Logistic Regression untuk Klasifikasi

Logistic Regression adalah salah satu algoritma machine learning paling populer untuk masalah klasifikasi. Dalam klasifikasi, tujuannya adalah memprediksi kategori atau kelas dari suatu data berdasarkan fitur-fitur yang diberikan. Misalnya, dalam kasus dataset Titanic, kita ingin memprediksi apakah seorang penumpang akan selama (Survived = 1) atau tidak (Survived = 0).

# Mengapa Logistic Regression?

Logistic Regression memiliki beberapa keunggulan, antara lain:

* Mudah Diinterpretasikan: Model menghasilkan probabilitas untuk setiap kelas, sehingga mudah untuk memahami bagaimana model membuat keputusan.
* Efisien: Algoritma cukup cepat untuk dilatih, bahkan pada dataset yang besar.
* Cocok untuk Masalah Klasifikasi Biner: Sangat baik untuk masalah dengan dua kelas (seperti hidup/mati, ya/tidak).

# Cara Kerja Logistic Regression

Secara sederhana, Logistic Regression bekerja dengan cara menghitung probabilitas suatu data termasuk dalam suatu kelas, lalu menggunakan *threshold* untuk memutuskan kelas mana yang akan diberikan. Misalnya, jika probabilitas selamat (Survived = 1) lebih besar dari 0.5, maka model akan memprediksi penumpang tersebut selamat.

# Penerapan Logistic Regression pada Dataset Titanic

Mari kita terapkan Logistic Regression untuk memprediksi apakah seorang penumpang Titanic akan selamat berdasarkan fitur-fitur yang ada.

## Tugas Hari 9:

1. Pembuatan Model:

* Gunakan LogisticRegression dari scikit-learn untuk membuat model Logistic Regression.
* Latih model menggunakan X\_train dan y\_train yang sudah di buat sebelumnya.

1. Prediksi:

* Gunakan model untuk membuat prediksi pada data uji (X\_train).

1. Evaluasi:

* Hitung akurasi model pada data uji.

## Contoh Kode (Scikit-Learn):

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import accuracy\_score

# ... (kode untuk membaca dan mengolah dataset Titanic, serta membagi data latih dan data uji)

# Membuat model Logistic Regression

model = LogisticRegression(random\_state=42)

# Melatih model

model.fit(X\_train, y\_train)

# Prediksi pada data uji

y\_pred = model.predict(X\_test)

# Evaluasi akurasi

accuracy = accuracy\_score(y\_test, y\_pred)

print(f'Akurasi model: {accuracy:.2f}')

**Catatan:**

* Pastikan telah menginstall library scikit-learn.
* Mencoba berbagai fitur (kolom) dari dataset Titanic sebagai input untuk model.
* Selain akurasi, dapat juga mengevaluasi model menggunakan metrik lain seperti precision, recall, dan F1-score.

**Selamat Mengerjakan Tugas! 😂**

### Tugas:

1. Pembuatan Model:

# Membuat Model Logistic Regression

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import accuracy\_score

model = LogisticRegression(random\_state=42)

# Melatih Model

model.fit(X\_train, y\_train)

1. Prediksi:

# Prediksi pada data uji

y\_pred = model.predict(X\_test)

1. Evaluasi:

# Evaluasi Akurasi

accuracy = accuracy\_score(y\_test, y\_pred)

print(f'Akurasi model: {accuracy:.2f}')

print(len(y\_test)) # Harus mencetak 524

print(len(y\_pred)) # Harus mencetak 524

Dengan Hasil:

Akurasi model: 0.85

524

524