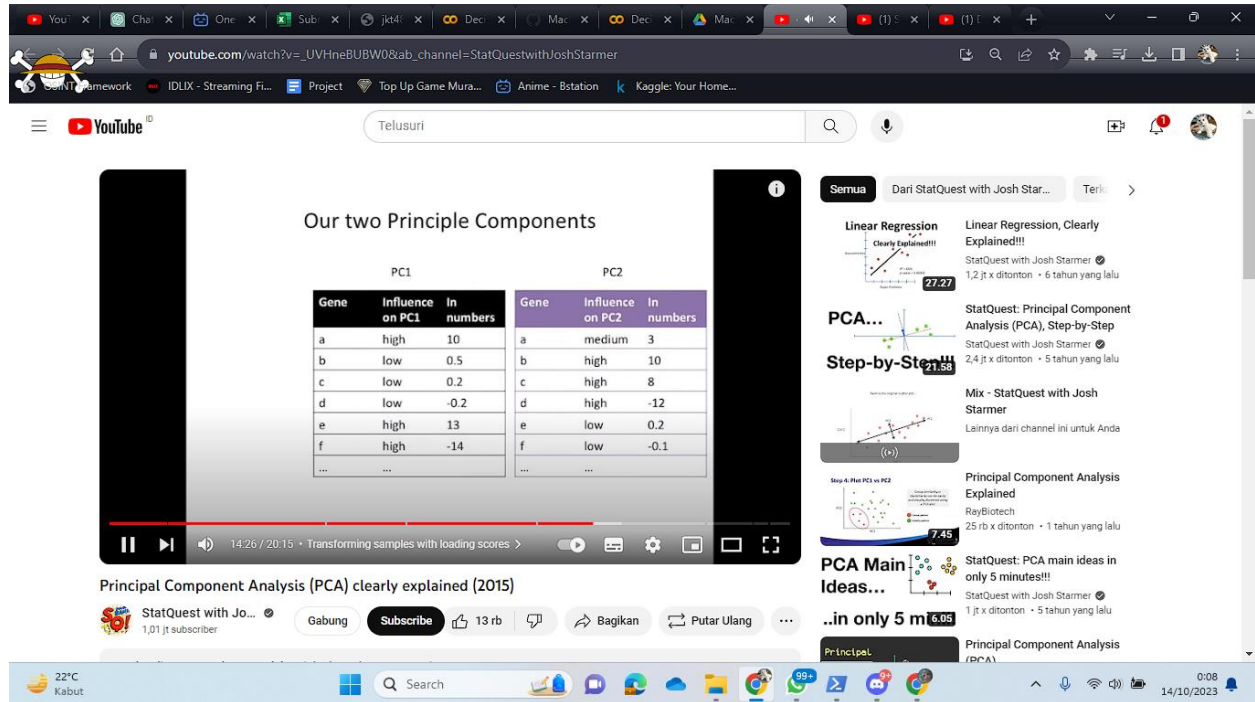


Nama : Nadhifi Qurrunul B F H

Nim : 1103204156

Principal Component Analysis (PCA) clearly explained



The video displays a table titled "Our two Principle Components" showing the influence of six genes on the first two principal components (PC1 and PC2). The table is organized into two columns: PC1 and PC2. Each column has a sub-header "Gene" and "In numbers".

PC1			PC2		
Gene	Influence on PC1	In numbers	Gene	Influence on PC2	In numbers
a	high	10	a	medium	3
b	low	0.5	b	high	10
c	low	0.2	c	high	8
d	low	-0.2	d	high	-12
e	high	13	e	low	0.2
f	high	-14	f	low	-0.1

PCA adalah metode statistik yang digunakan untuk mengurangi dimensi data dengan mentransformasikan data asli ke sistem koordinat baru. Dalam sistem koordinat baru ini, variabilitas data ditempatkan pada komponen utama atau "principal components," yang merupakan vektor-vektor orthogonal (tidak berkorelasi) yang menggambarkan variasi tertinggi dalam data. PCA mengidentifikasi komponen-komponen ini dengan memproyeksikan data ke dalam ruang vektor yang lebih kecil.

Tujuan dari PCA:

- **Reduksi Dimensi:** PCA membantu mengurangi dimensi data dengan menghapus komponen-komponen yang memiliki variasi rendah. Ini bermanfaat untuk mengatasi "kutub curse" (curse of dimensionality) dan menghemat sumber daya komputasi.
- **Visualisasi Data:** PCA membantu mengubah data ke dalam bentuk yang lebih mudah divisualisasikan. Dengan mengurangi dimensi, Anda dapat memvisualisasikan data dalam ruang dua atau tiga dimensi.

- Menghilangkan Korelasi: PCA memproyeksikan data ke dalam sistem koordinat baru yang tidak berkorelasi. Ini berguna untuk mengatasi masalah multikolinearitas di mana variabel-variabel dalam data asli berkorelasi kuat.
- Mengidentifikasi Pola dan Struktur: PCA membantu mengidentifikasi pola atau struktur dalam data dengan menyoroti komponen-komponen utama yang memiliki variasi tertinggi.
- Ekstraksi Informasi yang Signifikan: PCA memungkinkan Anda untuk mempertahankan sebagian besar informasi penting dalam data dengan hanya mempertimbangkan komponen-komponen utama.
- Pemrosesan Sinyal dan Kompresi Gambar: PCA digunakan dalam berbagai aplikasi pemrosesan sinyal dan kompresi gambar, di mana informasi yang kurang signifikan dapat dihilangkan tanpa mengorbankan kualitas.
- Deteksi Anomali: PCA dapat digunakan untuk mendeteksi anomali atau data yang jauh dari pola umum dengan memeriksa deviasi dari komponen utama.

K-nearest neighbors, Clearly Explained

The screenshot displays a YouTube video player interface. The main video is titled "StatQuest: K-nearest neighbors, Clearly Explained" by the channel "StatQuest with Josh Starmer". The video content shows a heatmap with a vertical line and text: "If our new cell ended up in the middle of the light blue cluster and...". The video player interface includes a search bar, a list of recommended videos on the right, and a Windows taskbar at the bottom.

Didalam video tersebut menjelaskan metode K-NN dalam machine learning secara rinci dan jelas. K-NN adalah algoritma yang digunakan untuk masalah klasifikasi dan regresi. Ini bekerja dengan cara mencari

K titik data terdekat dengan titik data yang akan diprediksi, dan kemudian melakukan mayoritas pemungutan suara (dalam klasifikasi) atau menghitung rata-rata (dalam regresi) dari tetangga terdekat untuk memprediksi hasilnya. StatQuest tentang K-NN menjelaskan konsep dasar K-NN, seperti cara menentukan jumlah K (jumlah tetangga terdekat yang akan diperhitungkan), penggunaan metrik jarak (misalnya Euclidean distance), dan bagaimana algoritma ini bekerja dalam praktek.

Decision and Classification Trees, Clearly Explained

The screenshot shows a YouTube video player with the title "Decision and Classification Trees, Clearly Explained" by StatQuest with Josh Starmer. The video content includes a table of data points and a decision tree diagram.

Loves Popcorn	Loves Soda	Age	Loves Cool As Ice
Yes	Yes	7	No
Yes	No	12	No
No	Yes	18	Yes
No	Yes	35	Yes
Yes	Yes	38	Yes
Yes	No	50	No
No	No	83	No

The decision tree diagram shows a root node "Soda" with a "True" branch leading to "Cool As Ice" (Yes/No) and a "False" branch leading to "Cool As Ice" (Yes/No). The "True" branch further splits into "Popcorn" and "Age < ???". The "False" branch further splits into "Cool As Ice" (Yes/No) and "Age < ???".

So let's see if we can reduce the Impurity by splitting the people that Love Soda based on Loves Popcorn or Age.

Pohon keputusan adalah model prediksi yang digunakan dalam machine learning untuk mengambil keputusan berdasarkan serangkaian aturan atau pertanyaan yang diajukan tentang data input. Pohon ini memiliki simpul yang mewakili aturan dan cabang yang mewakili pilihan atau keputusan. Dalam konteks klasifikasi, pohon keputusan digunakan untuk memisahkan data ke dalam kelas-kelas yang berbeda berdasarkan fitur-fitur input. Ini adalah algoritma pembelajaran yang sangat interpretable karena Anda dapat melihat langkah-langkah keputusan yang diambil oleh model.

Pohon klasifikasi adalah jenis pohon keputusan yang digunakan untuk masalah klasifikasi. Dalam konteks ini, pohon klasifikasi digunakan untuk mengkategorikan data input ke dalam kelas-kelas yang berbeda. Setiap simpul dalam pohon klasifikasi mewakili aturan atau pertanyaan yang digunakan untuk

memisahkan data menjadi kelompok kelas yang berbeda. Dengan menjawab serangkaian pertanyaan ini, data dapat diklasifikasikan ke dalam kelas yang sesuai.

Tujuan dari "Decision and Classification Trees, Clearly Explained!!!" adalah untuk menjelaskan konsep dasar pohon keputusan dan pohon klasifikasi dengan cara yang mudah dimengerti, termasuk bagaimana pohon tersebut dibangun, bagaimana aturan-aturan dipilih, dan bagaimana pohon ini digunakan dalam prediksi. Video atau materi tersebut mungkin akan menggunakan ilustrasi, contoh kasus, atau analogi sederhana untuk membuat konsep ini lebih mudah dipahami oleh pemirsa yang mungkin baru mengenalnya.