Evaluasi Model Klasifikasi

Setelah membangun model klasifikasi seperti Logistic Regression, penting untuk mengevaluasi kinerjanya. Evaluasi model membantu kita memahami seberapa baik model dapat membuat prediksi pada data yang belum pernah dilihatnya sebelumnya (data uji).

# Metrik Evaluasi Klasifikasi

Ada beberapa metrik yang umum digunakan untuk mengevaluasi model klasifikasi:

1. Akurasi (Accuracy): Proporsi prediksi yang benar dari semua prediksi.
2. Presisi (Precision): Proporsi prediksi positif yang benar dari semua prediksi positif.
3. Recall (Sensitivity): Proporsi prediksi positif yang benar dari semua data yang sebenarnya positif.
4. F1-Score: Rata-rata harmonik antara presisi dan recall.
5. Confusion Matrix: Tabel yang menunjukkan jumlah prediksi benar dan salah untuk setiap kelas.

# Memilih Metrik yang Tepat

Pemilihan metrik evaluasi yang tepat tepat tergantung pada tujuan dan karakteristik masalah klasifikasi Anda.

* Jika Anda hanya ingin mengetahui seberapa sering model membuat prediksi yang benar secara keseluruhan, akurasi mungkin sudah cukup.
* Jika Anda lebih peduli tentang menghindari kesalahan positif (misalnya, dalam diagnosis medis), presisi mungkin lebih penting.
* Jika Anda lebih peduli tentang menghindari kesalahan negatif (misalnya, dalam deteksi spam), recall mungkin lebih penting.
* F1-Score adalah metrik yang baik untuk digunakan jika Anda ingin menyeimbangkan antara presisi dan recall.

# Confusion Matrix

Confusion matrix adalah alat yang sangat berguna untuk memahami kinerja model klasifikasi secara lebih detail. Confusion matrix menunjukkan jumlah prediksi benar dan salah untuk setiap kelas.

| Prediksi  Positif Negatif  Aktual Positif TP FN  Aktual Negatif FP TN |
| --- |

* **TP (True Positive):** Prediksi positif yang benar.
* **TN (True Negative):** Prediksi negatif yang benar.
* **FP (False Positive):** Prediksi positif yang salah.
* **FN (False Negative):** Prediksi negatif yang salah.

## Contoh Kode (Scikit-Learn):

from sklearn.metrics import accuracy\_score, precision\_score, recall\_score, f1\_score, confusion\_matrix

# ... (kode untuk membaca, mengolah dataset, membagi data, dan membuat model)

# Prediksi pada data uji

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

# Hitung metrik evaluasi

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

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

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

f1 = f1\_score(y\_test, y\_pred)

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

print(f'Presisi: {precision:.2f}')

print(f'Recall: {recall:.2f}')

print(f'F1-Score: {f1:.2f}')

# Confusion matrix

cm = confusion\_matrix(y\_test, y\_pred)

print('\nConfusion Matrix:')

print(cm)

### Tugas Hari 10:

Lanjutkan dengan model Logistic Regression yang telah dibuat untuk dataset Titanic.

### Tugas:

1. Hitung Metrik Evaluasi:

* Hitung akurasi, presisi, recall, dan F1-score dari model pada data uji.

1. Buat Confusion Matrix:

* Buat confusion matrix untuk model pada data uji.
* Tampilkan confusion matrix dalam bentuk tabel.

1. Analisis:

* Interpretasikan hasil metrik evaluasi dan confusion matrix. Bagaimana kinerja model Anda? Apakah ada kelas yang lebih sulit diprediksi daripada yang lain?

Selamat Mengerjakan Tugas! 😀

**Tugas:**

1. **Hitung Metrik Evaluasi:**

from sklearn.metrics import accuracy\_score, precision\_score, recall\_score, f1\_score, confusion\_matrix

# Prediksi pada data uji

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

# Prediks metrik evaluasi

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

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

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

f1 = f1\_score(y\_test, y\_pred)

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

print(f'Presisi: {precision:.2f}')

print(f'Recall: {recall:.2f}')

print(f'F1-Score: {f1:.2f}')

Output:

Akurasi: 0.85

Presisi: 0.82

Recall: 0.59

F1-Score: 0.68

1. **Buat Confusion Matrix:**

# Confusion matrix

cm = confusion\_matrix(y\_test, y\_pred)

print('\nConfusion Matrix: ')

print(cm)

Output:

Confusion Matrix:

[[362 19]

[ 59 84]]

1. **Analisis:**

# Analisis

print("\nAnalisis:")

print(f"- Akurasi model cukup baik (85%), tetapi masih ada ruang untuk perbaikan.")

print(f"- Presisi (82%) menunjukkan bahwa dari semua penumpang yang diprediksi selamat, sebagian kecil sebenarnya tidak selamat.")

print(f"- Recall (59%) menunjukkan bahwa model cukup baik dalam mengidentifikasi penumpang yang benar-benar selamat.")

print(f"- Confusion matrix menunjukkan bahwa model lebih sering melakukan kesalahan false positive (19) daripada false negative (59). Ini mungkin berarti model terlalu optimis dalam memprediksi kelangsungan hidup.")