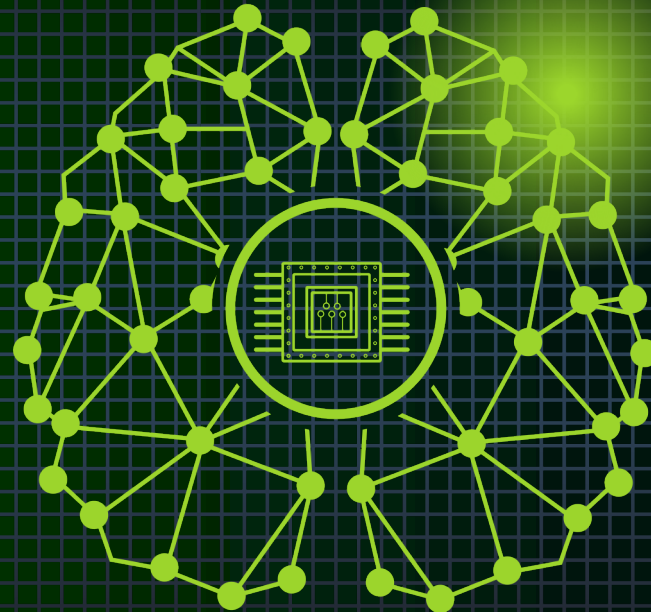


EIE6012

MACHINE LEARNING

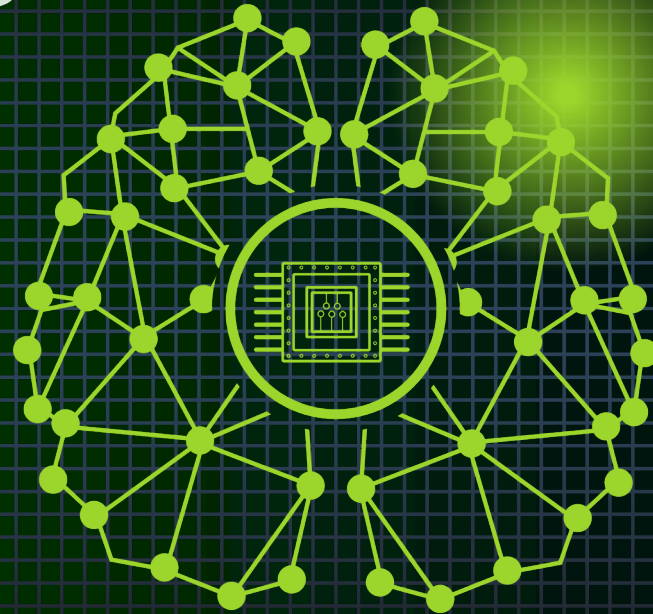
ENDANG SRI RAHAYU



3

Pre-Processing

- **Feature Selection**
- **Feature Extraction**



Tujuan Pembelajaran

Setelah mengikuti kuliah ini, mahasiswa diharapkan mampu:

- Memahami perbedaan antara feature selection dan feature extraction.
- Menggunakan Python untuk menunjukkan salah satu metode feature selection

Perbedaan Feature Selection vs Feature Extraction

Aspek	Feature Selection	Feature Extraction
Tujuan utama	Memilih fitur yang paling relevan dari data yang sudah ada	Menciptakan fitur baru dari kombinasi fitur asli
Apa yang dilakukan	Hanya memilih sebagian fitur yang penting dan menghapus yang tidak penting	Mengubah fitur asli menjadi fitur baru dengan representasi berbeda
Output	Subset dari fitur lama (misalnya dari 10 fitur → dipilih 5 terbaik)	Sekumpulan fitur baru (misalnya dari 10 fitur → jadi 3 komponen utama)
Contoh teknik	- Korelasi	

Contoh Ilustrasi Sederhana: Feature Selection

Umur	Gaji	Berat Badan	Warna Mobil	Tinggi Badan	Membeli Produk
25	5000	70	Merah	170	Ya
30	7000	80	Hitam	180	Tidak

Setelah analisis korelasi, Anda temukan bahwa hanya **Umur** dan **Gaji** berpengaruh pada keputusan membeli.

Maka Anda **memilih** fitur Umur dan Gaji saja

Feature Selection, pendekatan

“Tidak semua fitur penting – pilih yang paling berpengaruh terhadap target.”

Tujuan:

- Meningkatkan akurasi
- Mengurangi waktu pelatihan
- Menghindari overfitting

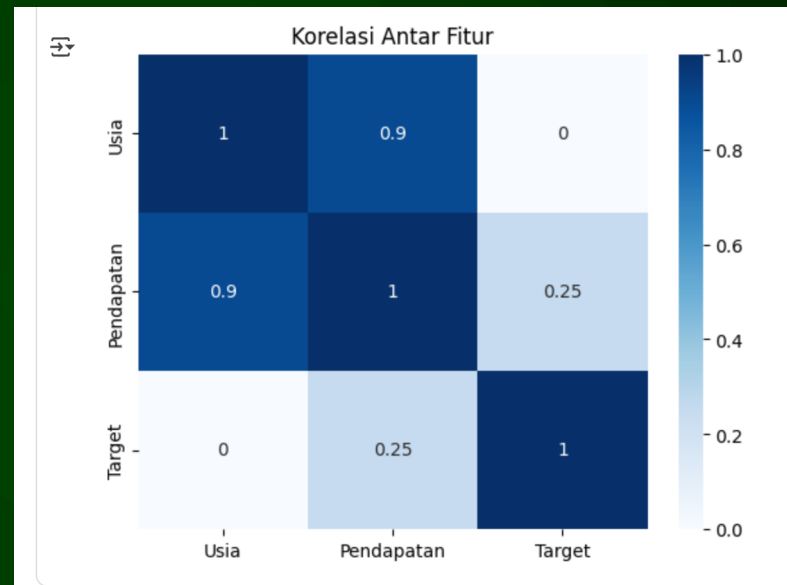
Jenis	Contoh	Penjelasan
Filter Method	Korelasi, Chi-square	Berdasarkan statistik antar fitur & target
Wrapper Method	Recursive Feature Elimination (RFE)	Menguji subset fitur menggunakan model
Embedded Method	Feature importance (Decision Tree)	Model memilih fitur saat training

(contoh) Filter Method (Korelasi)

```
import seaborn as sns
import matplotlib.pyplot as plt

# Tambah kolom target dummy
data['Target'] = [1, 0, 1, 0, 1]

# Korelasi antar fitur
corr = data.corr(numeric_only=True)
sns.heatmap(corr, annot=True, cmap='Blues')
plt.title("Korelasi Antar Fitur")
plt.show()
```



Contoh Ilustrasi Sederhana: Feature Extraction

Masih dataset yang sama. Anda ingin **mengurangi dimensi data** dengan *Principal Component Analysis (PCA)*.

PCA mengubah semua fitur (Umur, Gaji, Berat, Tinggi) menjadi **komponen baru**, misalnya:

Komponen_1	Komponen_2	Membeli Produk
0.95	0.12	Ya
0.88	-0.20	Tidak

Fitur baru (Komponen_1, Komponen_2) adalah **hasil kombinasi matematis** dari fitur asli

Kesimpulan

Aspek Penting

Menghapus fitur tidak relevan

Membuat fitur baru

Mudah diinterpretasi

Contoh algoritma

Tujuan utama

Feature Selection

✓ Ya

✗ Tidak

✓

RFE, Chi-square

Efisiensi & relevansi

Feature Extraction

✗ Tidak

✓ Ya

✗

PCA, Autoencoder

Reduksi dimensi & kompresi informasi

DISKUSI

Apa yang terjadi jika kita melatih model tanpa preprocessing?
Mengapa memilih fitur yang tepat bisa mempercepat proses training?

Gunakan dataset publik (misal *Iris Dataset* atau *Titanic Dataset*) dan lakukan:

- Feature selection (RFE atau korelasi)
- Jelaskan fitur mana yang paling relevan dan mengapa.

Kapan sebaiknya kita memilih feature selection, dan kapan feature extraction lebih cocok?

Bagaimana dampak jumlah fitur terhadap waktu training dan akurasi model?