

Nama : Zaky Ibnu Kusumah

NIM : 1103204213

Kelas : TK4404

Tugas PDF

Principal Component Analysis

Principal Component Analysis (PCA) adalah sebuah teknik statistik yang digunakan dalam analisis data untuk mengurangi dimensi dari data yang kompleks. PCA bertujuan untuk mengidentifikasi dan mengurutkan komponen utama dari data, sehingga memungkinkan representasi data yang lebih sederhana namun tetap mempertahankan sebagian besar variabilitas informasi. PCA bekerja dengan mentransformasi data asli ke dalam sistem koordinat baru yang disusun berdasarkan komponen utama, di mana komponen pertama adalah yang paling penting dalam menjelaskan variasi data. Teknik ini sering digunakan dalam berbagai bidang seperti analisis data, pengolahan citra, dan pengenalan pola untuk membantu mengidentifikasi pola-pola yang relevan dan menghilangkan kelebihan informasi yang tidak diperlukan.

K-Nearest Neighbors

K-Nearest Neighbors (K-NN) adalah salah satu algoritma machine learning yang digunakan untuk klasifikasi dan regresi. Konsep dasar K-NN adalah bahwa objek yang serupa cenderung berada dekat satu sama lain dalam ruang fitur. Dalam K-NN, setiap data point direpresentasikan sebagai titik dalam ruang fitur multidimensi, dan ketika kita ingin melakukan prediksi, algoritma ini mencari sejumlah K tetangga terdekat dari data point yang ingin diprediksi. Prediksi kemudian dibuat berdasarkan mayoritas kelas dari tetangga-tetangga ini dalam kasus klasifikasi, atau dengan menghitung rata-rata nilai dalam kasus regresi. K-NN memiliki fleksibilitas yang tinggi dan relatif mudah untuk diimplementasikan, tetapi kelemahannya termasuk sensitif terhadap skala fitur dan memerlukan perhitungan jarak antara data points, yang bisa menjadi komputasi yang mahal jika datasetnya besar.

Decision Tree

Decision Tree, atau pohon keputusan, adalah algoritma machine learning yang digunakan untuk pengambilan keputusan dan klasifikasi data. Algoritma ini mewakili struktur hirarki yang mirip dengan pohon, di mana setiap simpul dalam pohon mewakili keputusan berdasarkan atribut data. Pada setiap simpul, algoritma memilih atribut yang paling baik memisahkan data menjadi dua atau lebih kelompok yang paling homogen. Proses ini berlanjut hingga mencapai simpul daun (leaf node), di mana keputusan akhir atau label kelas ditentukan.