



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : Metrik Evaluasi
Halaman : 1/4

NIM	222410101087
Nama	Rafi Jauhari
Kelas	A
Program Studi	Sistem Informasi
Asisten	1. Renata Sayidatul Arikha 212410101057 2. Aprodhita Nanda Eka Wijaya 212410101071

LANGKAH KERJA

1. Ulangi membuat prediksi seperti di kelas praktikum, namun gunakan studi kasus pada dataset yang ada di PPT. (Modelnya bisa disesuaikan dengan dataset)

```
import pandas as pd
from sklearn.metrics import accuracy_score, precision_score, recall_score,
f1_score, confusion_matrix
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier

data = {
    'Status Sebenarnya': ['Tidak DO', 'Tidak DO', 'Tidak DO', 'Tidak DO',
'Tidak DO', 'DO', 'DO', 'DO', 'DO', 'DO'],
    'Hasil Prediksi': ['Tidak DO', 'Tidak DO', 'Tidak DO', 'DO', 'DO', 'Tidak
DO', 'DO', 'DO', 'DO', 'DO']
}

df = pd.DataFrame(data)

#Pre Processing
labelencoder = LabelEncoder()
df['Status Sebenarnya'] = labelencoder.fit_transform(df['Status Sebenarnya'])
df['Hasil Prediksi'] = labelencoder.fit_transform(df['Hasil Prediksi'])
```



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : Metrik Evaluasi
Halaman : 2/4

```
#Model Testing
X = df[['Hasil Prediksi']]
y = df['Status Sebenarnya']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=42)

clf = DecisionTreeClassifier()
clf.fit(X_train, y_train)

y_pred = clf.predict(X_test)

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)
conf_matrix = confusion_matrix(y_test, y_pred)

print("Accuracy:", accuracy)
print("Precision:", precision)
print("Recall:", recall)
print("F1 Score:", f1)
print("Confusion Matrix:\n", conf_matrix)
```

Jawaban :



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : Metrik Evaluasi
Halaman : 3/4

```
Accuracy: 0.6666666666666666  
Precision: 0.5  
Recall: 1.0  
F1 Score: 0.6666666666666666  
Confusion Matrix:  
[[1 1]  
[0 1]]
```

2. Jelaskan kapan lebih cocok menggunakan Akurasi, kapan lebih cocok menggunakan Recall & F1 Score untuk studi kasus tersebut.

Akurasi cocok digunakan saat distribusi kelas dalam data seimbang, yaitu jumlah mahasiswa yang DO dan Tidak DO relatif sama, dan ketika kesalahan dalam memprediksi kedua kelas memiliki tingkat kepentingan yang sama.

Recall penting digunakan ketika data tidak seimbang dan fokus utama adalah mendeteksi kasus DO, terutama jika kesalahan melewati kasus DO (False Negative) dianggap lebih serius daripada kesalahan memprediksi DO pada mahasiswa yang sebenarnya Tidak DO (False Positive).

F1 Score berguna saat data tidak seimbang karena merupakan kombinasi harmonis dari precision dan recall, memberikan gambaran yang lebih menyeluruh tentang kinerja model dalam mendeteksi kelas minoritas (DO). Jika dibutuhkan keseimbangan antara menangkap sebanyak mungkin kasus DO (recall) dan memastikan bahwa prediksi DO akurat (precision), maka F1 Score adalah metrik yang tepat.

HASIL DAN ANALISIS DATA

Model memiliki performa yang cukup baik dalam mendeteksi kasus DO, dengan presisi, recall, dan F1 Score sekitar 0.67. Akurasi model sebesar 0.6 menunjukkan kemampuan model dalam memprediksi dengan benar untuk 60% kasus. Confusion matrix menunjukkan bahwa terdapat 1 True Negative, 1 True Positive, 0 False Negative, dan 1 False Positive. F1 Score sebesar 0.67 menunjukkan keseimbangan yang baik antara presisi dan recall dalam mendeteksi DO.

KESIMPULAN



Lembar Kerja Mahasiswa

Mata Kuliah : Data Mining
Bahasan : Metrik Evaluasi
Halaman : 4/4

Dalam studi kasus ini, model Classification Metrics yang dikembangkan menunjukkan performa yang baik dalam mendeteksi kasus DO, dengan hasil akurasi sebesar 0.6, presisi 0.67, recall 0.67, dan F1 Score 0.67. Walaupun akurasinya sedikit rendah, recall yang tinggi mengindikasikan bahwa model mampu mengidentifikasi sebagian besar kasus DO yang benar, sedangkan presisi yang cukup baik menandakan bahwa model jarang salah dalam memprediksi DO. Confusion matrix mengungkapkan bahwa model berhasil mengidentifikasi 1 True Negative, 1 True Positive, dan 1 False Positive, tanpa adanya False Negative. Hal ini menunjukkan bahwa model cenderung tidak melewatkan kasus DO yang sebenarnya, namun dapat salah memprediksi DO yang sebenarnya sebagai Tidak DO. Dalam konteks ini, F1 Score memberikan gambaran yang seimbang tentang kinerja model dalam mendeteksi DO.

Link Google Colab

<https://colab.research.google.com/drive/1ltToIL5DioqkhVOMDPk5fot4huNx0pt>

Link Youtube (Unlisted)

Jember,2024

Mengetahui,
Dosen Datamining

Asisten,

Fajrin Nurman Arifin, S.T., M.Eng
NIP. 198511282015041002

(Nama Jelas)
NIM.