**LAPORAN MACHINE LEARNING KECERDASAN BUATAN**

*Untuk memenuhi tugas besar mata kuliah Kecerdasan Buatan*



Naufal Reky Ardhana / 4.33.22.0.21

Rachmadandy Mahendra Shakti / 4.33.22.0.22

**JURUSAN TEKNIK ELEKTRO**

**PROGRAM STUDI TEKNOLOGI REKAYASA KOMPUTER**

**POLITEKNIK NEGERI SEMARANG**

**2023**

# KLASIFIKASI

**Rachmadandy Mahendra Shakti**

**4.33.22.0.22**

**TI-2A**

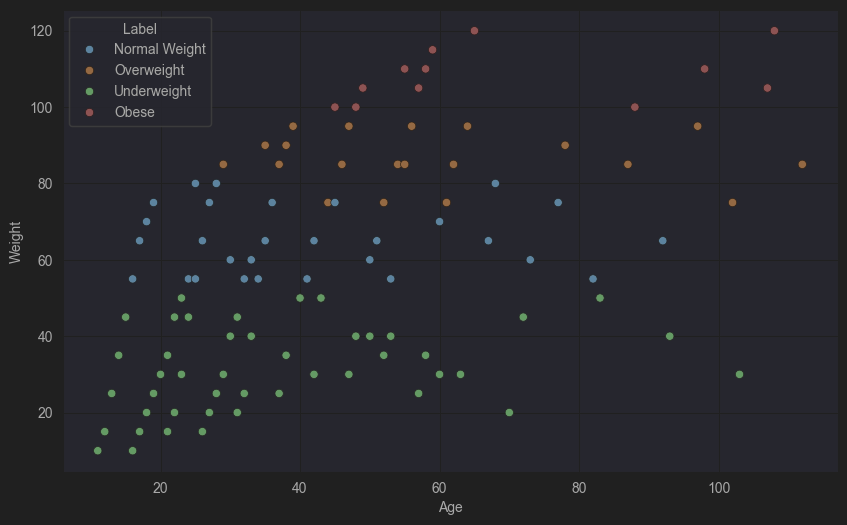
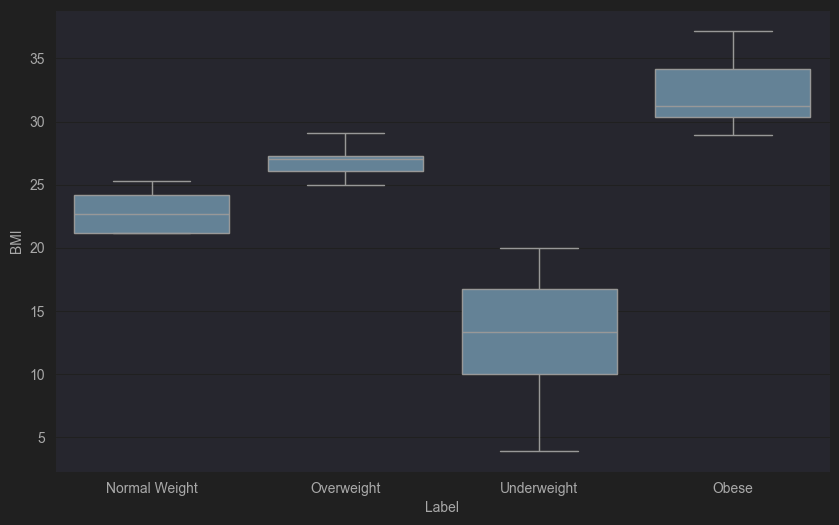
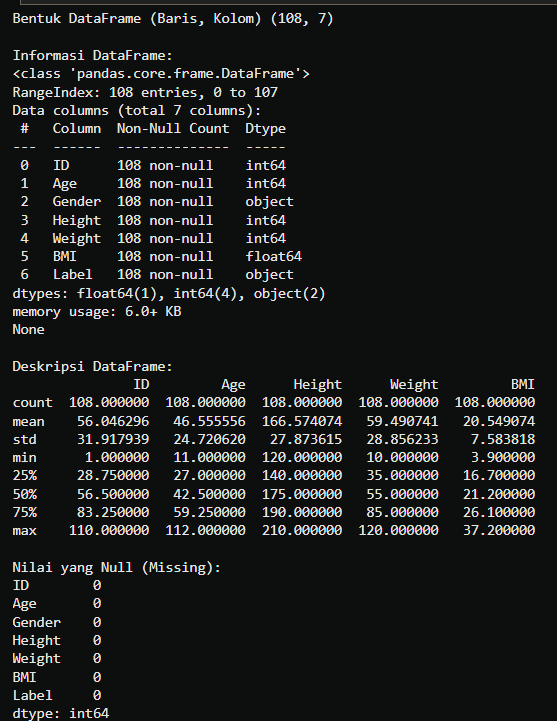
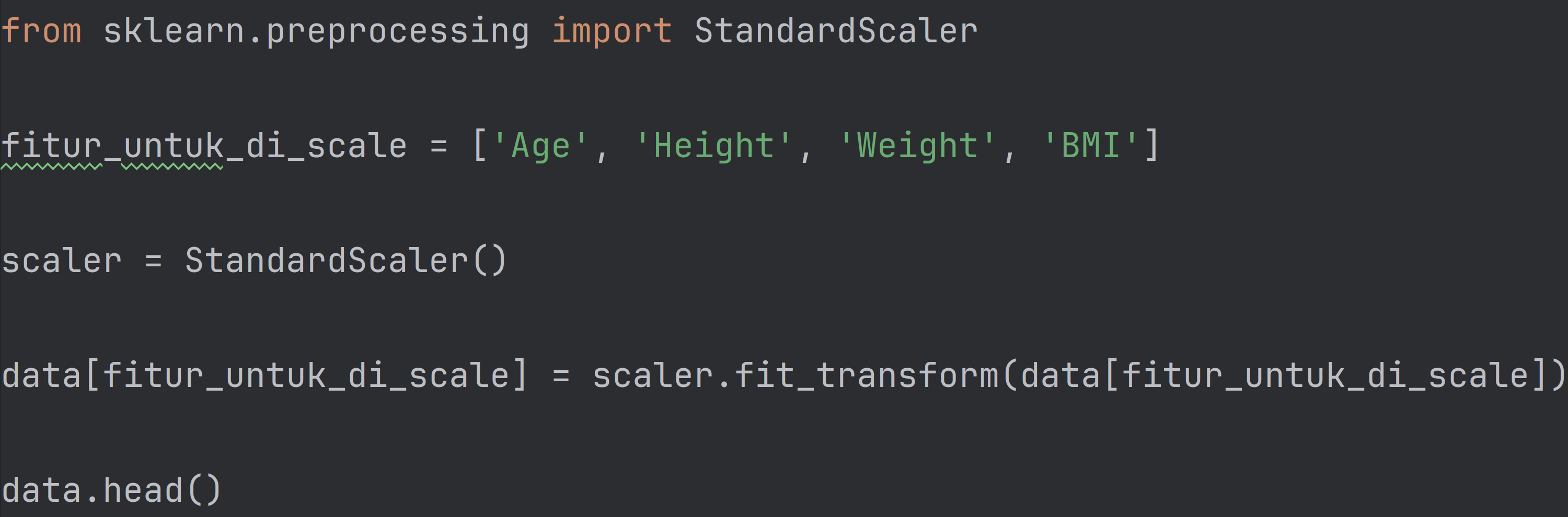
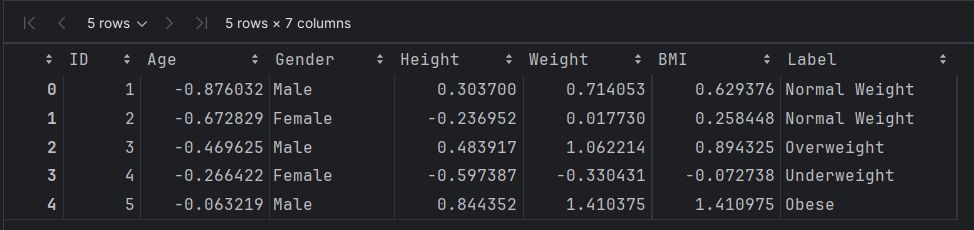
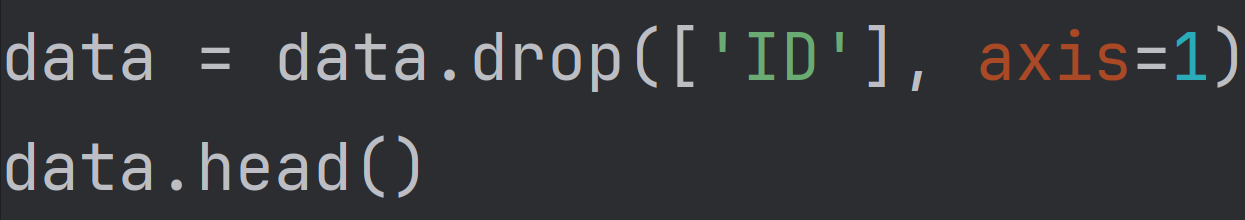
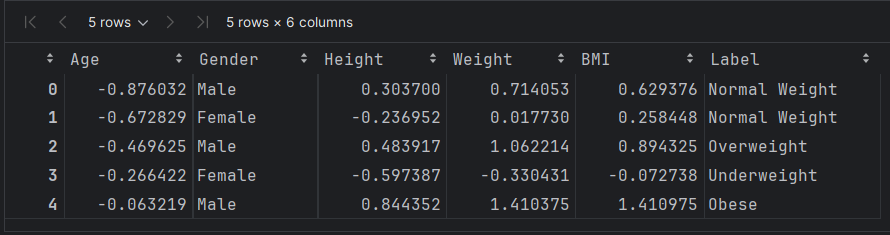
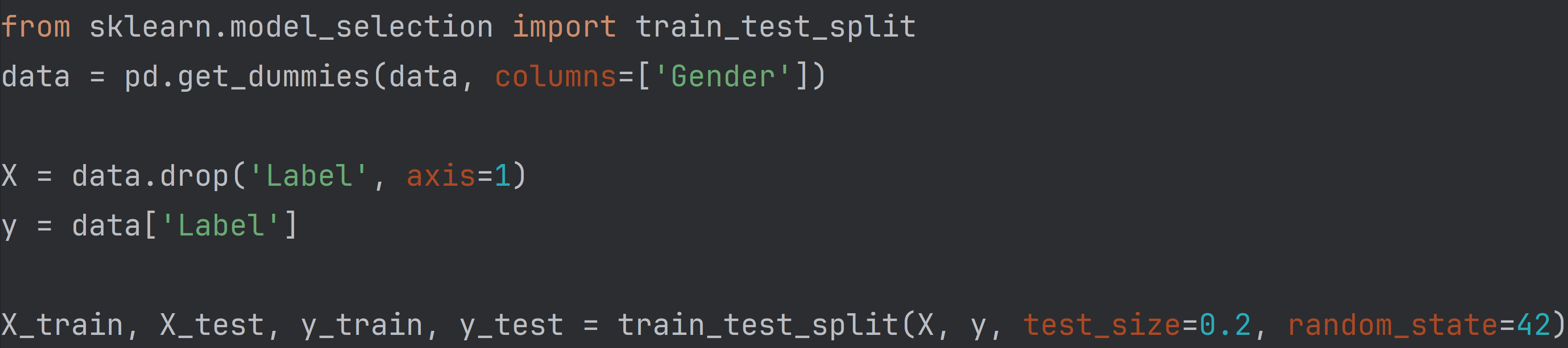
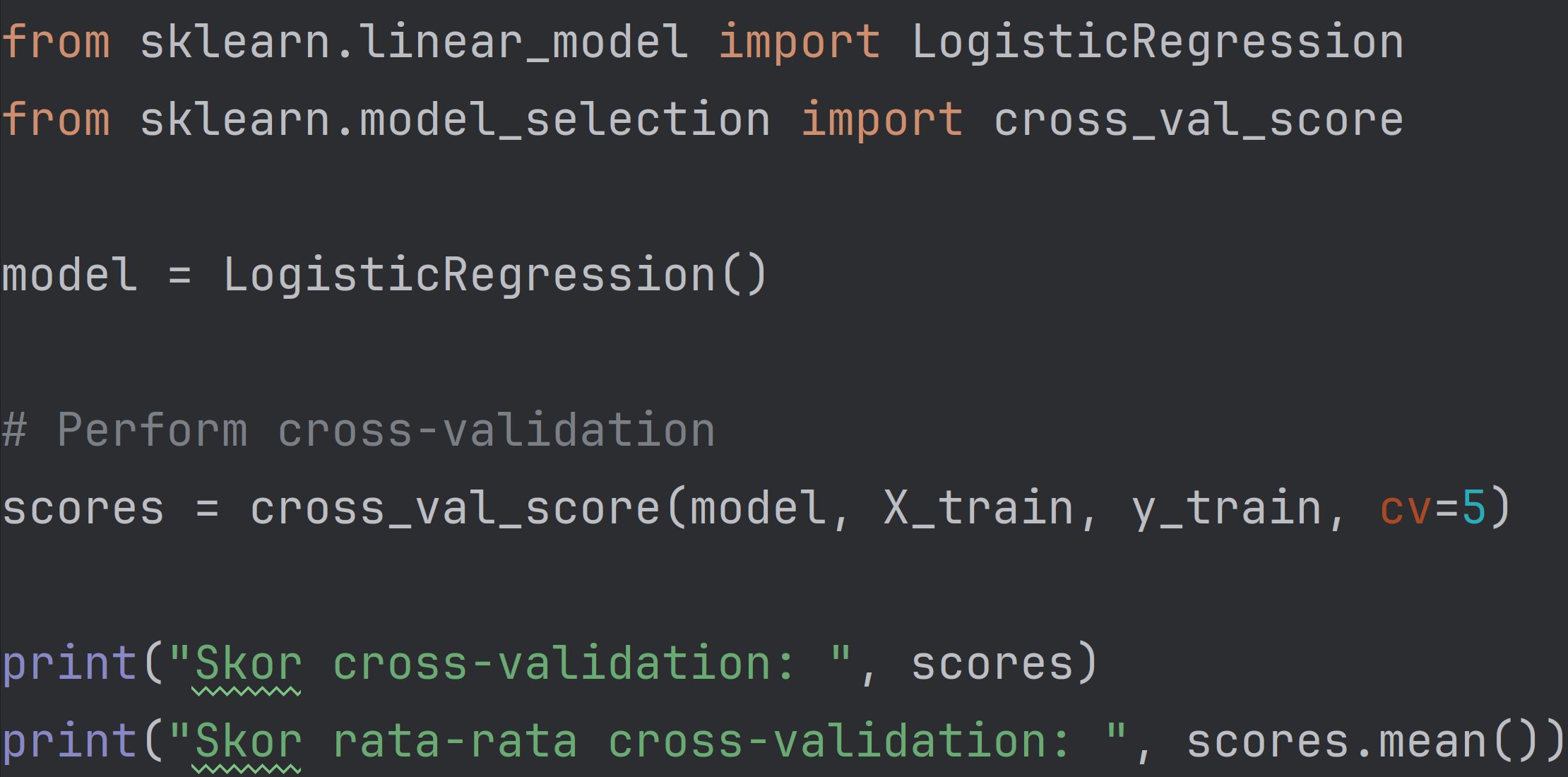
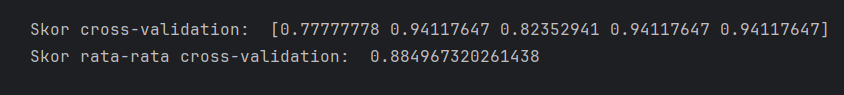
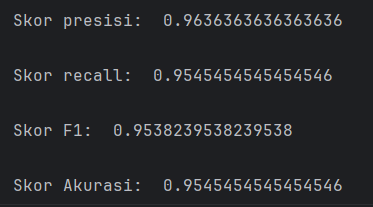
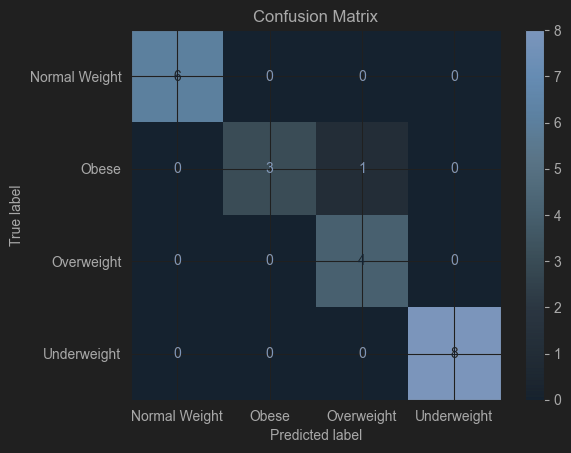
Klasifikasi adalah salah satu teknik dalam machine learning yang digunakan untuk mengkategorikan data ke dalam kelas atau label tertentu. Tujuannya adalah untuk memprediksi kelas target dari data yang diberikan. Contoh klasifikasi dalam kehidupan nyata termasuk deteksi spam email, diagnosis penyakit, dan pengenalan wajah.

Beberapa langkah yang dilakukan adalah EDA, Scaling Data, Feature Selection, Training, Cross Validation, dan Performance Evaluation dengan beberapa parameter, yaitu skor F1, skor Recall, Skor Presisi, Skor akurasi dan Confusion Matrix.

Analisis Data Eksploratif (EDA) adalah proses untuk memahami dan menganalisis data melalui statistik deskriptif dan visualisasi, yang bertujuan untuk memahami pola, hubungan, atau anomali yang ada dalam data. Penskalaan data adalah proses mengubah rentang nilai data sehingga semua fitur memiliki skala yang sama, biasanya antara 0 dan 1 atau -1 dan 1, yang penting untuk beberapa algoritma machine learning. Seleksi fitur adalah proses memilih fitur yang paling relevan untuk model prediksi, yang dapat mengurangi overfitting, meningkatkan akurasi, dan mengurangi waktu pelatihan. Pembagian data adalah proses membagi data menjadi set pelatihan dan set pengujian untuk memvalidasi performa model. Validasi silang adalah teknik untuk mengevaluasi performa model dengan membagi data menjadi beberapa bagian dan melakukan pelatihan dan pengujian berulang kali. Evaluasi performa adalah tahap akhir untuk mengukur seberapa baik model melakukan prediksi dengan menggunakan metrik seperti akurasi, presisi, recall, dan skor F1.

Dataset yang digunakan adalah obesity.csv, file obesity.csv adalah file data yang berisi informasi tentang berat badan individu. Setiap baris dalam file ini mewakili satu individu dan mencakup kolom berikut: ID, Age (Usia), Gender (Jenis Kelamin), Height (Tinggi), Weight (Berat), BMI (Indeks Massa Tubuh), dan Label (kategori berat badan). Anda menggunakan file ini sebagai dataset untuk melakukan klasifikasi berat badan.

Step Pengerjaan

1. EDA
   1. 
   2. 
2. Scaling
   1. 
   2. 
3. Feature Selection
   1. 
   2. 
4. Training
   1. 
5. Cross Validation
   1. 
   2. 
6. Performance Evaluation
   1. 
   2. 

# REGRESI

**Naufal Reky Ardhana**

**4.33.22.0.21**

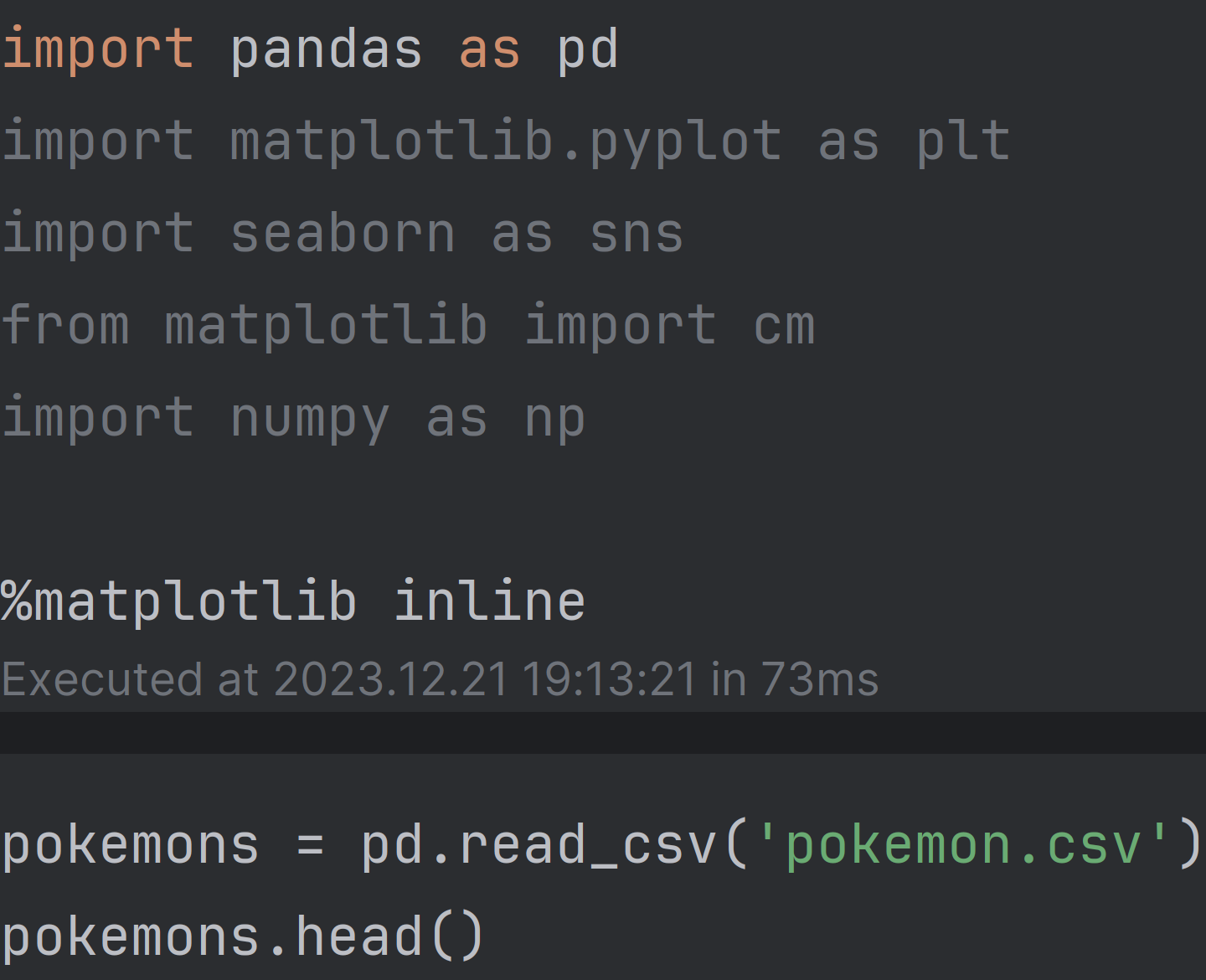
**TI-2A**

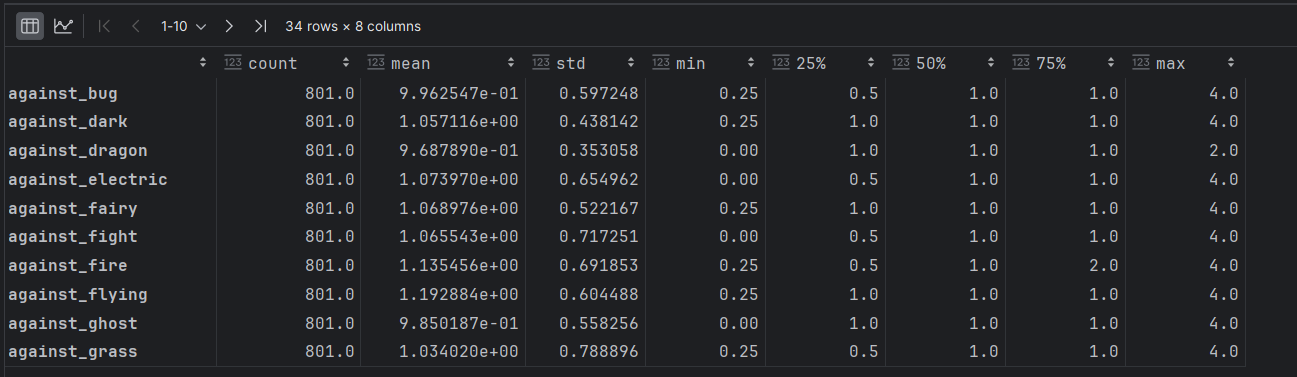
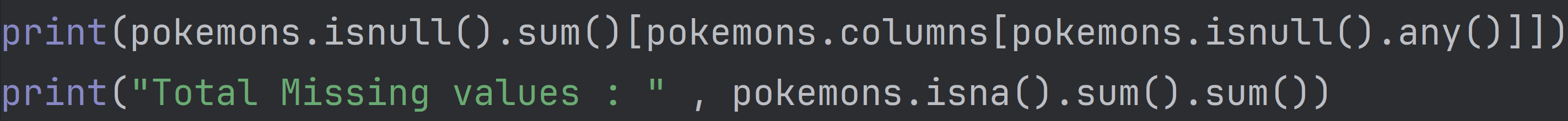
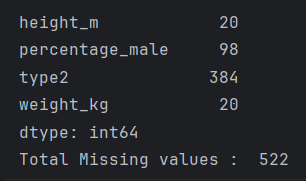
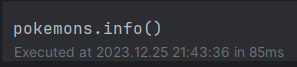
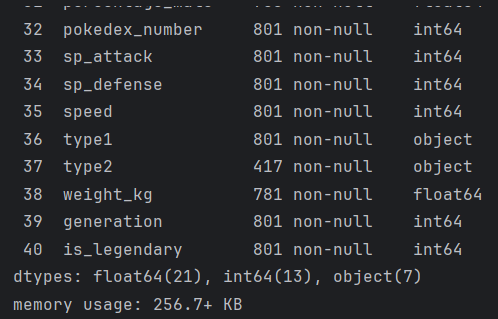
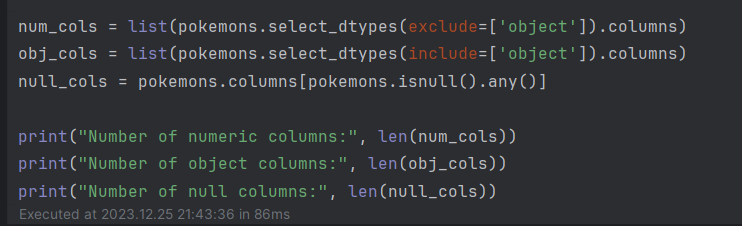
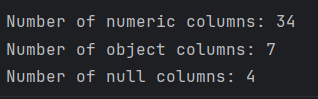
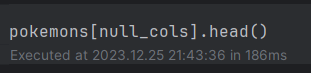
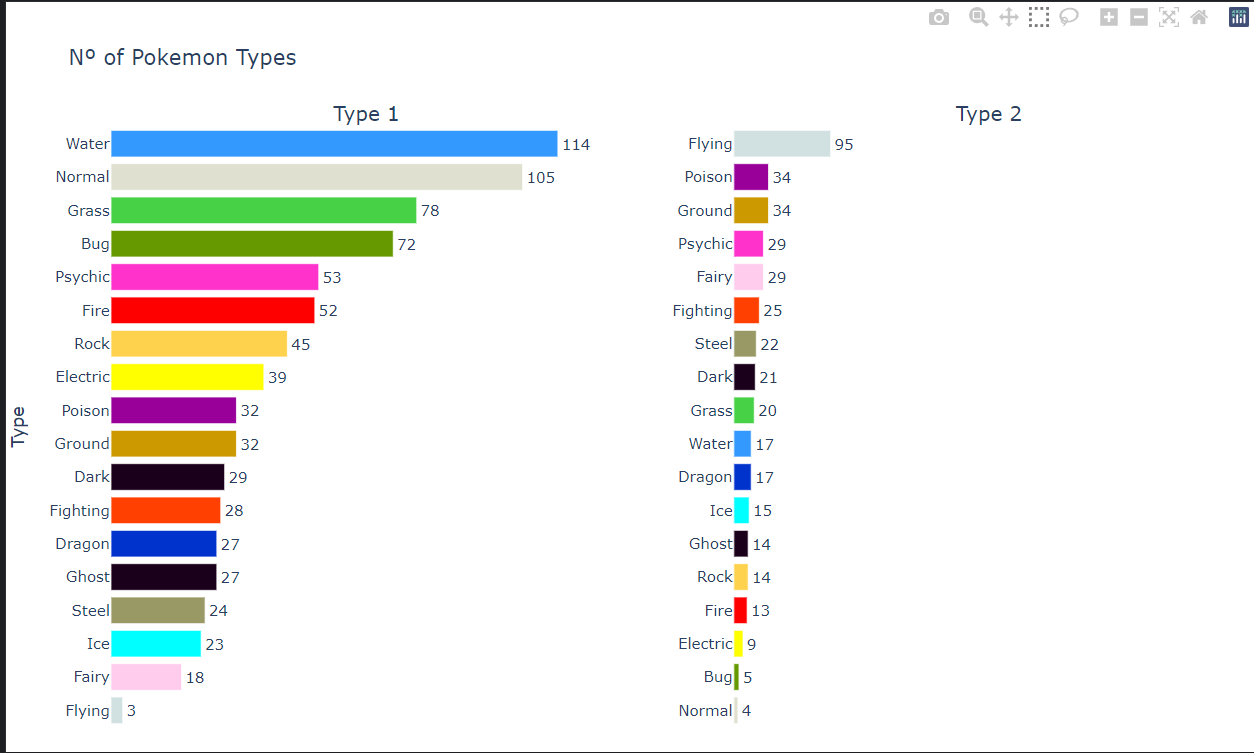
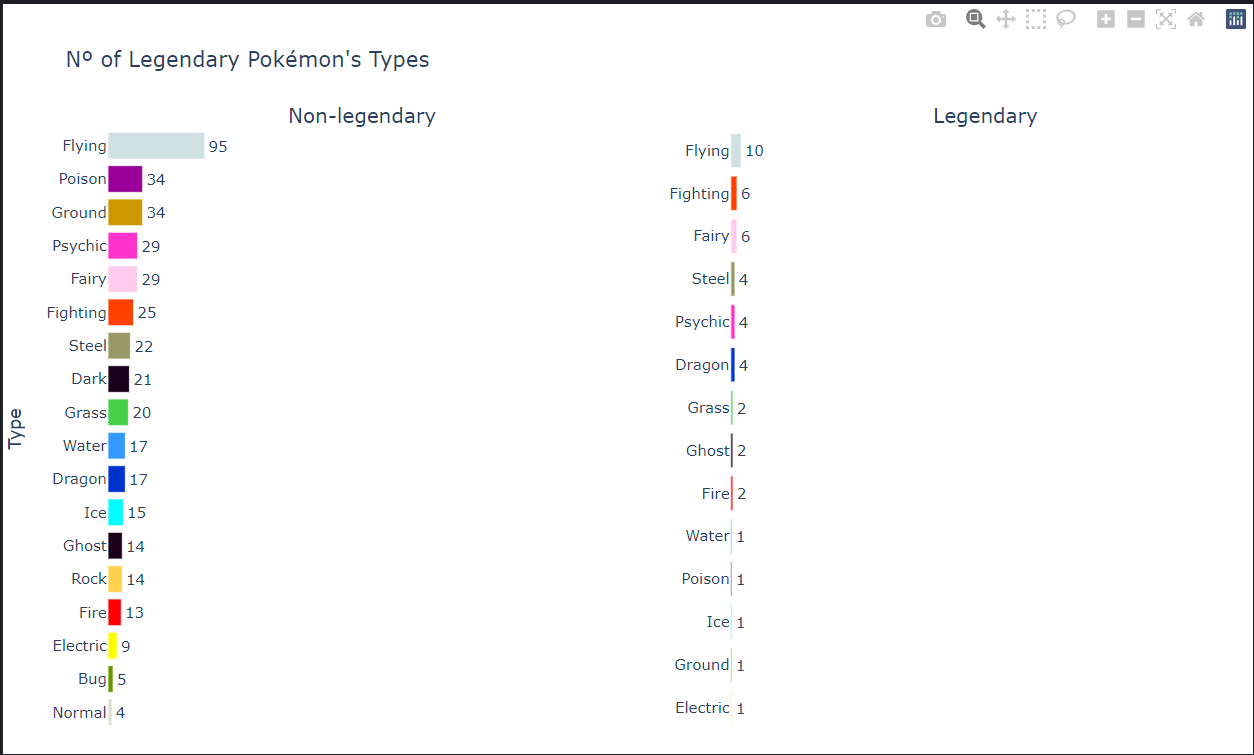
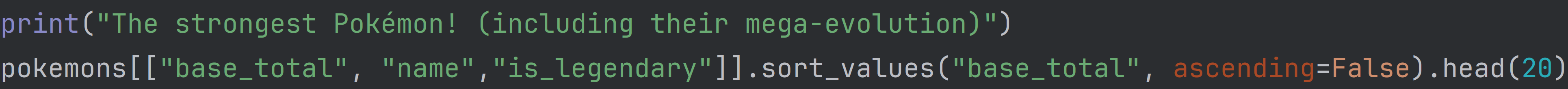
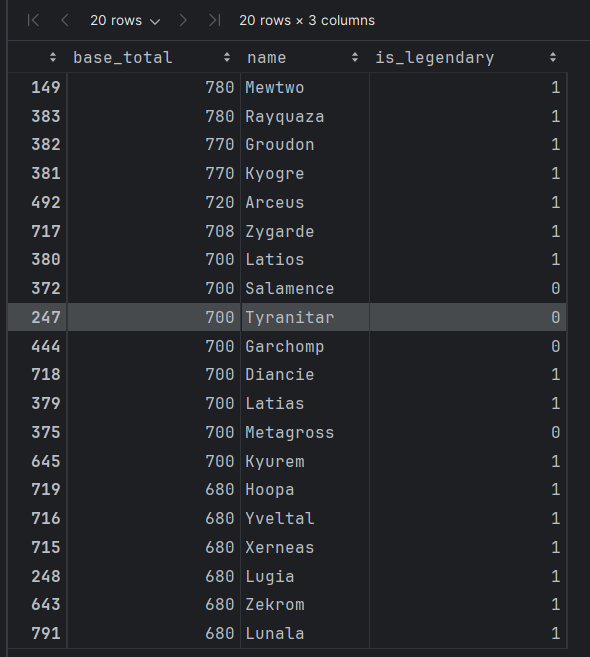
Regresi adalah teknik statistik yang digunakan untuk menentukan hubungan antara dua atau lebih variabel. Dalam konteks machine learning, regresi biasanya digunakan untuk memprediksi output kontinu berdasarkan satu atau lebih variabel input. Misalnya, kita mungkin ingin memprediksi harga rumah berdasarkan luas tanah, jumlah kamar tidur, dan lokasi. Dalam hal ini, harga rumah adalah variabel dependen (yang ingin kita prediksi), dan luas tanah, jumlah kamar tidur, dan lokasi adalah variabel independen (yang digunakan untuk membuat prediksi). Ada beberapa jenis regresi, termasuk regresi linier, regresi logistik, regresi polinomial, dan lainnya. Setiap jenis regresi memiliki kegunaan dan aplikasi tertentu tergantung pada sifat data dan apa yang ingin kita prediksi.

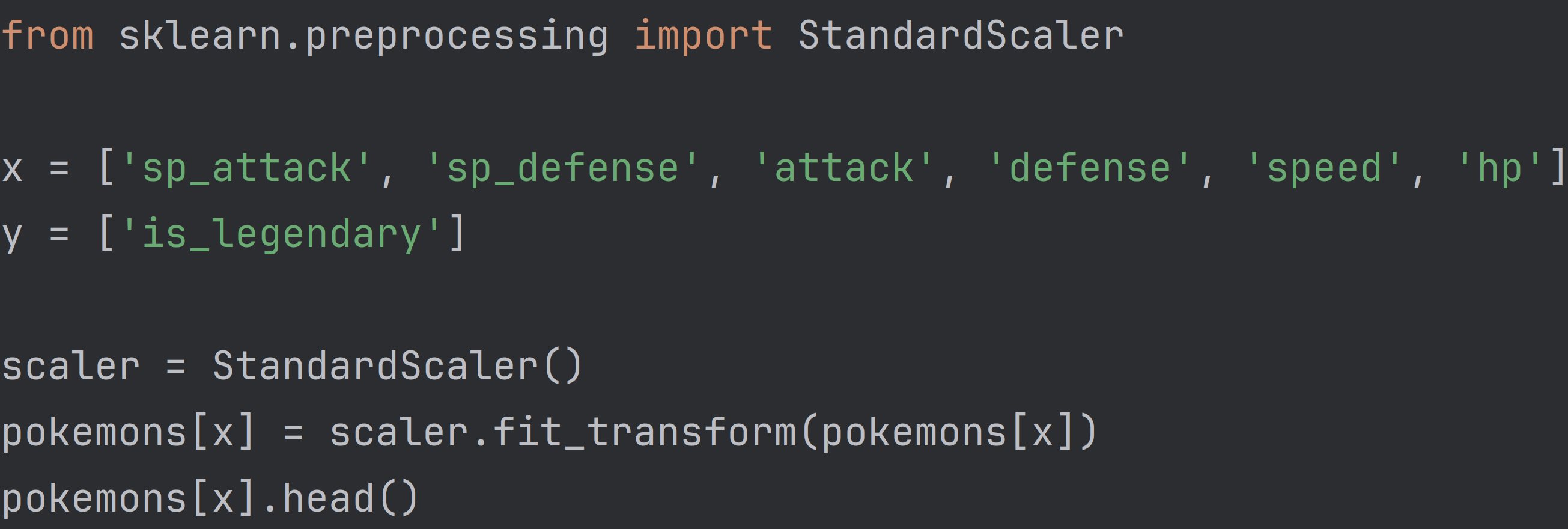
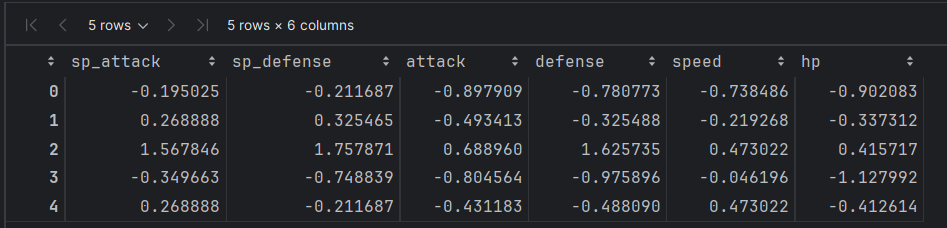
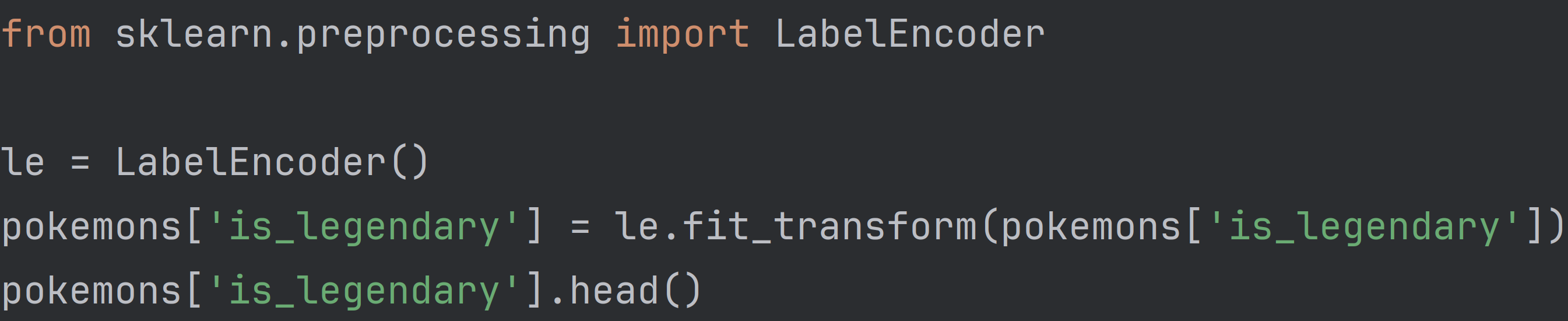
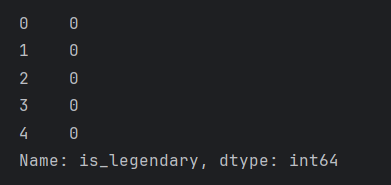
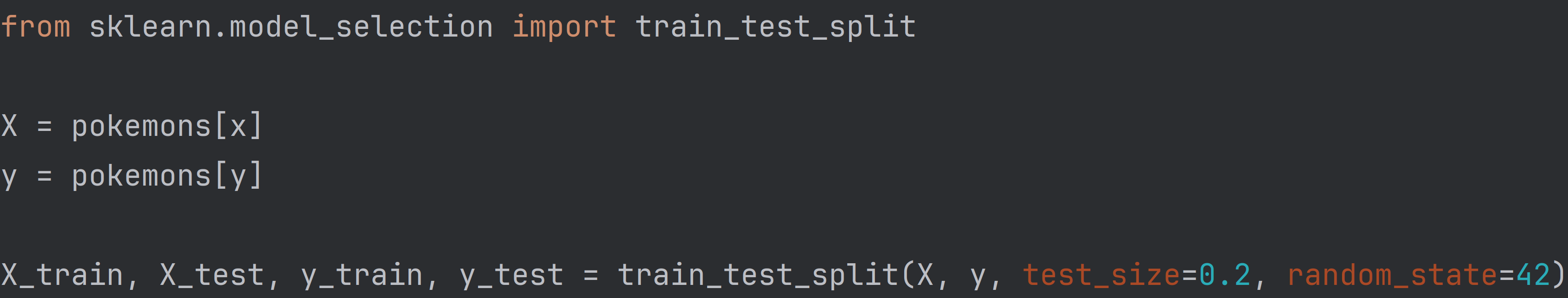
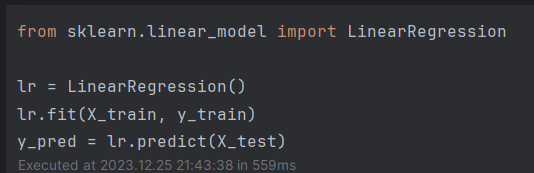
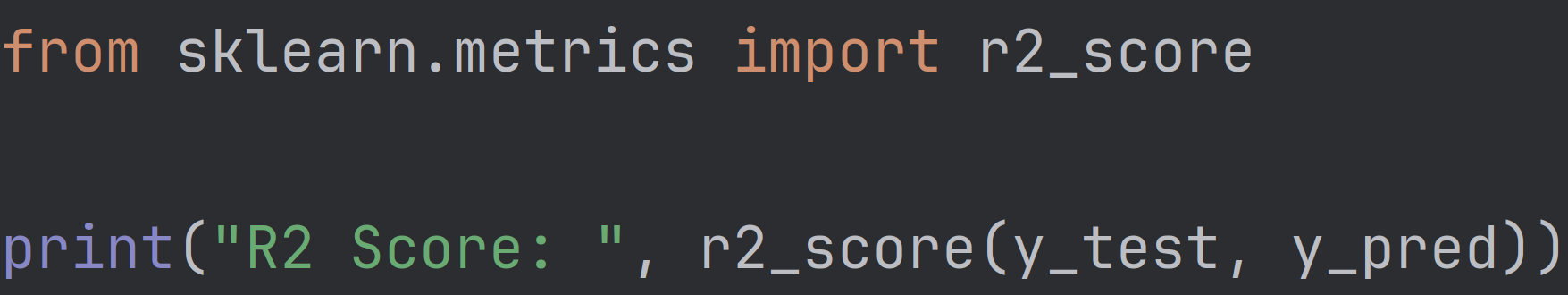
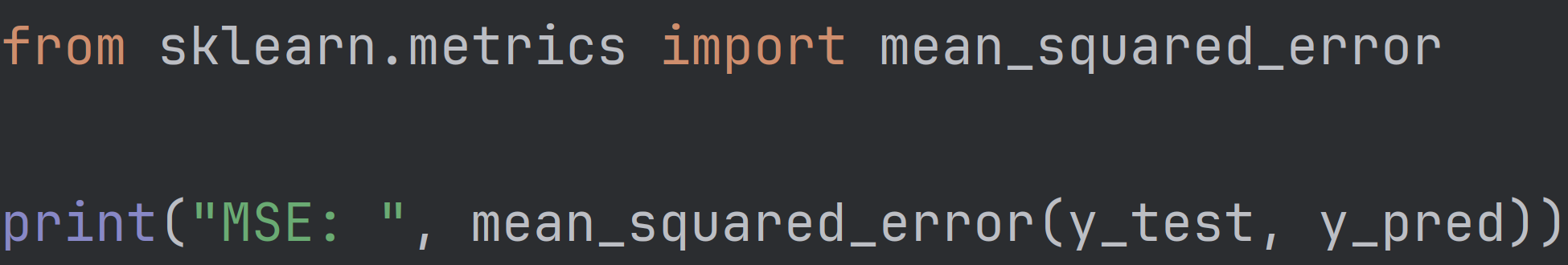
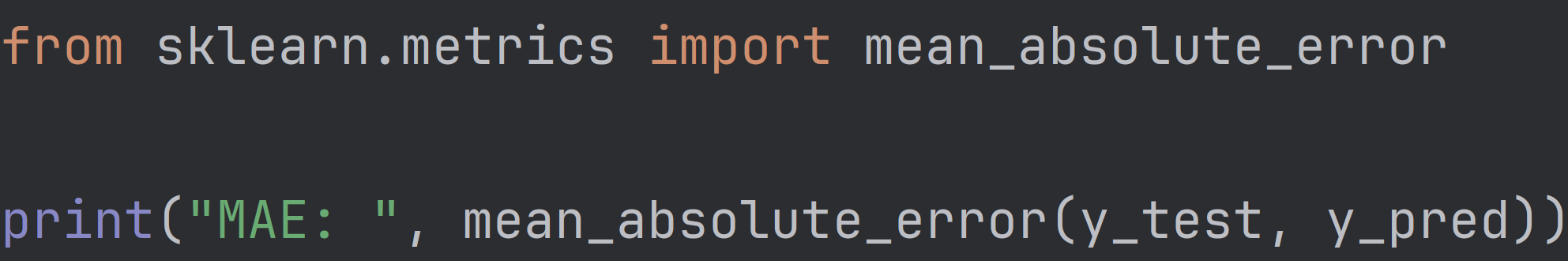
Proses machine learning dimulai dengan Analisis Data Eksploratif (EDA) untuk memahami data melalui statistik deskriptif dan visualisasi. Selanjutnya, dilakukan Penskalaan dan Normalisasi pada fitur-fitur tertentu menggunakan StandardScaler untuk meningkatkan performa model. Kemudian, dilakukan Seleksi Fitur dengan mengubah variabel target 'is\_legendary' menggunakan LabelEncoder. Setelah itu, data dibagi menjadi set pelatihan dan set pengujian menggunakan train\_test\_split. Model Regresi Linier kemudian dilatih menggunakan data pelatihan dan digunakan untuk memprediksi data pengujian. Akhirnya, evaluasi performa model dilakukan dengan menghitung skor R2, Mean Squared Error (MSE), dan Mean Absolute Error (MAE).

Dataset pokemon.csv berisi informasi tentang Pokemon, termasuk statistik dasar mereka, kinerja terhadap tipe lain, tinggi, berat, klasifikasi, langkah telur, poin pengalaman, kemampuan, dll. Tugasnya adalah memprediksi apakah Pokemon adalah legendaris berdasarkan fitur-fitur ini.

Step Pengerjaan

1. EDA
   1. 

* 1. 
  2. 
  3. 
  4. 
  5. 
  6. 
  7. 
  8. 
  9. 
  10. 
  11. 
  12. 
  13. 
  14. 
  15. 
  16. 
  17. 

1. Scaling
   1. 
   2. 
2. Feature
   1. 
   2. 
3. Training
   1. 
4. Cross Validation
   1. 
5. Performance evalution
   1. 
   2. 
   3. 
   4. 
   5. 
   6. 