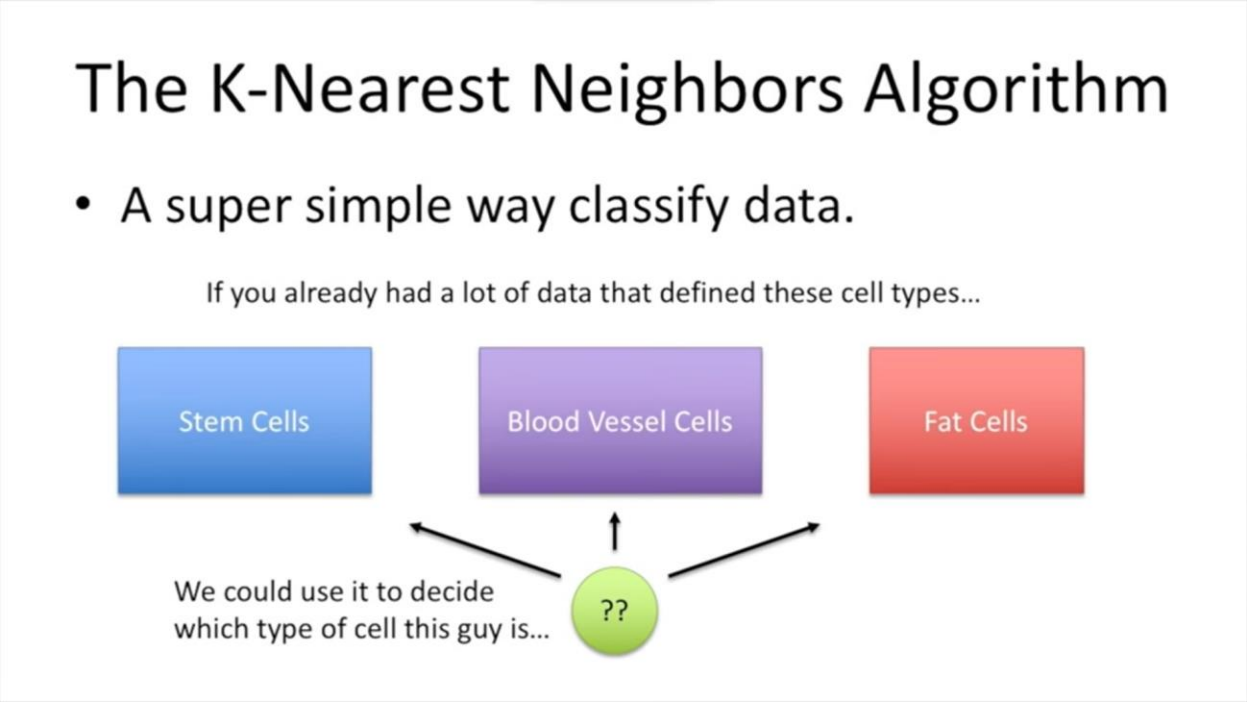
Nama : April Hamonangan Marbun

Kelas : TK4401

NIM : 1103202039



K-Nearest Neighbors (K-NN) adalah salah satu algoritma pembelajaran mesin yang digunakan dalam klasifikasi dan regresi. Ini adalah metode yang sederhana namun efektif yang digunakan untuk memprediksi kelas atau nilai berdasarkan kedekatan dengan tetangga terdekat dalam data.

Cara kerja K-NN adalah sebagai berikut:

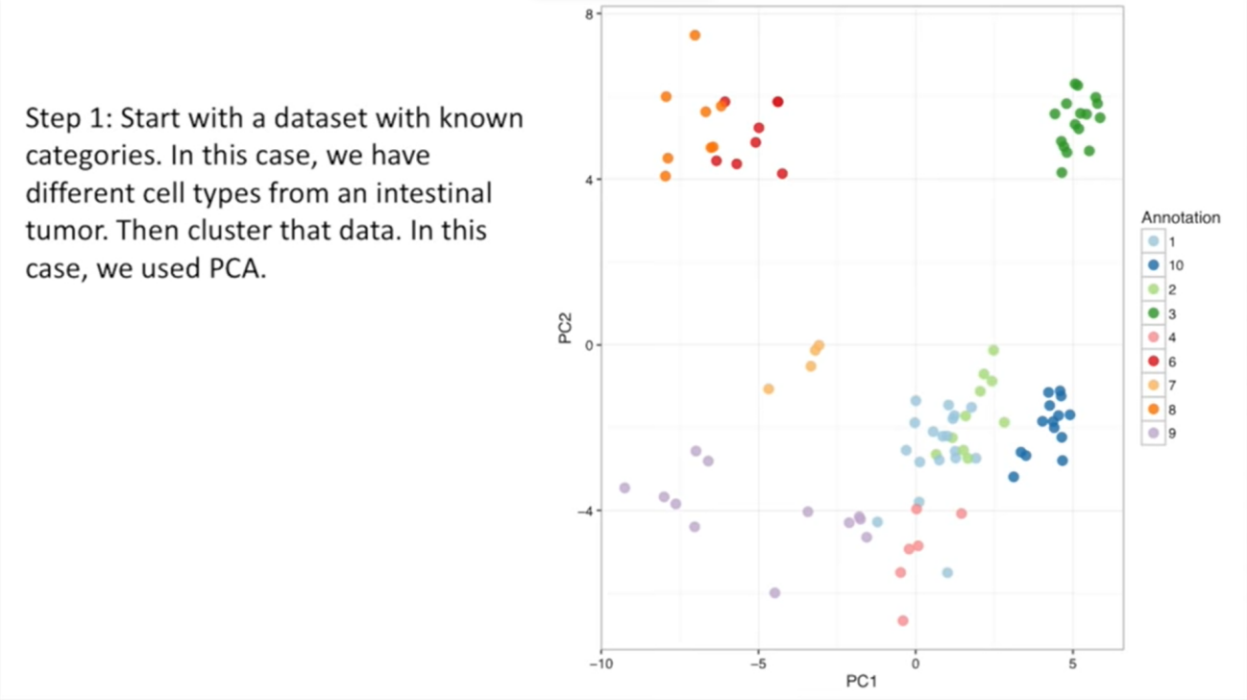
Pemilihan Nilai K: Anda harus memilih nilai K, yang merupakan jumlah tetangga terdekat yang akan digunakan untuk membuat prediksi. Nilai K biasanya dipilih secara manual dan berpengaruh terhadap kualitas prediksi.

Mengukur Jarak: K-NN mengukur jarak antara titik data yang ingin diprediksi dengan semua titik data lain dalam dataset. Pada dasarnya, ia mencari tetangga terdekat berdasarkan metrik jarak, seperti jarak Euclidean atau Manhattan.

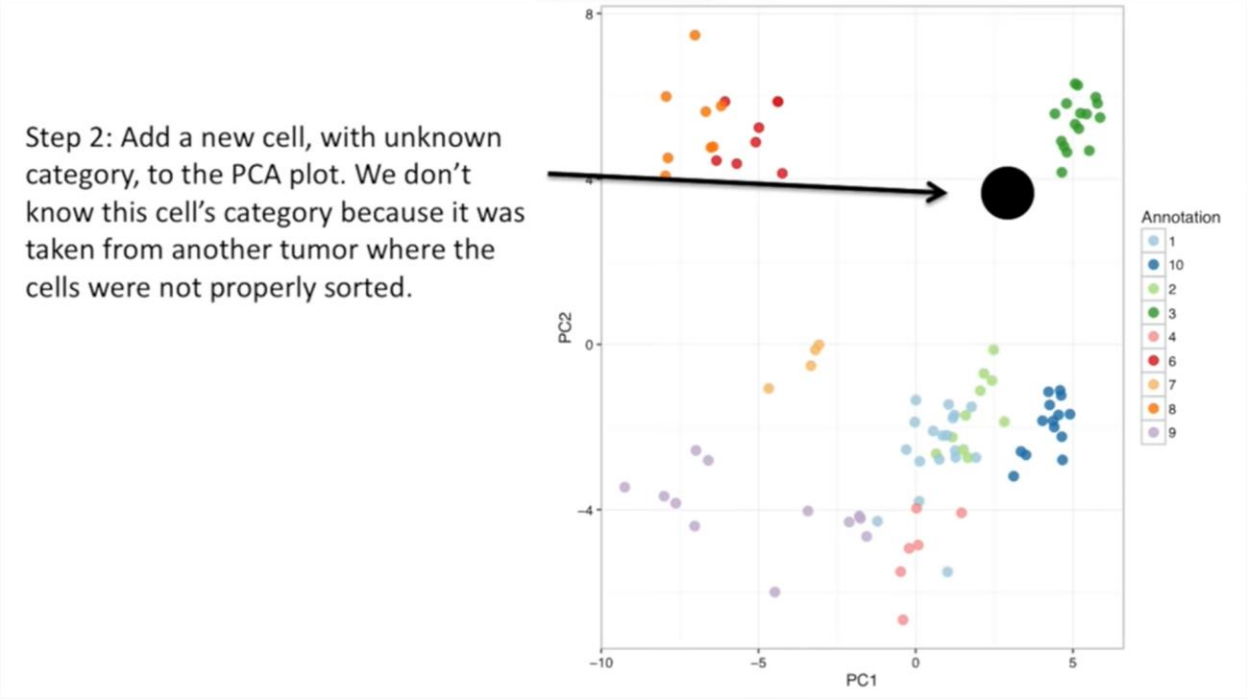
Pemilihan Tetangga Terdekat: Algoritma ini memilih K tetangga terdekat yang memiliki jarak terpendek dari titik data yang ingin diprediksi.

Klasifikasi atau Regresi: Untuk tugas klasifikasi, K-NN melakukan mayoritas suara di antara K tetangga terdekat untuk menentukan kelas prediksi. Sedangkan untuk regresi, ia mengambil rata-rata (atau metrik lain seperti median) dari nilai-nilai tetangga terdekat untuk menghasilkan nilai prediksi.

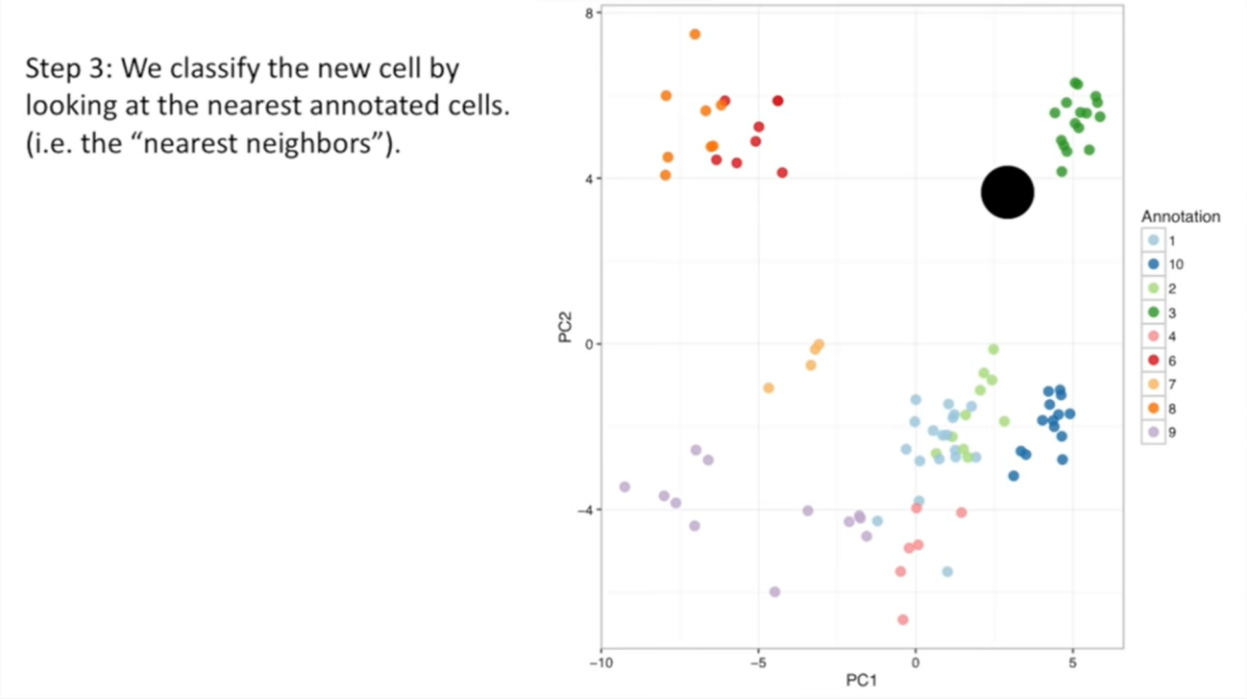
Keuntungan K-NN adalah kemudahan implementasinya dan kinerjanya yang sering kali cukup baik dalam kasus dataset yang relatif kecil. Namun, K-NN bisa menjadi komputasi intensif ketika memiliki banyak data karena harus menghitung jarak ke semua titik dalam dataset. Selain itu, pemilihan nilai K adalah faktor penting yang perlu dipertimbangkan dalam penggunaan algoritma ini.



Mulailah dengan Dataset yang Memiliki Kategori yang Diketahui: Pada tahap awal, Anda akan memiliki sebuah dataset yang berisi data tentang berbagai jenis sel dari tumor usus. Dataset ini biasanya mencakup berbagai informasi tentang sel-sel ini, seperti ekspresi gen, tingkat protein, atau fitur lain yang relevan. Selain itu, setiap entri dalam dataset diketahui termasuk dalam salah satu dari beberapa kategori atau jenis sel yang berbeda.



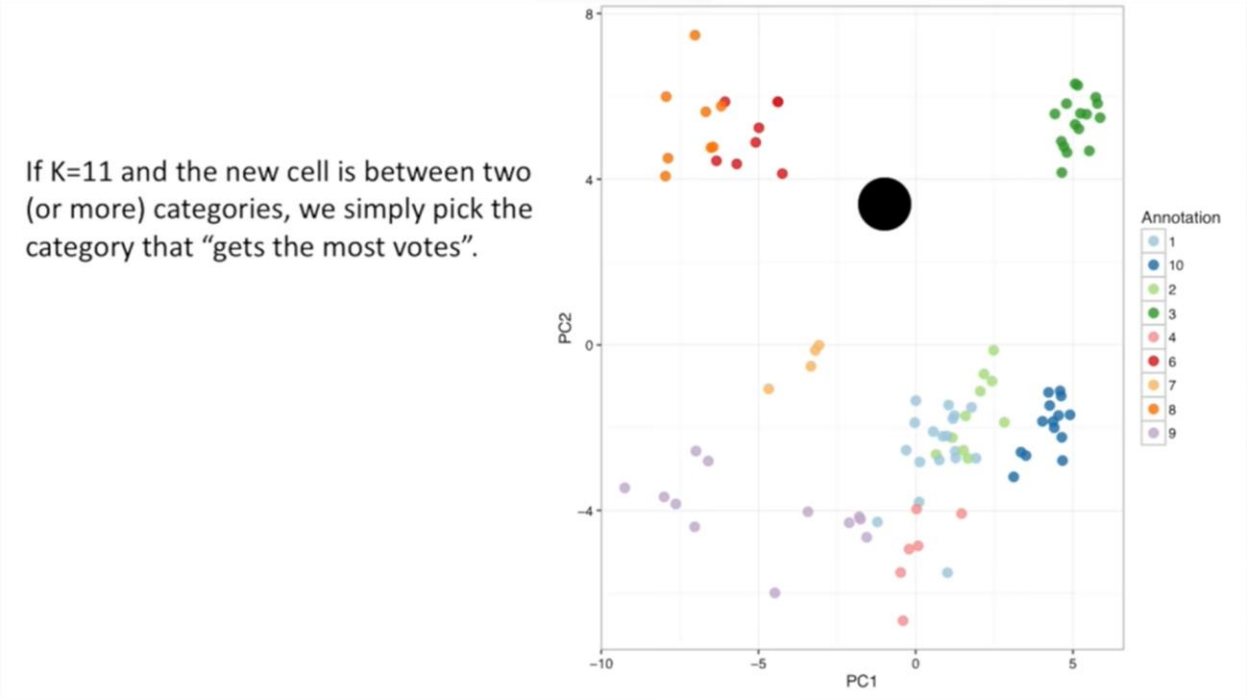
PCA Plot: Sebelumnya, Anda telah melakukan analisis PCA pada dataset asli yang terdiri dari sel-sel dengan kategori yang diketahui. PCA menghasilkan plot atau representasi visual dari data dalam dimensi yang lebih rendah (biasanya 2D atau 3D) di mana Anda dapat melihat hubungan antara jenis sel yang berbeda berdasarkan komponen utama yang ditemukan.



Menambahkan Sel ke dalam Plot PCA: Langkah ini berarti Anda ingin memasukkan sel yang tidak diketahui kategorinya ke dalam plot PCA yang sudah ada. Anda akan melakukan ini dengan mengidentifikasi posisi atau letak sel tersebut dalam ruang yang direpresentasikan oleh plot PCA. Ini

akan memberi Anda gambaran tentang di mana sel tersebut akan berada dalam konteks hubungan dengan sel-sel yang sudah ada dalam plot.

Menggunakan K-Nearest Neighbors (K-NN): Selanjutnya, Anda dapat menggunakan algoritma K-NN untuk menentukan kategori atau jenis sel yang tidak diketahui ini. K-NN akan memeriksa sel yang tidak diketahui dan melihat seberapa dekatnya dengan sel-sel yang sudah ada dalam plot PCA, kemudian menentukan kategori berdasarkan mayoritas tetangga terdekatnya.



Menggunakan K-Nearest Neighbors (K-NN): Selanjutnya, Anda dapat menggunakan algoritma K-NN untuk menentukan kategori atau jenis sel yang tidak diketahui ini. K-NN akan memeriksa sel yang tidak diketahui dan melihat seberapa dekatnya dengan sel-sel yang sudah ada dalam plot PCA, kemudian menentukan kategori berdasarkan mayoritas tetangga terdekatnya.