewaan sepeda (cnt), lalu diuji menggunakan data testing.

Hasil evaluasi menunjukkan nilai R² sebesar 0.6822, artinya model mampu menjelaskan sekitar 68% variasi data. Nilai MAE (795.64) dan RMSE (1072.87) menggambarkan tingkat kesalahan prediksi model. Koefisien menunjukkan pengaruh tiap variabel, sedangkan intercept (1831.74) merupakan nilai dasar prediksi saat semua variabel bernilai nol.

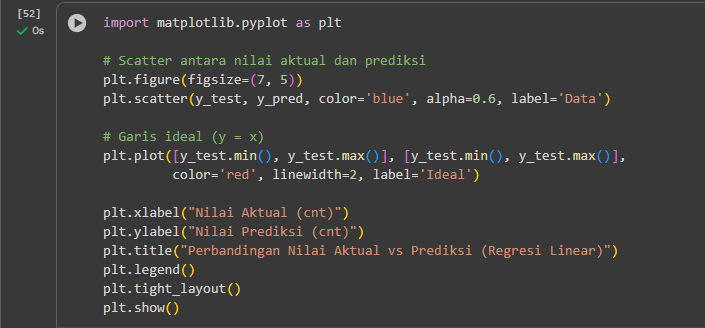
Secara keseluruhan, model regresi linear ini sudah cukup baik dalam memprediksi jumlah penyewaan sepeda, meskipun masih bisa ditingkatkan lagi untuk mengurangi error.

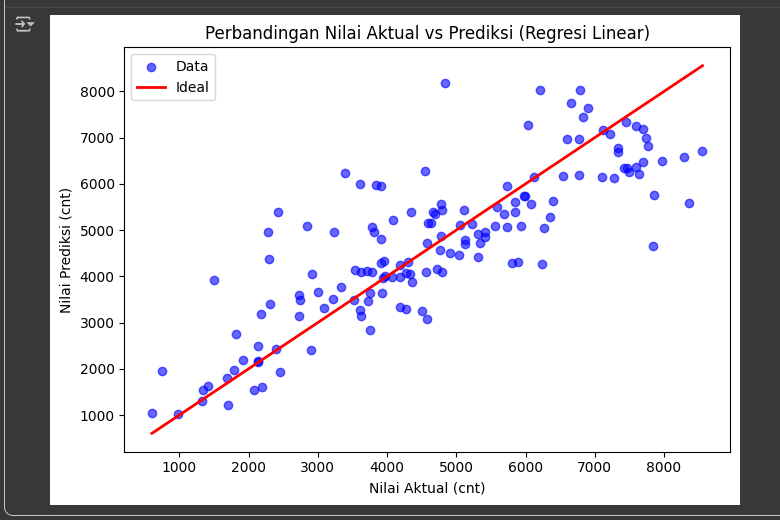


Pada tahap ini dilakukan pembuatan persamaan regresi linear berganda berdasarkan hasil pelatihan model. Pertama, nilai intercept dan koefisien (slope) diambil dari model menggunakan model.intercept\_ dan model.coef\_. Nilai-nilai ini kemudian digunakan untuk membentuk persamaan regresi yang menggambarkan hubungan antara variabel independen (fitur) dengan variabel dependen yaitu jumlah penyewaan sepeda (cnt).

Dan hasil persamaan menunjukkan bahwa setiap variabel memiliki pengaruh berbeda terhadap jumlah penyewaan sepeda. Nilai koefisien positif (seperti temp, atemp, season, dan yr) menandakan bahwa peningkatan nilai variabel tersebut akan meningkatkan jumlah penyewaan. Sebaliknya, koefisien bernilai negatif (seperti hum, windspeed, dan weathersit) menunjukkan bahwa semakin tinggi nilai variabel tersebut, maka jumlah penyewaan cenderung menurun.

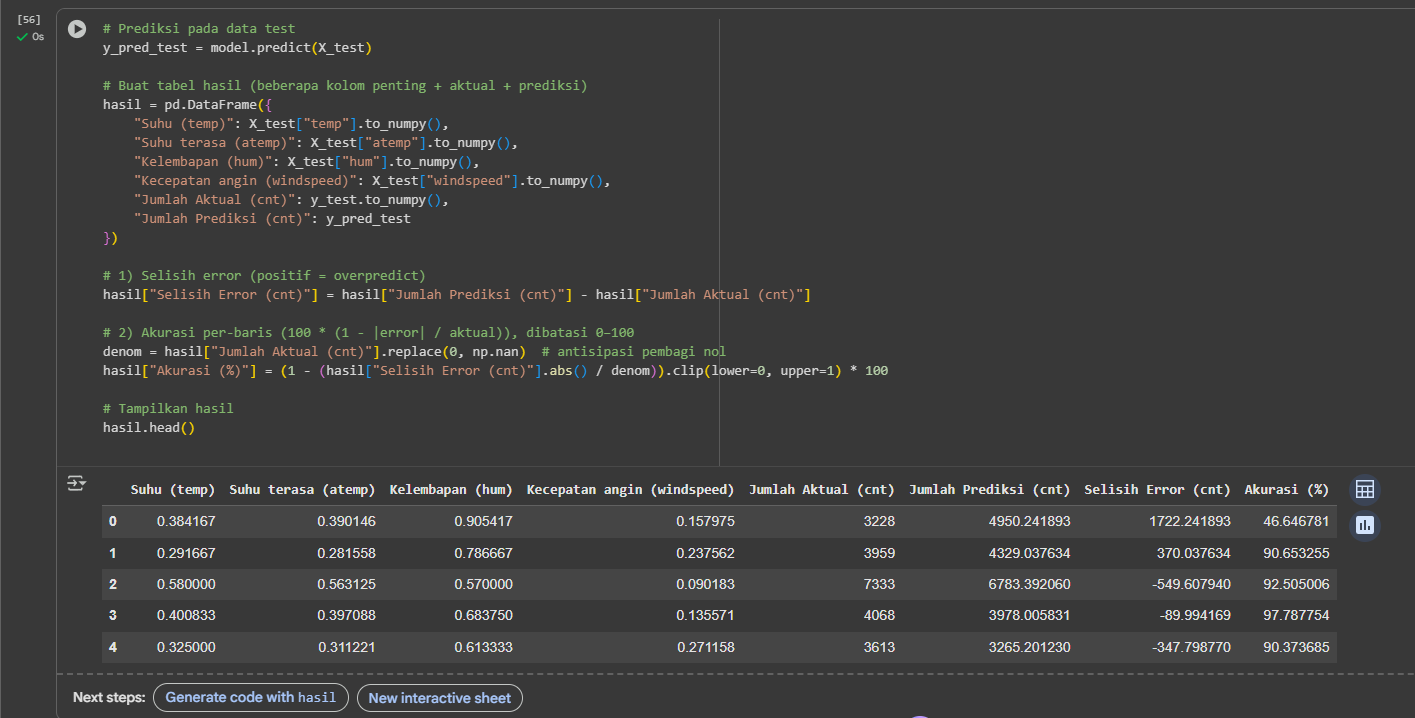
Secara keseluruhan, persamaan ini menjadi representasi matematis dari model regresi linear berganda yang telah dibangun, dan bisa digunakan untuk memperkirakan jumlah penyewaan sepeda berdasarkan kondisi cuaca dan faktor lainnya.





Kode di atas digunakan untuk membuat visualisasi perbandingan antara nilai aktual dan nilai prediksi dari model regresi linear. Fungsi plt.scatter() menampilkan titik-titik biru yang menggambarkan seberapa jauh hasil prediksi dari nilai sebenarnya, sedangkan plt.plot() menampilkan garis merah sebagai garis ideal (y = x), yaitu kondisi ketika prediksi sama dengan nilai aktual.

Dari grafik terlihat bahwa sebagian besar titik biru berada cukup dekat dengan garis merah, menandakan bahwa model memiliki akurasi yang cukup baik dalam memprediksi jumlah penyewaan sepeda. Meski masih ada beberapa penyimpangan, pola datanya sudah menunjukkan hubungan yang cukup konsisten antara nilai aktual dan hasil prediksi.



Kode di atas digunakan untuk mengevaluasi hasil prediksi model regresi linear secara lebih detail dengan membandingkan nilai aktual dan nilai prediksi pada setiap baris data. Pertama, model melakukan prediksi terhadap data uji (X\_test) menggunakan model.predict(), lalu hasilnya dimasukkan ke dalam sebuah tabel DataFrame bersama dengan beberapa kolom penting seperti suhu, kelembapan, kecepatan angin, serta jumlah penyewaan aktual dan prediksi.Selanjutnya, dibuat kolom “Selisih Error (cnt)” yang menunjukkan selisih antara hasil prediksi dan nilai aktual. Jika nilainya positif, berarti model memprediksi lebih tinggi dari nilai sebenarnya (overpredict), sedangkan nilai negatif berarti sebaliknya (underpredict).

Kemudian dihitung juga akurasi per baris menggunakan rumus 100 × (1 - |error| / aktual), yang menggambarkan seberapa dekat hasil prediksi terhadap nilai sebenarnya. Nilai akurasi dibatasi antara 0 hingga 100 agar lebih mudah dibaca.

Dari tabel hasil, terlihat bahwa tiap baris menunjukkan nilai aktual, hasil prediksi, selisih error, dan akurasi dalam persen. Misalnya, pada baris pertama model memiliki akurasi sekitar 46.6%, sedangkan pada baris lain akurasinya bisa mencapai lebih dari 90%, menandakan bahwa sebagian besar prediksi model cukup mendekati nilai sebenarnya, meskipun masih ada beberapa data dengan error yang cukup besar.

Link GitHub :

<https://github.com/Ariakristallnacht/TI03_AriaKS_0110222076/tree/main/Praktikum03>