**MERAMAL CUACA**

Kelompok : 1. M Fauzi Eka A11.2021.13357

2. Aryasatya Wistara Adhi A11.2021.13678

3. Muhammad Hafidz Al H A11.2021.13664

1. **Penjelasan Dataset**

Dataset Meramal Cuaca (MC) terdiri atas dua data, yaitu `Data Harian` dan `Data Jam`. Data Harian memiliki **\*\*14730 baris\*\*** dan Data Jam memiliki **\*\*353520 baris\*\***. Terdapat **\*\*17 kolom\*\*** pada data harian dan **\*\*27 kolom\*\*** pada data jam. detail kolom-kolom tersebut adalah sebagai berikut:

**\*\*Data harian:\*\***

\* `Id` - id cuaca

\* `time` - Tanggal pencatatan

\* `temperature\_2m\_max (°C)` - Temperatur udara tertinggi pada ketinggian 2 m di atas permukaan

\* `temperature\_2m\_min (°C)` - Temperatur udara terendah pada ketinggian 2 m di atas permukaan

\* `apparent\_temperature\_max (°C)` - Temperatur semu maksimum yang terasa

\* `apparent\_temperature\_min (°C)` - Temperatur semu minimum yang terasa

\* `sunrise (iso8601)` - Waktu matahariterbit pada hari itu dengan format ISO 8601

\* `sunset (iso8601)` - Waktu mataharitenggelam pada hari itu dengan format ISO 8601

\* `shortwave\_radiation\_sum (MJ/m²)` - Total radiasi matahari pada hari tersebut

\* `rain\_sum (mm)` - Jumlah curah hujan pada hari tersebut - Target Regresi

\* `snowfall` - Jumlah salju pada haritersebut

\* `precipitation` - Jumlah curah hujan harian (termasuk hujan, hujan lebat, dan salju)

\* `windspeed\_10m\_max (km/h)` - Kecepatan angin maksimum pada ketinggian 10 m

\* `windgusts\_10m\_max (km/h)` - Kecepatan angin minimum pada ketinggian 10 m

\* `winddirection\_10m\_dominant (°)` - Arah angin dominan pada haritersebut

\* `city` - Nama kota yang tercatat

\* `class` - Jenis/kelompok curah hujan harian - Target Klasifikasi

**\*\*Data Jam:\*\***

\* `time` - Tanggal dan jam pencatatan

\* `temperature\_2m (°C)` - Temperatur pada ketinggian 2 m

\* `relativehumidity\_2m (%)` - Kelembapan pada ketinggian 2 m

\* `winddirection\_100m (°)` - Arah angin pada ketinggian 100 m

\* `windgusts\_10m (km/h)` - Kecepatan angin ketika terdapat angin kencang

\* `et0\_fao\_evapotranspiration (mm)` - Jumlah evapotranspirasi (evaporasi dan transpirasi) pada jam tersebut

\* `vapor\_pressure\_deficit (kPa)` - Perbedaan tekanan uap air dari udar

dengan tekanan uap air ketika udara tersaturasi

\* `soil\_temperature\_0\_to\_7cm (°C)` - Rata-rata temperatur tanah pada kedalaman 0-7 cm

\* `soil\_temperature\_7\_to\_28cm (°C)` - Rata-rata temperatur tanah pada kedalaman 7-28 cm

\* `soil\_temperature\_28\_to\_100cm (°C)` - Rata-rata temperatur tanah pada kedalaman 28-100 cm

\* `soil\_temperature\_100\_to\_255cm (°C)` - Rata-rata temperatur tanah pada kedalaman 100-255 cm

\* `soil\_moisture\_0\_to\_7cm (m³/m³)` - Rata-rata kelembapan air pada tanah untuk kedalaman 0-7 cm

\* `soil\_moisture\_7\_to\_28cm (m³/m³)` - Rata-rata kelembapan air pada tanah untuk kedalaman 7-28 cm

\* `dewpoint\_2m (°C)` - Titik embun; suhu ambang udara mengembun

\* `apparent\_temperature (°C)` - Temperatur semu yang dirasakan

\* `pressure\_msl (hPa)` - Tekanan udara pada ketinggian permukaan air laut rata-rata (mean sea level)

\* `surface\_pressure (hPa)` - Tekanan udara pada ketinggian permukaan daerah tersebut

\* `snowfall (cm)` - Jumlah hujan salju pada jam tersebut

\* `cloudcover\_low (%)` - Persentase cloud cover pada awan sampai ketinggian 2 km

\* `cloudcover\_mid (%)` - Persentase cloud cover pada ketinggian 2-6 km

\* `cloudcover\_high (%)` - Persentase cloud cover pada ketinggian di atas 6 km

\* `windspeed\_10m (km/h)` - Kecepatan angin pada ketinggian 10 m

\* `windspeed\_100m (km/h)` - Kecepatan angin pada ketinggian 100 m

\* `winddirection\_10m (°)` - Arah angin pada ketinggian 10 m

\* `soil\_moisture\_28\_to\_100cm (m³/m³)` - Rata-rata kelembapan air pada tanah untuk kedalaman 28-100 cm

\* `soil\_moisture\_100\_to\_255cm (m³/m³)` - Rata-rata kelembapan air pada tanah untuk kedalaman 100-255 cm

\* `city` - Nama kota

1. **Permasalahan dan Tujuan**

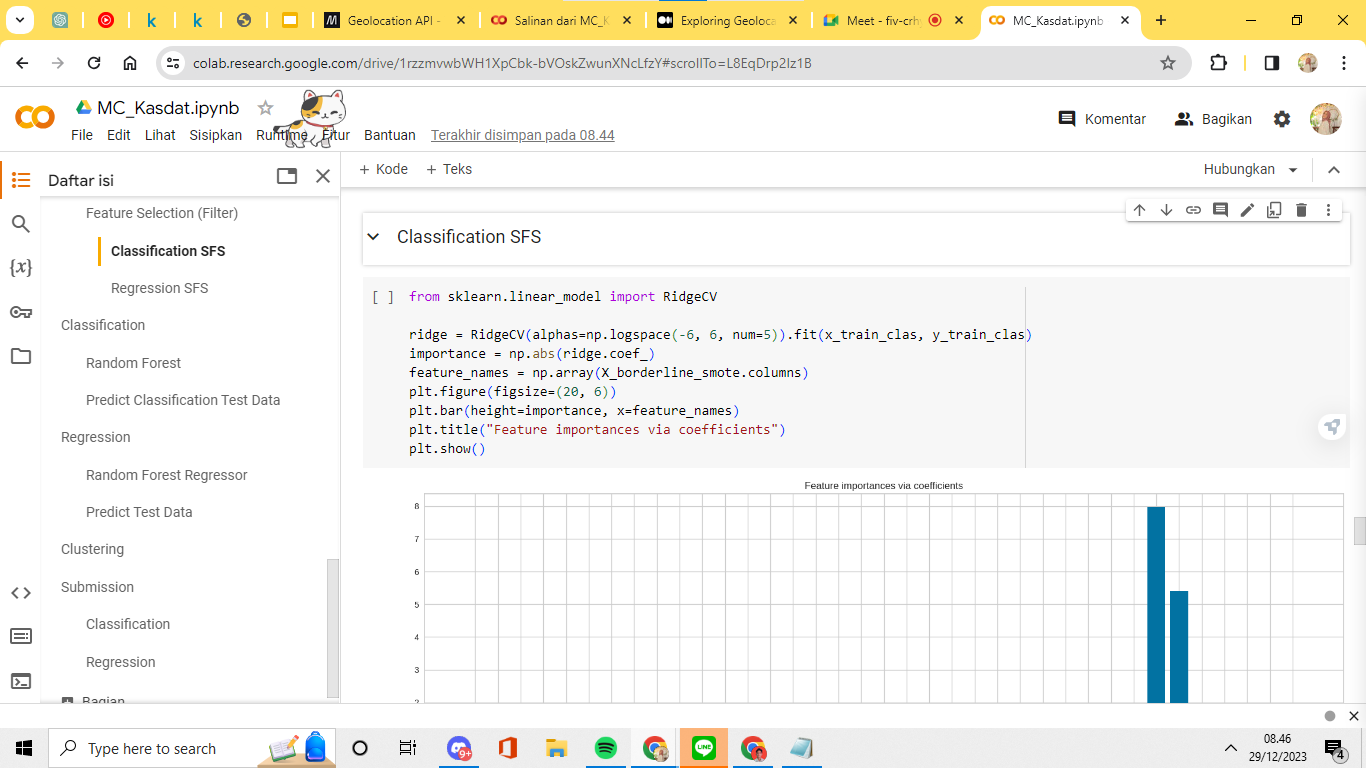
Peneliti melakukan penelitian ini dengan tujuan sebagai berikut :

* Menemukan hubungan ke empat kolom mengenai `soil\_moisture` dengan jumlah curah hujan harian
* Pasangan kota yang dekat secara geografis
* Mencari letak daerah berdasarkan kota
* Mencari 3 kota yang memiliki rata-rata jumlah curah hujan tahunan tertinggi beserta nilainya

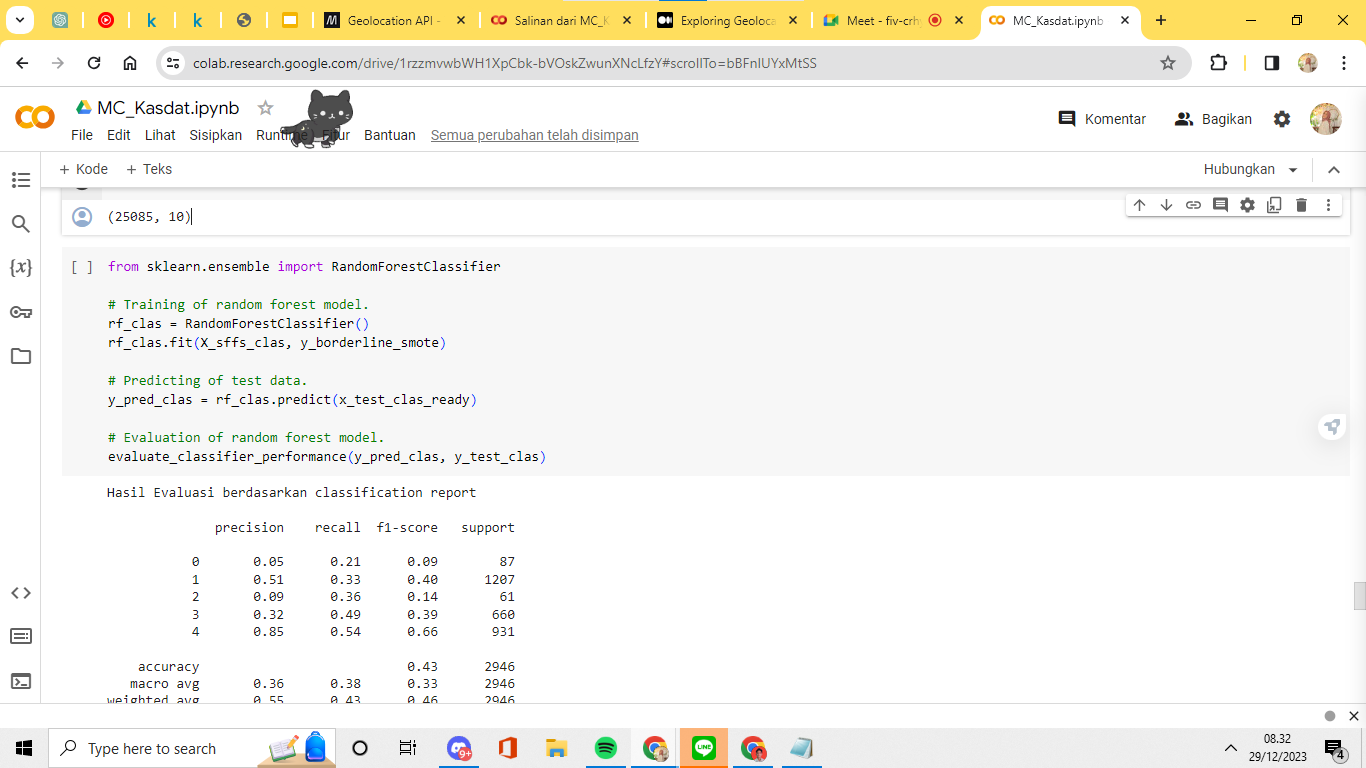
1. **Model**
2. **Klasifikasi**

* **Feature Selection using SFFS**

Sebelum menguji performa, dilakukan feature selection menggunakan metode SFFS



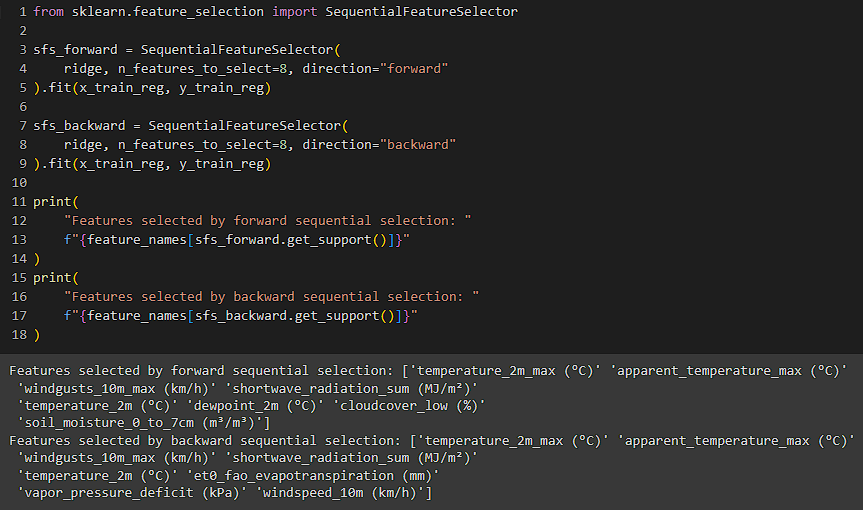
* **Model (Random Forest Classifier)**

****

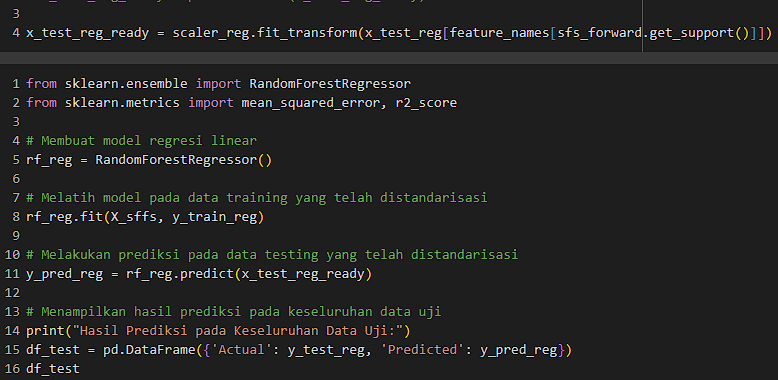
1. **Regresi**

* **Feature Selection using SFFS**

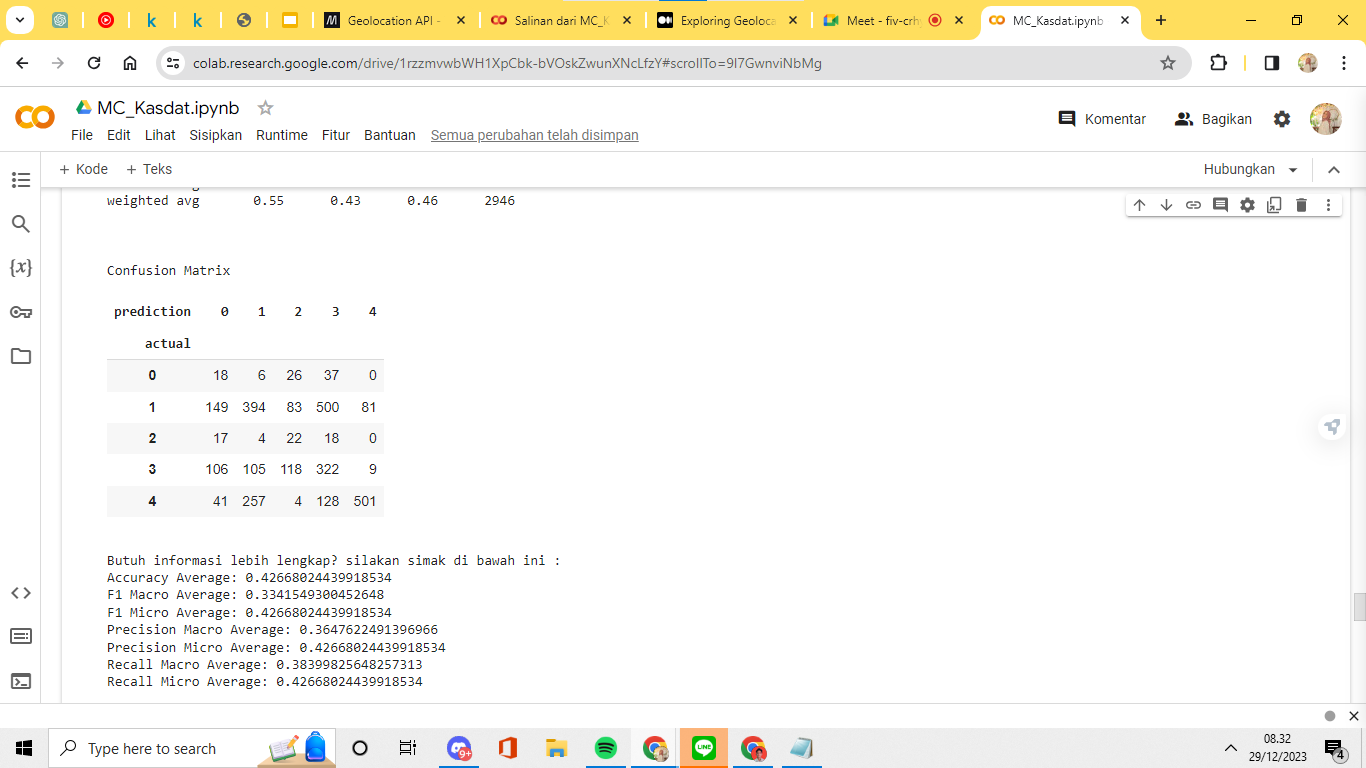
Sebelum menguji performa, dilakukan feature selection menggunakan metode SFFS



* **Model (Random Forest Regressor)**



1. **Performa Model**
2. **Klasifikasi**

****

Hasil uji performa model memperoleh hasil sebesar 0.42668024439918534

1. **Regresi**

****

Hasil uji performa menggunakan R-Square menunjukkan hasil sebesar 0.39758

1. **Kesimpulan & Rekomendasi**

Kolom soil moisture memiliki hubungan korelasi yang kuat satu sama lain. Hal ini tidak kita inginkan dalam pelatihan model, karena akan membuat informasi menjadi redundan dan dapat menyebabkan overfit. Oleh karena itu, apabila ingin menggunakan soil moisture sebagai fitur, cukup kita pilih salah satu.Adapun hubungan antara keempat kolom soil moisture dengan curah hujan terbilang cukup rendah, sehingga perlu kita pertimbangkan apakah akan menggunakan kolom soil moisture sebagai fitur. Namun, untuk mendapatkan daerah dengan kota tertentu, kita perlu melakukan scraping map menggunakan API yang mana hal ini akan cukup sulit.

Diperlukan feature engineering untuk menemukan daerah dengan kota - kota tertentu, seperti penemuan fitur latitude, longitude dan lain - lain.