

NAMA :Alief Fauzi Luqmanul Hakim

NIM : 110320129

- **PCA**

PCA (Principal Component Analysis) adalah sebuah metode dalam machine learning yang digunakan untuk analisis data multivariat. Tujuan utama dari PCA adalah untuk mengurangi dimensi data dengan tetap mempertahankan sebagian besar informasi yang terkandung dalam data asli. PCA mencoba untuk menemukan kombinasi linear baru dari variabel asli yang disebut komponen utama (principal components) sehingga komponen utama pertama menyimpan sebanyak mungkin variasi dalam data, komponen utama kedua menyimpan sisa variasi yang paling besar, dan seterusnya.

1. **Reduksi Dimensi:** PCA memungkinkan Anda untuk mengurangi dimensi data dengan menghilangkan komponen yang kurang penting. Ini dapat membantu dalam mengatasi masalah "kutukan dimensi" di mana jumlah fitur atau variabel sangat besar.
2. **Visualisasi Data:** Dengan mengurangi dimensi data ke 2 atau 3 komponen utama pertama, Anda dapat memvisualisasikan data dalam plot yang lebih mudah dipahami.
3. **Filtrasi Noise:** PCA dapat membantu dalam menghilangkan komponen yang memiliki sedikit variasi dan oleh karena itu dianggap sebagai noise dalam data.
4. **Kompresi Data:** Dalam beberapa kasus, PCA dapat digunakan untuk kompresi data dengan mempertahankan sebagian besar informasi penting.
5. **Pengenalan Pola:** PCA dapat membantu dalam mengidentifikasi pola dan hubungan antara variabel dalam data.

- **K-Nearest Neighbors**

K-Nearest Neighbors (K-NN) adalah salah satu algoritma dalam machine learning yang digunakan untuk masalah klasifikasi dan regresi. Ini adalah algoritma yang sangat sederhana dan intuitif yang dapat digunakan untuk mengklasifikasikan data atau memprediksi nilai berdasarkan kedekatan mereka dengan data pelatihan yang telah ada.

Cara kerja K-NN adalah sebagai berikut:

1. **Memilih nilai K:** K adalah jumlah tetangga terdekat yang akan digunakan untuk membuat prediksi. Nilai K ini dapat ditentukan sebelumnya oleh pengguna.
2. **Mengukur jarak:** K-NN menggunakan metrik jarak, seperti jarak Euclidean atau jarak Manhattan, untuk mengukur seberapa dekat titik data uji dengan titik data pelatihan yang ada. Ini mengukur kedekatan antara data yang akan diprediksi dengan data pelatihan yang telah ada.
3. **Menentukan K tetangga terdekat:** Algoritma ini akan mengidentifikasi K titik data pelatihan terdekat dengan data yang akan diprediksi berdasarkan metrik jarak yang digunakan.
4. **Klasifikasi (untuk masalah klasifikasi) atau regresi (untuk masalah regresi):** Jika masalahnya adalah klasifikasi, maka K-NN akan melakukan pemungutan suara mayoritas dari K tetangga terdekat untuk menentukan kelas mana yang akan diberikan kepada data yang akan diprediksi. Jika masalahnya adalah regresi, K-NN akan mengambil rata-rata nilai dari K tetangga terdekat sebagai prediksi nilai kontinu.

K-NN adalah salah satu algoritma yang sangat bergantung pada data pelatihan dan pemilihan parameter K yang tepat. K-NN cenderung berkinerja baik jika data Anda memiliki cluster yang jelas dan jika Anda memilih nilai K yang sesuai. Namun, algoritma ini juga bisa rentan terhadap data pencilan (outliers) dan membutuhkan penghitungan jarak yang dapat menjadi mahal komputasinya jika Anda memiliki banyak data pelatihan.

- **Decision tree**

Decision tree (pohon keputusan) dan classification tree (pohon klasifikasi) adalah dua konsep penting dalam machine learning, khususnya dalam konteks supervised learning. Berikut penjelasan singkat tentang keduanya:

1. Decision Tree (Pohon Keputusan)

- Decision tree adalah model prediktif yang digunakan dalam machine learning untuk mengambil keputusan atau membuat prediksi dengan cara yang mirip dengan mengambil keputusan berdasarkan serangkaian pertanyaan dan aturan.
- Model ini mewakili data sebagai pohon dengan node internal yang merepresentasikan pengujian atribut, cabang yang merepresentasikan hasil dari pengujian ini, dan daun yang merepresentasikan label klasifikasi.
- Decision tree berfungsi dengan membagi data ke dalam subset yang semakin kecil dan homogen berdasarkan atribut-atribut yang ada, sehingga akhirnya mencapai prediksi atau keputusan yang tepat.

2. Classification Tree (Pohon Klasifikasi)

- Classification tree adalah bentuk khusus dari decision tree yang digunakan untuk melakukan klasifikasi, yaitu mengelompokkan instance data ke dalam kelas atau kategori yang berbeda berdasarkan atribut-atribut tertentu.
- Dalam pohon klasifikasi, setiap daun dari pohon mewakili kelas atau kategori yang berbeda, sedangkan setiap node internal mewakili aturan atau keputusan yang memisahkan instance data ke dalam subset yang berbeda.
- Classification tree sangat umum digunakan dalam tugas-tugas seperti klasifikasi email spam, diagnosis medis, atau pengklasifikasian objek dalam pengolahan citra.

Decision tree dan classification tree adalah alat yang berguna dalam machine learning karena mereka mudah diinterpretasikan dan dapat digunakan dalam berbagai aplikasi. Mereka membantu mengambil keputusan atau membuat prediksi berdasarkan aturan-aturan yang jelas dan dapat diuraikan.