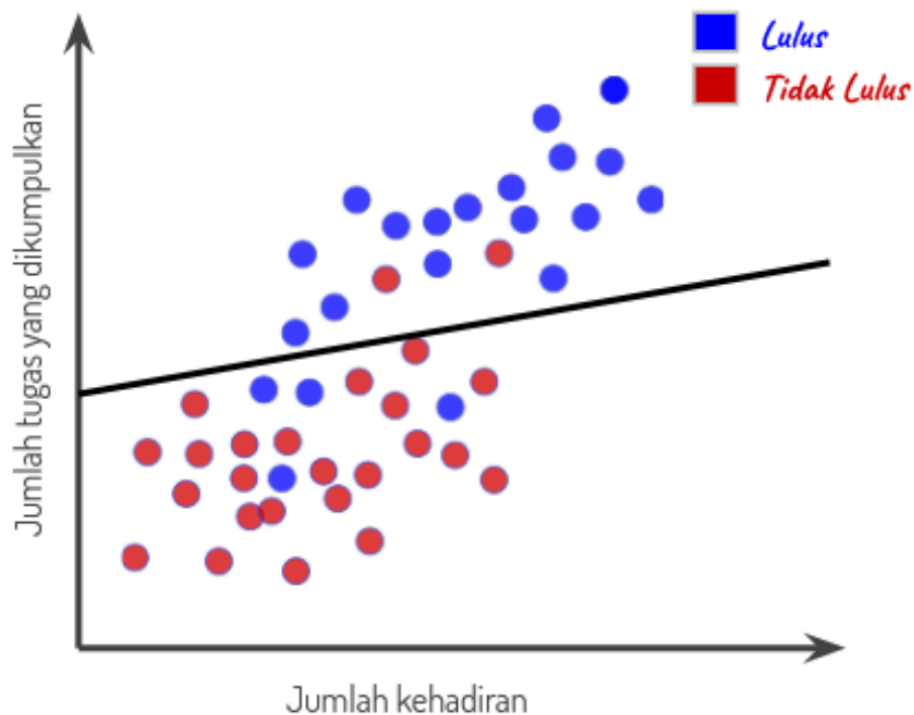


CLASSIFICATION(Klasifikasi)

Klasifikasi adalah proses pengelompokkan atau pembagian sesuatu menjadi kelas-kelas atau kelompok-kelompok berdasarkan ciri-ciri atau persamaan yang sama.



Gambar 1.1 klasifikasi

Terhadap 2 tipe klasifikasi :

- **Binary Classification** : Menentukan data ke dalam dua kelas, seperti 'spam' atau 'tidak spam'.
- **Multiclass Classification**: Menentukan data ke dalam lebih dari dua kelas, seperti memprediksi jenis buah (apel, pisang, jeruk).

Proses klasifikasi dapat berjalan dengan baik, terdapat beberapa komponen penting yang harus dipahami, yaitu fitur dan label. Fitur adalah karakteristik dari data yang menjadi dasar pengambilan keputusan model, misalnya panjang kelopak bunga atau suhu tubuh pasien. Label adalah kelas yang ingin diprediksi oleh model. Proses klasifikasi terdiri dari dua tahap utama: pelatihan (training) dan pengujian (testing). Pada tahap pelatihan, model mempelajari hubungan antara fitur dan label. Sementara itu, pada tahap pengujian, model diuji dengan data baru untuk mengukur keakuratannya.

Metode algoritma

No	Metode Klasifikasi	Keunggulan	Kelemahan
1	Logistic Regression	- Sederhana dan cepat- Mudah diinterpretasi	- Tidak cocok untuk data non-linear- Sensitif terhadap outlier
2	K-Nearest Neighbors	- Mudah dipahami- Tidak perlu pelatihan (lazy learning)	- Boros memori dan waktu- Terpengaruh skala dan outlier
3	Decision Tree	- Mudah divisualisasikan- Bisa untuk data kategorikal dan numerik	- Rentan overfitting- Tidak stabil pada dataset kecil
4	Random Forest	- Lebih akurat dari decision tree- Tahan terhadap overfitting	- Kompleks dan lambat- Sulit diinterpretasi
5	Support Vector Machine	- Efektif pada data dimensi tinggi- Bagus untuk margin yang jelas	- Lambat pada data besar- Perlu tuning kernel
6	Naive Bayes	- Cepat dan sederhana- Baik untuk data teks dan kecil	- Asumsi independensi fitur sering tidak realistis- Kurang cocok untuk fitur kontinu