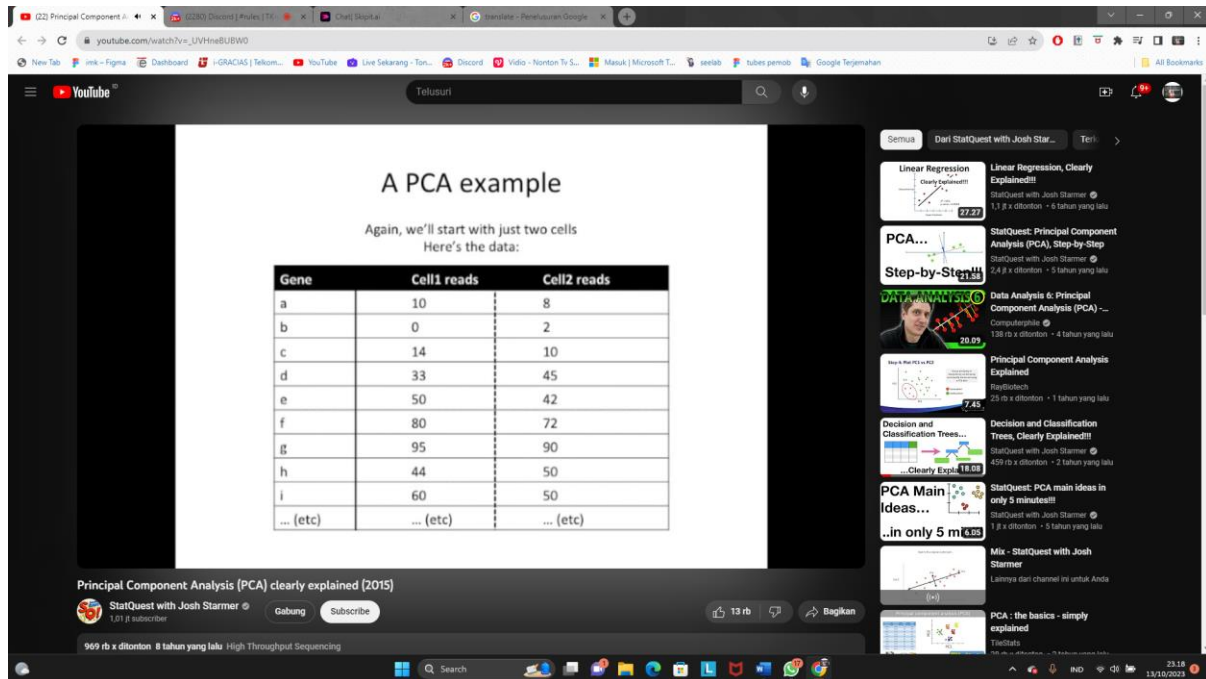


Nama : Fahmi Adhiwangsa

NIM : 1103204142

Kelas : TK4401

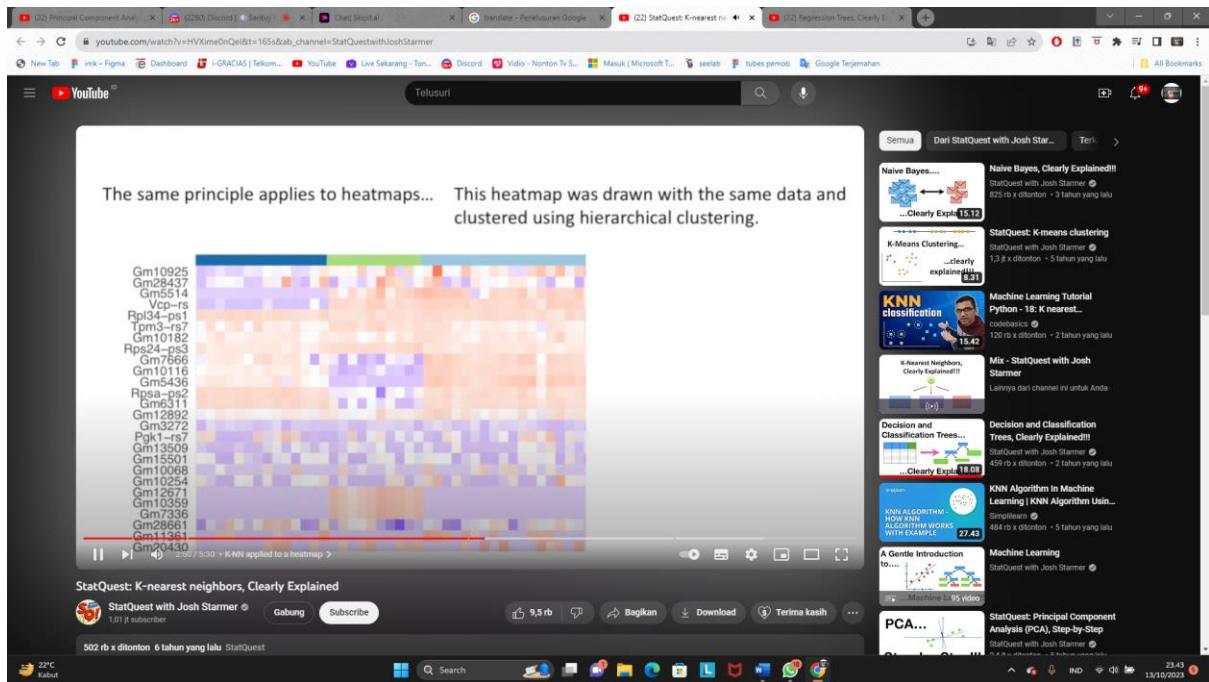


The screenshot shows a YouTube video player with the title "A PCA example". The video content displays a table of gene expression data with the following structure:

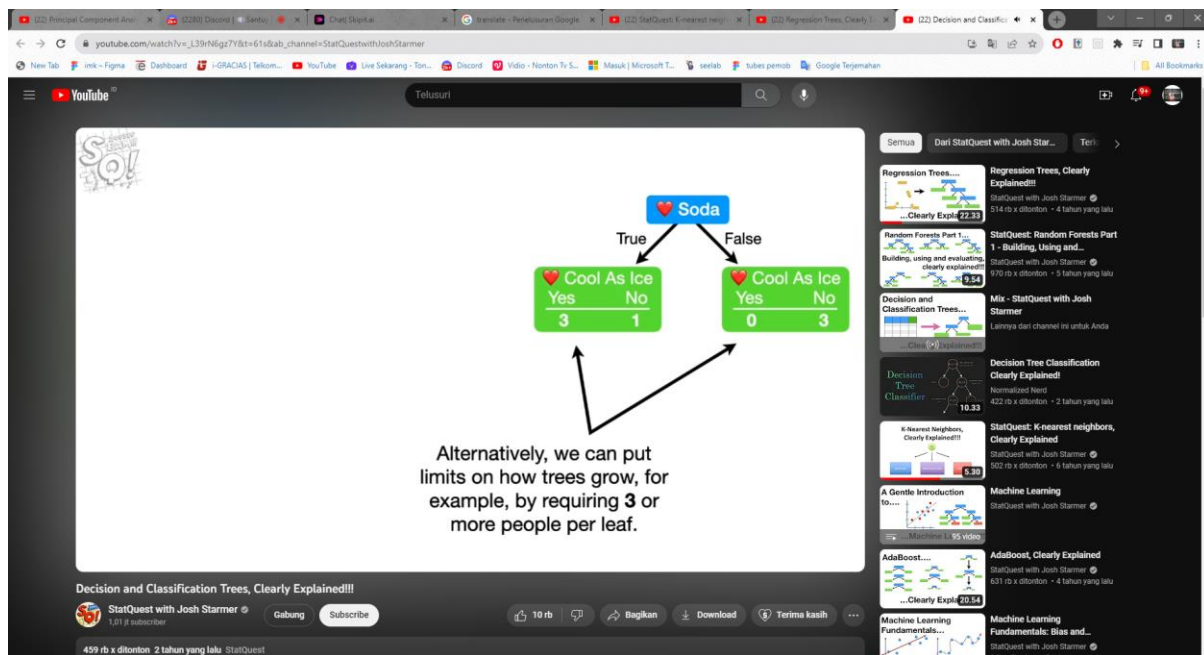
Gene	Cell1 reads	Cell2 reads
a	10	8
b	0	2
c	14	10
d	33	45
e	50	42
f	80	72
g	95	90
h	44	50
i	60	50
... (etc)	... (etc)	... (etc)

Below the table, the video title "Principal Component Analysis (PCA) clearly explained (2015)" is visible, along with the channel name "StatQuest with Josh Starmer" and a subscriber count of 1,011. The right sidebar shows a list of related videos, including "Linear Regression, Clearly Explained!!!", "StatQuest: Principal Component Analysis (PCA), Step-by-Step", "PCA Main Ideas...", and "PCA: the basics - simply explained".

1. Analisis Komponen Utama (PCA) adalah teknik statistik yang digunakan untuk menganalisis pola dan hubungan dalam data.
2. PCA dapat digunakan untuk memvisualisasikan persamaan dan perbedaan antar sel berdasarkan pola transkripsinya.
3. Sel dengan pola transkripsi serupa akan berkumpul pada plot PCA.
4. Dengan memeriksa skor pengaruh pada komponen utama nomor satu, kita dapat mengidentifikasi gen yang mempunyai dampak signifikan dalam membedakan sel dermal di sisi kiri dan sel saraf di sisi kanan grafik.
5. Demikian pula, memeriksa skor pengaruh pada komponen utama nomor dua dapat membantu mengidentifikasi gen yang membedakan sel darah dari sel saraf dan sel dermal.
6. Data asli dapat digunakan untuk memperoleh nilai X dan Y untuk memplot sel pada gambar PCA.
7. Scree plot adalah plot diagnostik yang digunakan untuk menilai kegunaan plot PCA, yang menunjukkan seberapa besar variasi yang dapat dijelaskan oleh setiap komponen utama.
8. Scree plot yang diinginkan akan menunjukkan bahwa sebagian besar variasi dijelaskan oleh dua komponen utama pertama.
9. Dalam terminologi PCA, bobot yang menggambarkan pentingnya setiap gen untuk komponen utama disebut pemuatan.
10. Array pemuatan disebut sebagai vektor eigen dalam terminologi PCA.



1. Algoritma K-nearest neighbours digunakan untuk mengkategorikan data yang tidak diketahui berdasarkan kesamaan dengan data yang diketahui.
2. Nilai K menentukan berapa banyak titik data terdekat yang dipertimbangkan saat membuat klasifikasi.
3. Nilai K yang rendah, seperti 1 atau 2, dapat menimbulkan noise dan sensitif terhadap outlier.
4. Nilai K yang besar menghaluskan variasi tetapi mungkin mengabaikan kategori dengan sedikit sampel.
5. Untuk menemukan nilai K yang optimal, nilai yang berbeda harus diuji dan dievaluasi.
6. Algoritma bekerja dengan membandingkan data yang tidak diketahui dengan K tetangga terdekat dalam kumpulan data yang diketahui.
7. Kategori mayoritas di antara K tetangga menentukan klasifikasi data yang tidak diketahui.
8. Algoritma dapat digunakan dalam berbagai aplikasi, seperti pengenalan gambar, sistem rekomendasi, dan deteksi anomali.
9. Pemilihan K harus mempertimbangkan karakteristik spesifik data dan tingkat akurasi yang diinginkan.
10. Algoritme K-nearest neighbours adalah metode serbaguna dan intuitif untuk tugas klasifikasi.



1. Saat membuat pohon keputusan, jumlah optimal orang yang dibutuhkan per daun ditentukan melalui validasi silang, yang menguji berbagai nilai untuk menemukan nilai yang paling sesuai.
2. Nilai keluaran pohon keputusan ditentukan oleh simpul daun, yang mewakili prediksi atau klasifikasi akhir.
3. Proses pembangunan pohon keputusan tidak mengetahui terlebih dahulu jumlah orang yang ideal per daun, sehingga perlu dilakukan validasi silang untuk mencari nilai optimal.
4. Validasi silang melibatkan pemisahan kumpulan data menjadi beberapa subkumpulan, melatih pohon keputusan tentang kombinasi berbeda dari subkumpulan ini, dan mengevaluasi kinerja untuk memilih nilai parameter terbaik.
5. Konsep "leaf love" mengacu pada pentingnya node daun dalam model pohon keputusan, karena mereka berkontribusi langsung pada prediksi akhir.
6. Ungkapan "sedingin es" mungkin bersifat metaforis, yang menunjukkan bahwa keakuratan dan efisiensi pohon keputusan adalah kualitas yang diinginkan.
7. Panduan belajar StatQuest di [statquest.org](https://statquest.org) menyediakan sumber daya offline untuk meninjau statistik dan pembelajaran mesin.
8. Mendukung StatQuest dapat dilakukan melalui kontribusi Patreon, menjadi anggota saluran, membeli lagu, kaos, atau hoodies asli, atau memberikan donasi.

9. Terakhir dengan frasa "pencarian", yang menekankan pencarian pengetahuan dan pemahaman berkelanjutan dalam statistik dan pembelajaran mesin.