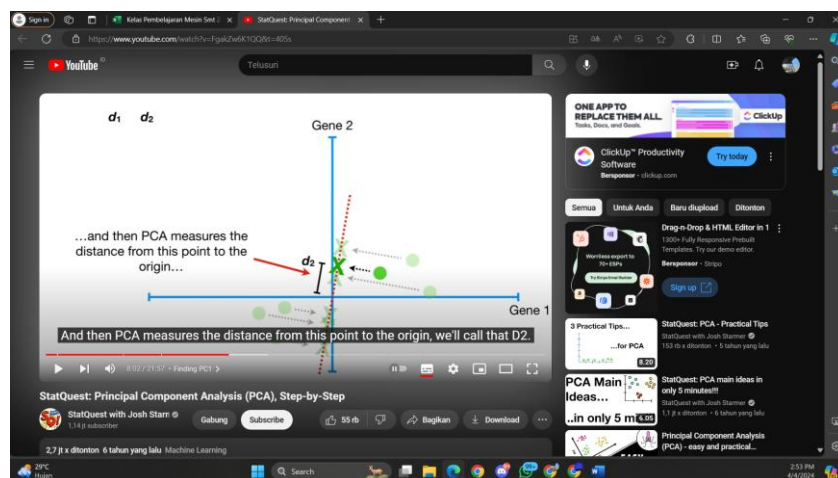


Dheandra Alfarrelwijaya 1103213048

(Principal Component Analysis/PCA) adalah sebuah metode statistik yang digunakan untuk mengurangi dimensi dari kumpulan data yang kompleks, sambil mempertahankan sebagian besar informasi yang terkandung dalam data tersebut. Tujuan utamanya adalah untuk mengidentifikasi pola-pola dalam data dengan mereduksi jumlah dimensi (variabel) yang digunakan untuk menggambarkan data tersebut, sehingga memudahkan pemahaman dan interpretasi. PCA dapat dijelaskan secara konseptual dengan menguraikan langkah-langkahnya menggunakan dekomposisi nilai singular (SVD). SVD adalah teknik matematis yang mendasari PCA untuk memecah data menjadi komponen-komponen utama.

PCA mencari transformasi linier dari data asli ke ruang dimensi yang lebih rendah, di mana setiap dimensi baru (komponen) adalah kombinasi linear dari dimensi-dimensi asli yang ada dalam data. Komponen-komponen ini diurutkan berdasarkan jumlah variasi dalam data yang dapat dijelaskan oleh masing-masing komponen, sehingga komponen pertama mengandung variasi yang paling besar, diikuti oleh komponen kedua, dan seterusnya. Jika kita mempertimbangkan data yang hanya memiliki dua gen, kita dapat memplotnya pada grafik dua dimensi. Gen 1 akan menjadi sumbu x dan Gen 2 akan menjadi sumbu y. Titik-titik pada grafik mewakili sampel-sampel yang diukur. Dalam konteks ini, PCA akan membantu kita memahami gen mana yang paling penting dalam mengelompokkan data. Misalnya, PCA dapat menunjukkan bahwa Gen 3 berperan penting dalam memisahkan sampel di sepanjang sumbu x

Berdasarkan video tersebut yang dapat disimpulkan bagaimana PCA melakukan Langkah demi Langkah oleh StatQuest. PCA digunakan untuk mendapat wawasan yang lebih dalam tentang suatu data melalui SVD. PCA membantu mengelompokkan data ke dalam komponen utama yang menjelaskan variasi data dengan baik



Berikut Langkah-langkah PCA:

- **Memusatkan Data:** Langkah awal PCA adalah memusatkan data pada titik asal grafik, dengan membuat nilai rata-rata dari setiap variabel menjadi nol.
- **Mencari Komponen Utama (PC1):** PCA mencari garis terbaik untuk memproyeksikan data. Garis ini disebut komponen utama pertama (PC1), yang mencerminkan arah dengan variasi terbesar dalam data. Proyeksi data ke PC1 membantu mengurangi dimensi data.
- **Iterasi untuk Komponen Utama Berikutnya:** Proses ini dilakukan secara iteratif untuk menemukan komponen utama berikutnya (PC2, PC3, dst.), yang tegak lurus satu sama lain dan memperjelas pola-pola yang lebih kompleks dalam data.
- **Penilaian Kualitas dengan Nilai Eigen:** Penilaian kualitas komponen utama dapat dilakukan dengan menggunakan nilai eigen. Nilai eigen mewakili jumlah kuadrat jarak antara titik proyeksi dan titik asal untuk setiap komponen utama. Semakin besar nilai eigen, semakin banyak variasi yang dijelaskan oleh komponen utama tersebut.
- **Visualisasi dengan Plot PCA:** Memahami konsep PCA memungkinkan pembuatan plot PCA dua dimensi, yang membantu memvisualisasikan pola-pola yang mungkin sulit dilihat dalam data asli.
- **Proses Serupa untuk Dimensi Lebih Tinggi:** PCA juga dapat dilakukan untuk data tiga dimensi atau lebih, dengan menggunakan PC1, PC2, PC3, dan seterusnya untuk mereduksi dimensi dan memahami struktur dalam data.