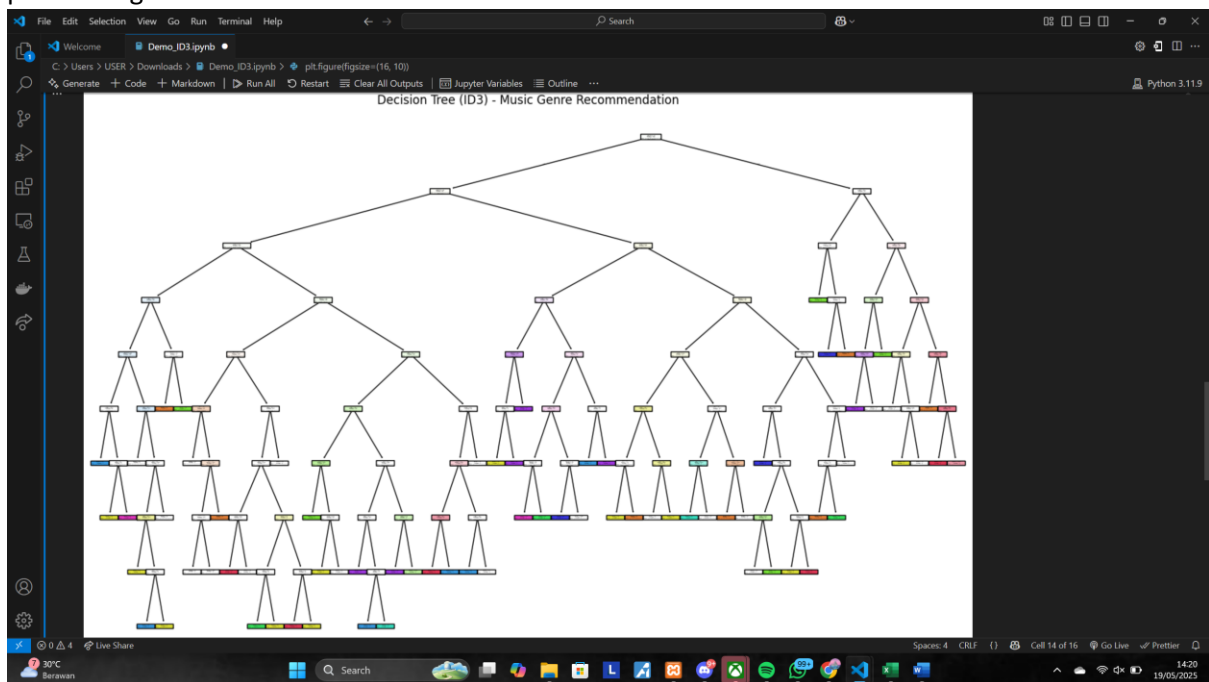


Nama : M Arifin Ilham

Nim: 1302223061

perbandingan 100 data dan 50 data



Contoh eksekusi inference

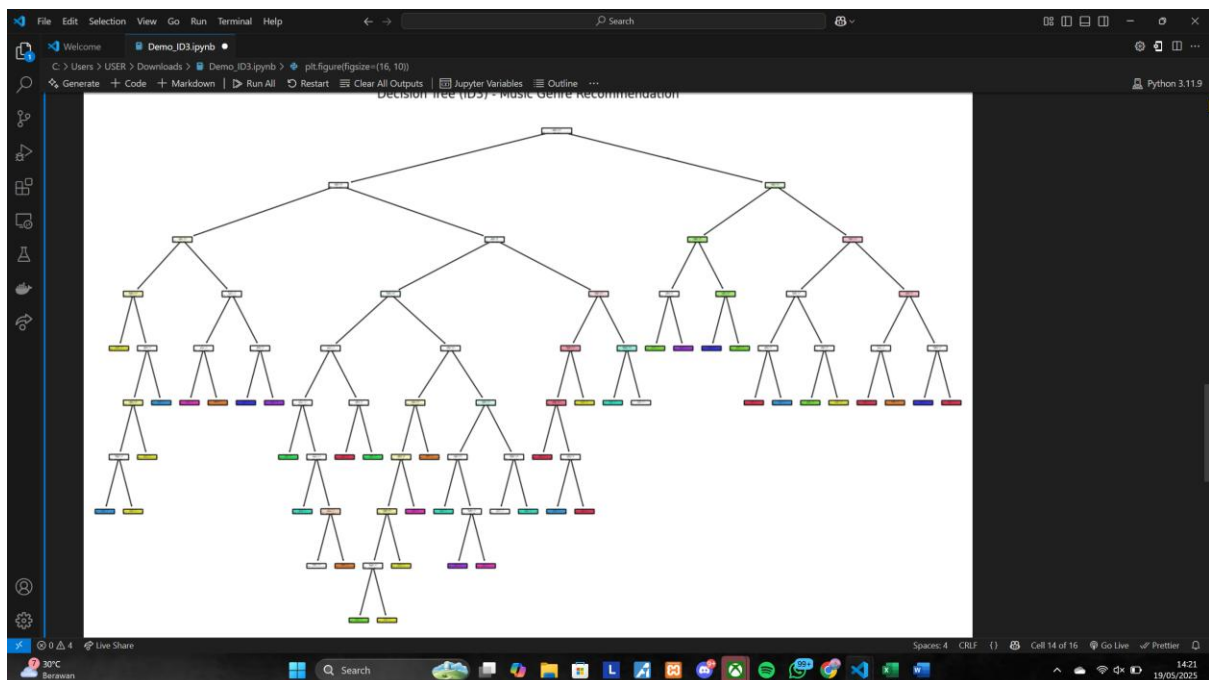
```
example = pd.DataFrame({
    "hari": ["Sabtu"],
    "cuaca": ["Cerah"],
    "aktifitas": ["Belajar"]
})

# Encode the example using the same label encoders
for column in example.columns:
    example[column] = label_encoders[column].transform(example[column])

# Predict and decode result
pred_encoded = clf.predict(example)
pred_genre = label_encoders["Genre"].inverse_transform(pred_encoded)
print(f"Recommended genre: {pred_genre[0]}")
```

[18] ✓ 0.0s

... Recommended genre: R&B



Contoh eksekusi inference

```
example = pd.DataFrame({
    "hari": ["Senin"],
    "cuaca": ["Hujan"],
    "aktivitas": ["Belajar"]
})

# Encode the example using the same label encoders
for column in example.columns:
    example[column] = label_encoders[column].transform(example[column])

# Predict and decode result
pred_encoded = clf.predict(example)
pred_genre = label_encoders["Genre"].inverse_transform(pred_encoded)
print(f"Recommended genre: {pred_genre[0]}")
```

[25] ✓ 0.0s

... Recommended genre: Rock

1. Ukuran Dataset

- **50 data** lebih cepat diproses dan dianalisis, cocok untuk uji coba awal (baseline).
- **100 data** memberikan **lebih banyak variasi**, sehingga hasil analisis atau prediksi bisa lebih akurat dan stabil.

2. Distribusi Variabel

Jika dilihat dari data sebelumnya (hari, cuaca, aktivitas, genre lagu):

- Dengan 100 data, distribusi kategori pada setiap variabel cenderung lebih **merata**.
- Dengan 50 data, beberapa kategori bisa **tidak muncul** atau terlalu sedikit, sehingga bisa mempengaruhi akurasi model (jika dipakai untuk machine learning).

3. Kesimpulan

- Untuk **analisis eksploratif awal**, 50 data cukup.
- Untuk **pelatihan model klasifikasi genre lagu**, 100 data lebih disarankan karena:
 - Distribusi kategori lebih lengkap.
 - Model memiliki lebih banyak pola untuk belajar.
 - Hasil evaluasi (akurasi, presisi, dll.) lebih stabil.