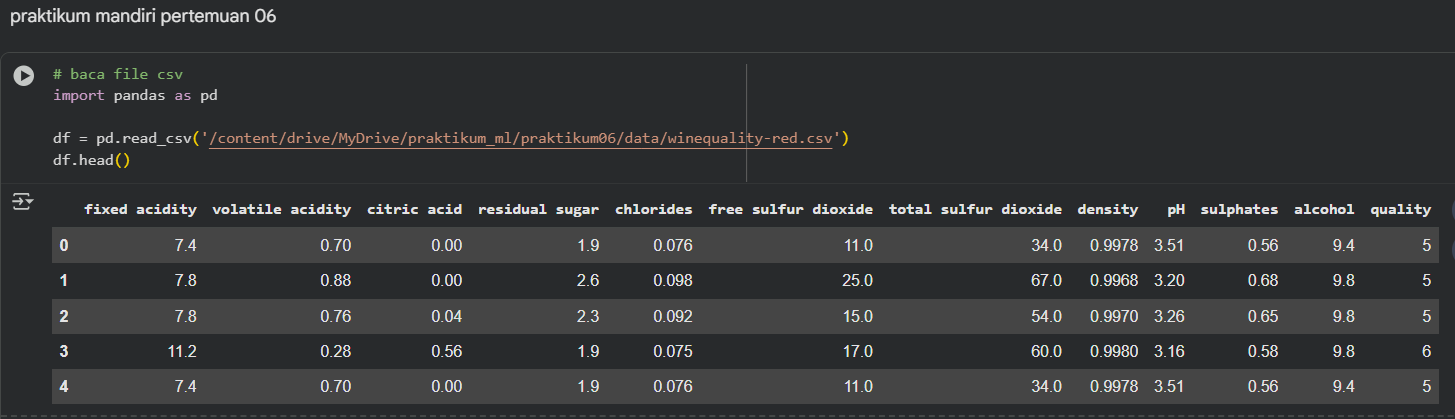
Tugas 6: Tugas Praktimum 6 Machine Learning

**Jamilatun Khoerunnisa - 010222254**

Teknik Informatika, STT Terpadu Nurul Fikri, Depok

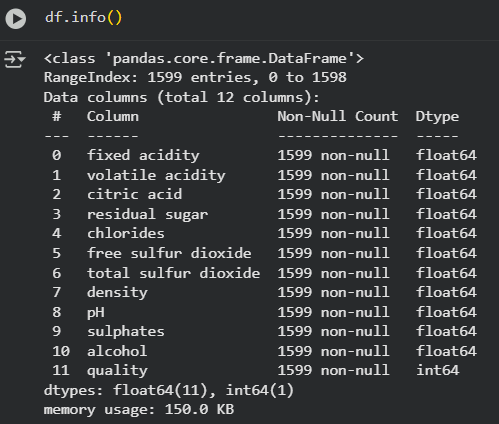
\*E-mail: [Jami22254ti@student.nurulfikri.ac.id](mailto:Jami22254ti@student.nurulfikri.ac.id)

1. Langkah 1



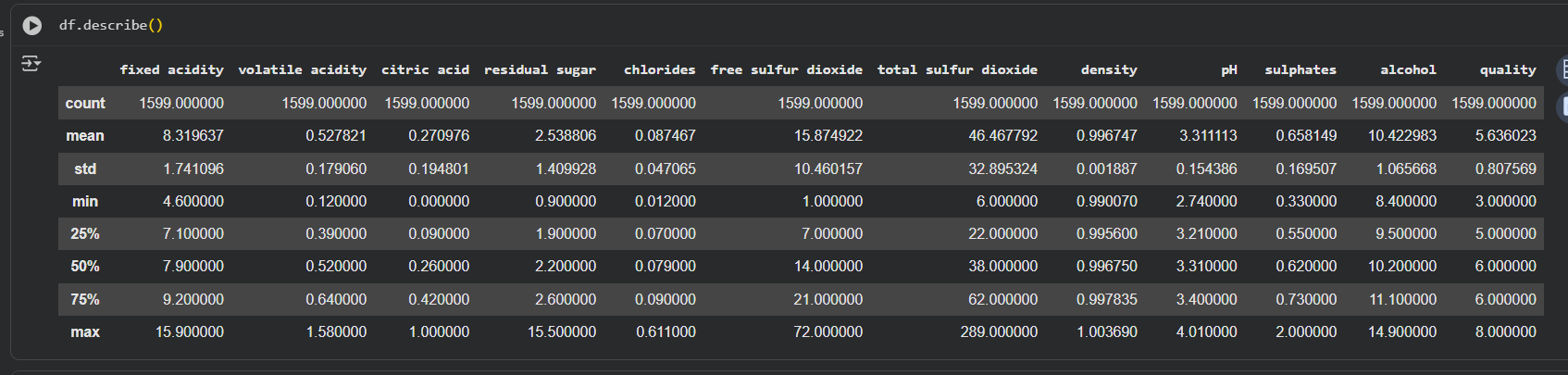
**pd.read\_csv** untuk membaca file CSV yang tersimpan di Google Drive berdasarkan path yang telah ditentukan. Hasil pembacaan ditampilkan dalam bentuk DataFrame, yang memperlihatkan data awal dari file winequality-red.csv, termasuk kolom seperti fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol, dan quality. Hasilnya, kita dapat melihat beberapa baris pertama data yang berhasil dimuat dengan mudah.

1. Langkah 2



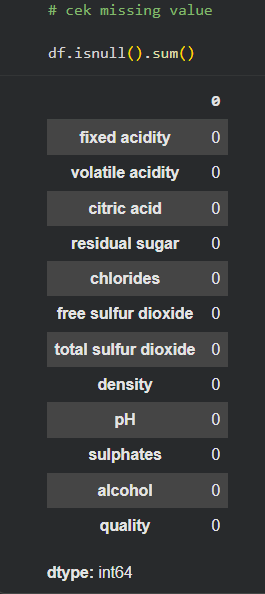
**df.info()** digunakan untuk menampilkan informasi ringkas mengenai struktur dataset yang telah dimuat ke dalam DataFrame. Berdasarkan hasil yang ditampilkan pada gambar, dataset memiliki 1599 baris data dan 12 kolom. Sebelas kolom bertipe data float64 seperti fixed acidity, volatile acidity, citric acid, residual sugar, hingga alcohol, sedangkan satu kolom bertipe int64 yaitu quality, yang berperan sebagai label atau target klasifikasi. Seluruh kolom memiliki nilai non-null, artinya tidak terdapat data yang kosong, sehingga dataset ini sudah bersih dan siap digunakan untuk proses analisis maupun pembuatan model. Ukuran memori yang digunakan sekitar 150.0 KB, menunjukkan bahwa dataset ini cukup ringan untuk diolah.

1. **Langkah 3**

****

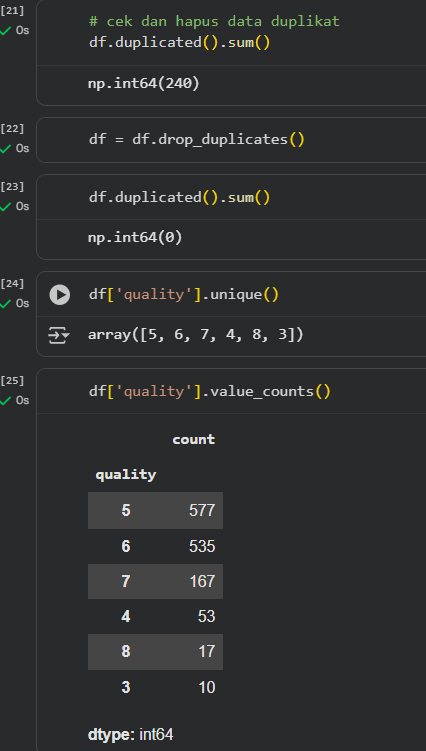
**df.describe()** digunakan untuk menampilkan statistik deskriptif dari setiap kolom numerik pada dataset. Berdasarkan hasil yang ditampilkan pada gambar, setiap kolom seperti fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol, dan quality memiliki nilai statistik seperti mean (rata-rata), std (standar deviasi), min (nilai minimum), max (nilai maksimum), serta 25%, 50%, dan 75% yang menunjukkan nilai kuartil.

1. **Langkah 4**

****

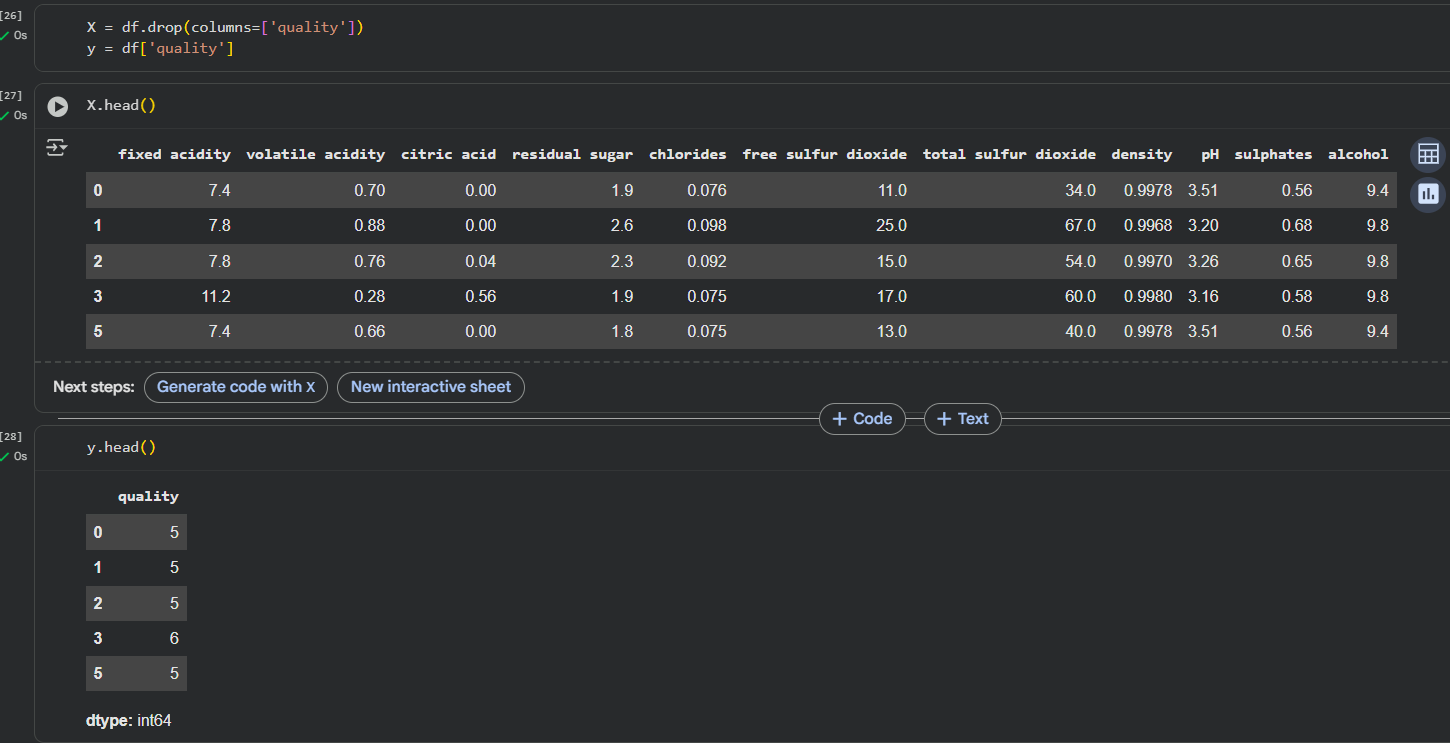
**df.isnull().sum()** digunakan untuk memeriksa jumlah nilai yang hilang (missing values) pada setiap kolom dalam dataset. Berdasarkan hasil yang ditampilkan pada gambar, semua kolom seperti fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol, dan quality memiliki nilai 0, yang berarti tidak terdapat data yang kosong atau hilang. Dengan demikian, dataset winequality-red.csv dapat langsung digunakan untuk proses analisis dan pelatihan model tanpa perlu melakukan tahap pembersihan data terlebih dahulu.

1. **Langkah 5**



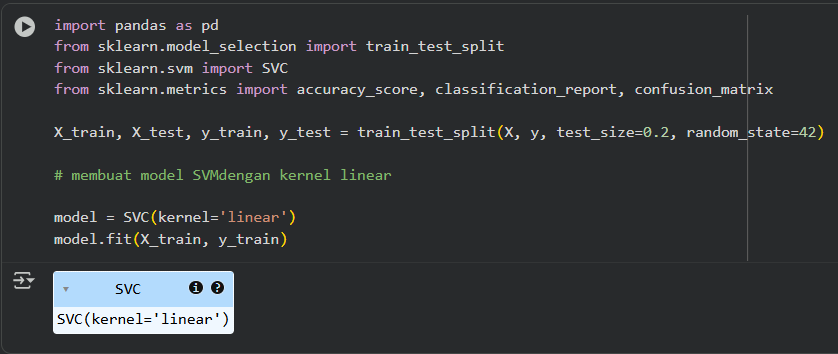
**df.duplicated().sum()** digunakan untuk mengecek jumlah data duplikat dalam dataset, dan hasilnya menunjukkan terdapat 240 baris duplikat. Setelah dihapus menggunakan **df.drop\_duplicates(),** dilakukan pengecekan ulang yang menunjukkan hasil 0, menandakan seluruh data duplikat telah berhasil dihapus. Selanjutnya, perintah **df['quality'].unique()** menampilkan nilai unik pada kolom quality, yaitu [5, 6, 7, 4, 8, 3], sedangkan **df['quality'].value\_counts()** digunakan untuk menghitung jumlah data di setiap kelas kualitas. Dari hasilnya terlihat bahwa nilai kualitas 5 dan 6 memiliki jumlah terbanyak, sedangkan 3, 4, dan 8 paling sedikit, menunjukkan bahwa distribusi data pada kolom quality tidak seimbang.

1. **Langkah 6**

****

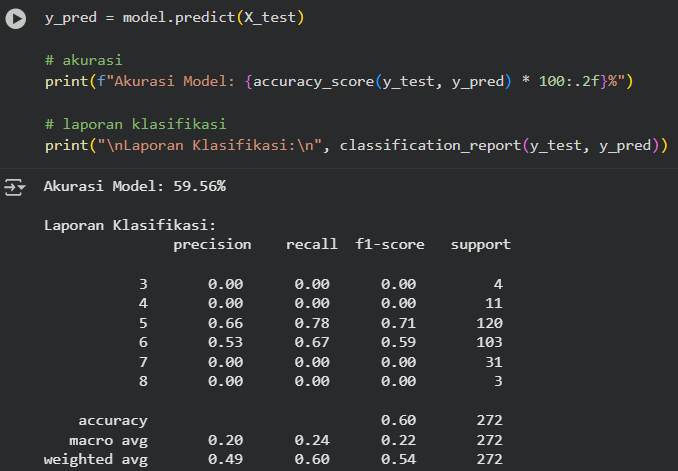
**X = df.drop(columns=['quality'])** untuk memisahkan variabel independen dari dataset dengan cara menghapus kolom quality, sedangkan **y = df['quality']** digunakan untuk mengambil kolom target (variabel dependen) yang akan diprediksi. Hasil perintah **X.head()** menampilkan beberapa baris pertama dari fitur seperti fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, dan alcohol. Sementara itu, hasil **y.head()** menampilkan nilai awal dari kolom quality yang berisi angka kualitas wine, yaitu 5 dan 6. Dengan pemisahan ini, data siap digunakan untuk proses pelatihan model machine learning.

1. **Langkah 7**

****

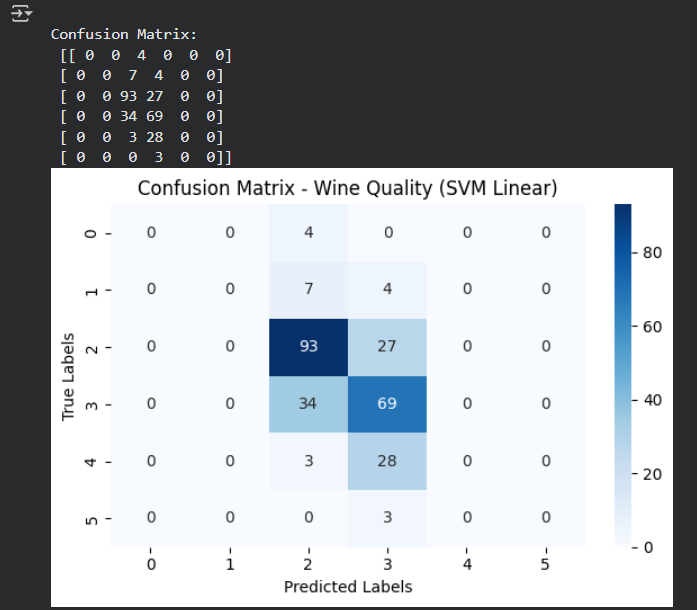
Kode di atas digunakan untuk membuat model Support Vector Machine (SVM) dengan kernel linear untuk klasifikasi. Library pandas, train\_test\_split, SVC, dan metrik evaluasi dari sklearn diimpor untuk pengolahan data, pembagian dataset, pembuatan model, serta evaluasi hasil. Dataset dibagi menjadi 80% data latih dan 20% data uji, kemudian model SVM dengan kernel linear dibuat menggunakan SVC(kernel='linear') dan dilatih dengan model.fit(X\_train, y\_train) agar dapat mempelajari pola dari data latih.

1. **Langkah 8**

****

Perintah **model.predict(X\_test)** menghasilkan prediksi data uji, lalu **accuracy\_score** menghitung akurasi sebesar 59.56%. Sementara **classification\_report** menampilkan nilai precision, recall, dan f1-score untuk tiap kelas, yang menunjukkan bahwa model dengan kernel linear belum optimal karena beberapa kelas masih memiliki skor rendah akibat ketidakseimbangan data.

1. **Langkah 9**



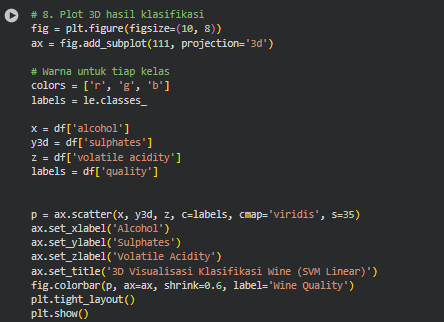
Gambar di atas menampilkan Confusion Matrix hasil prediksi model SVM dengan kernel linear. Nilai diagonal utama menunjukkan prediksi yang benar untuk tiap kelas wine, model paling akurat pada kelas 5 dan 6, sedangkan kelas lain masih sering salah diprediksi, jadi model belum optimal dalam membedakan semua kelas kualitas wine.

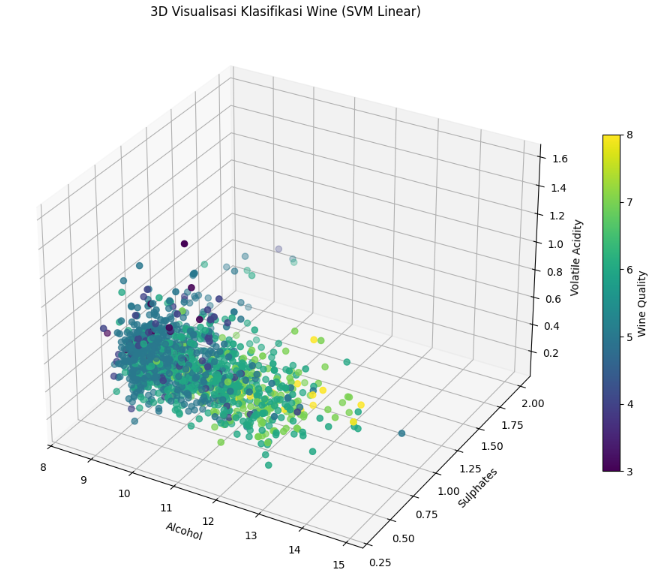
1. **Langkah 10**

|  |  |
| --- | --- |
|  |  |
|  |  |

Gambar diatas adalah hasil visualisasi 2D hubungan antara kadar alkohol dan volatile acidity terhadap kualitas wine. Titik-titik pada grafik menunjukkan sampel wine, di mana sumbu X merepresentasikan Alcohol, sumbu Y Volatile Acidity, dan warna titik menggambarkan nilai kualitas wine. Gambar tersebut menunjukan bahwa wine dengan kadar alkohol tinggi umumnya memiliki kualitas lebih baik, sedangkan wine dengan volatile acidity tinggi cenderung berkualitas rendah.

1. **Langkah 11**





Gambar di atas menampilkan visualisasi 3D hasil klasifikasi wine menggunakan SVM dengan kernel linear. Dari hasil terlihat hubungan antara tiga fitur utama, yaitu Alcohol (X), Sulphates (Y), dan Volatile Acidity (Z), dengan warna titik mewakili kualitas wine. Dari grafik terlihat bahwa sampel wine dengan kadar alkohol dan sulphates lebih tinggi cenderung memiliki kualitas yang lebih baik (warna kuning), sedangkan wine dengan volatile acidity tinggi umumnya memiliki kualitas lebih rendah (warna ungu atau biru tua).

**Link Github:**