

PEMBELAJARAN MESIN

Oleh: Dr Kamree & Mohd Fadil

Kandungan

- Pengenalan
- Langkah membina Model Pembelajaran Mesin
- Jenis-jenis Pembelajaran Mesin
- Alat untuk Membina Pembelajaran Mesin
- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
- Jenis-jenis Masalah Pembelajaran Mesin
- Jenis-jenis Algoritma
- Menilai Modal
- Q&A

Pengenalan

Apakah maksud pembelajaran mesin?

Pembelajaran mesin adalah cabang kecerdasan buatan (AI) yang difokuskan pada pembangunan aplikasi yang belajar dari data dan meningkatkan ketepatannya dari masa ke masa tanpa diprogramkan untuk melakukannya.

Contoh-contoh pembelajaran mesin

1. Pembantu Digital (Digital Assistants)
2. Chatbots
3. Pemanduan kereta sendiri (Self-driving cars)
4. Pengesanan Imej Perubatan (Medical Image Detection)
5. Pengesanan Penipuan (Fraud Detection)
6. Keselamatan siber (Cybersecurity)

Langkah membina Model Pembelajaran Mesin



Jenis-jenis Pembelajaran Mesin

Supervised

Pembelajaran mesin yang diselia melatih dirinya pada set data berlabel. Artinya, data dilabel dengan informasi yang model pembelajaran mesin sedang dibangun untuk menentukan dan bahkan mungkin diklasifikasikan dengan cara model tersebut seharusnya mengklasifikasikan data

Unsupervised

Pembelajaran mesin yang tidak diawasi memakan data yang tidak berlabel – banyak dan banyak – dan menggunakan algoritma untuk mengekstrak ciri-ciri bermakna yang diperlukan untuk melabel, menyusun, dan mengklasifikasikan data dalam masa nyata, tanpa campur tangan manusia.

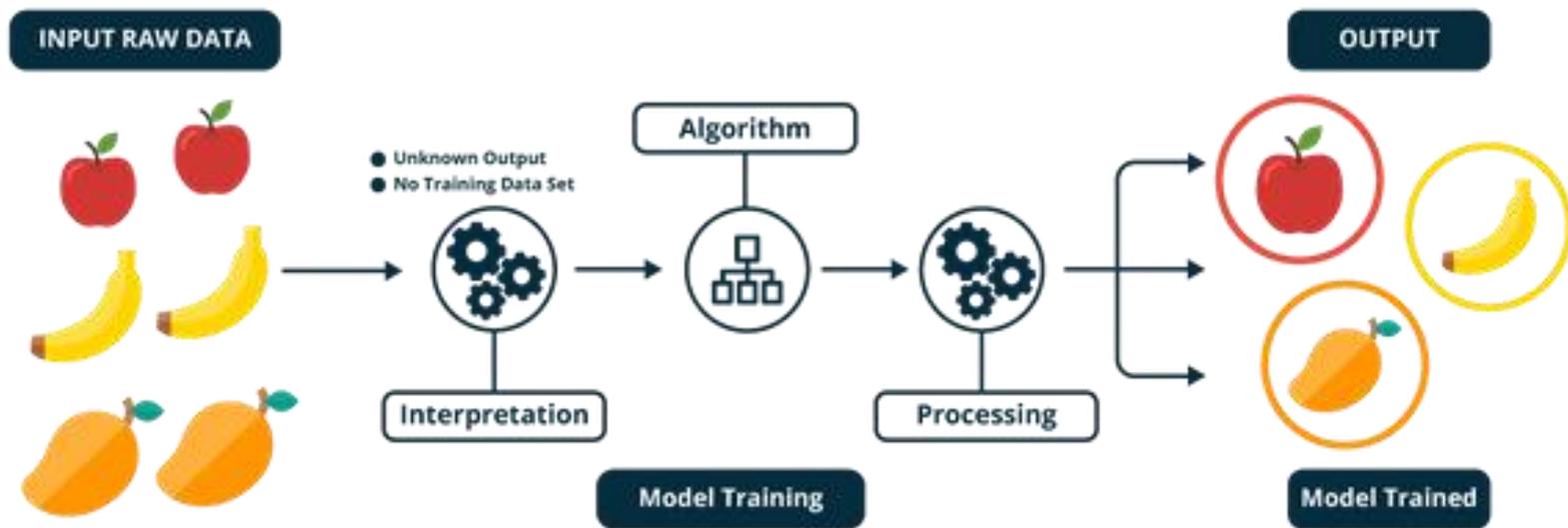
Reinforcement

Pembelajaran mesin penguhan adalah model pembelajaran mesin tingkah laku yang serupa dengan pembelajaran yang diselia, tetapi algoritma tidak dilatih menggunakan data sampel. Model ini belajar seiring dengan menggunakan percubaan dan kesilapan.

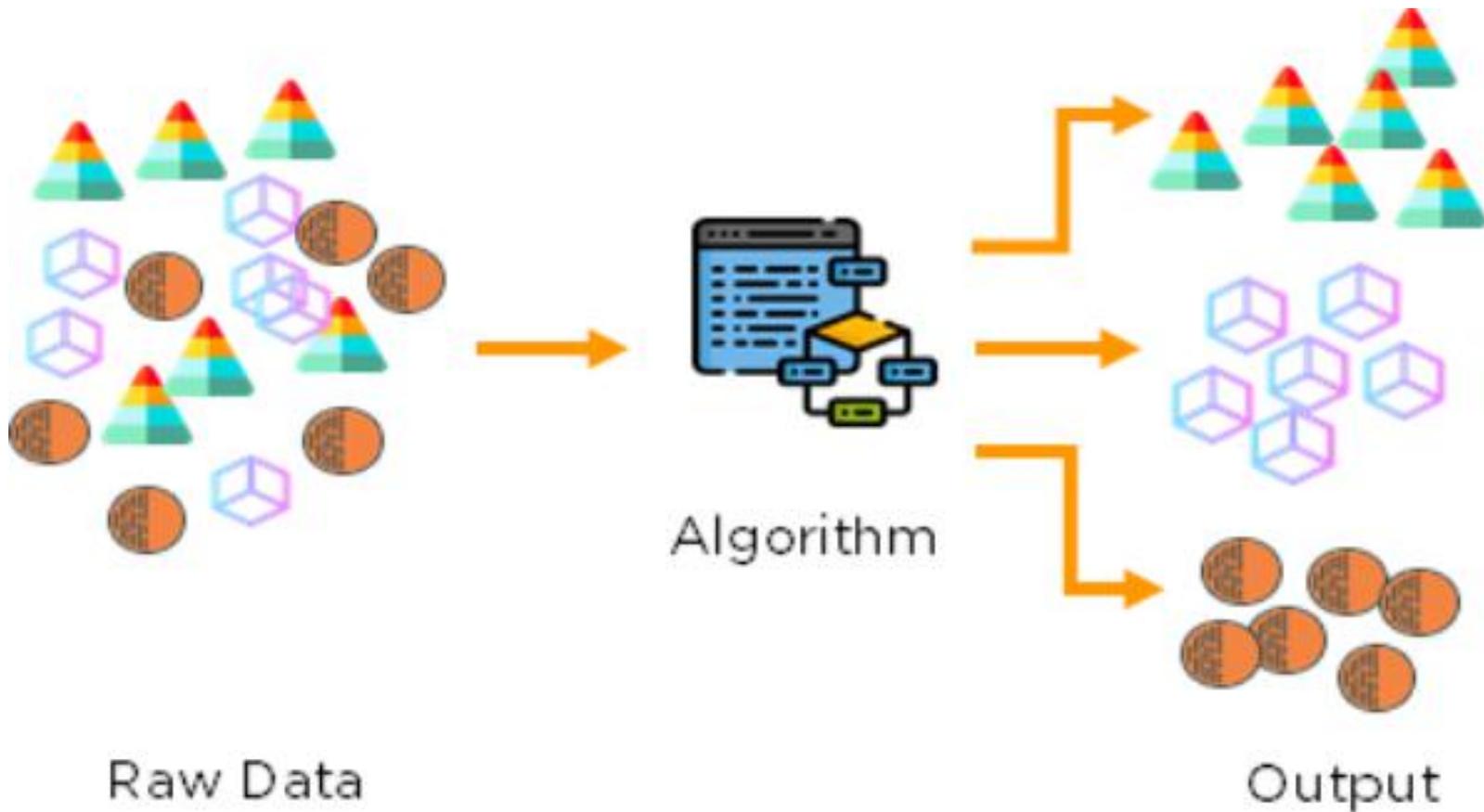
Alat untuk Membina Pembelajaran Mesin



Supervised Learning



Unsupervised Learning



Reinforcement Learning



Jenis-jenis Masalah Pembelajaran Mesin

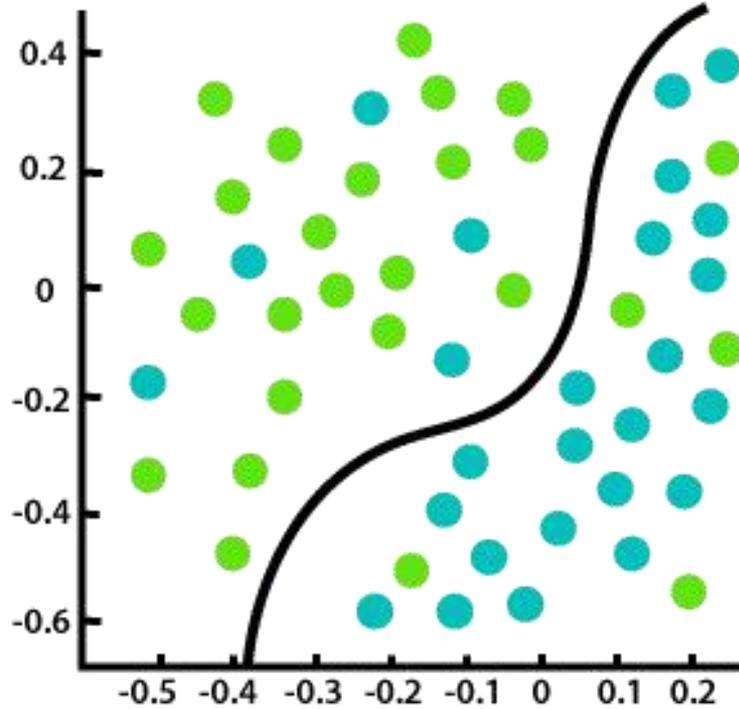
Regresi (Regression)

Masalah regresi adalah apabila pemboleh ubah output adalah nilai nyata atau berterusan, seperti "gaji" atau "berat". Banyak model yang boleh digunakan, yang paling mudah adalah regresi linier. Ia cuba menyesuaikan data dengan hiper-pesawat terbaik yang melalui titik.

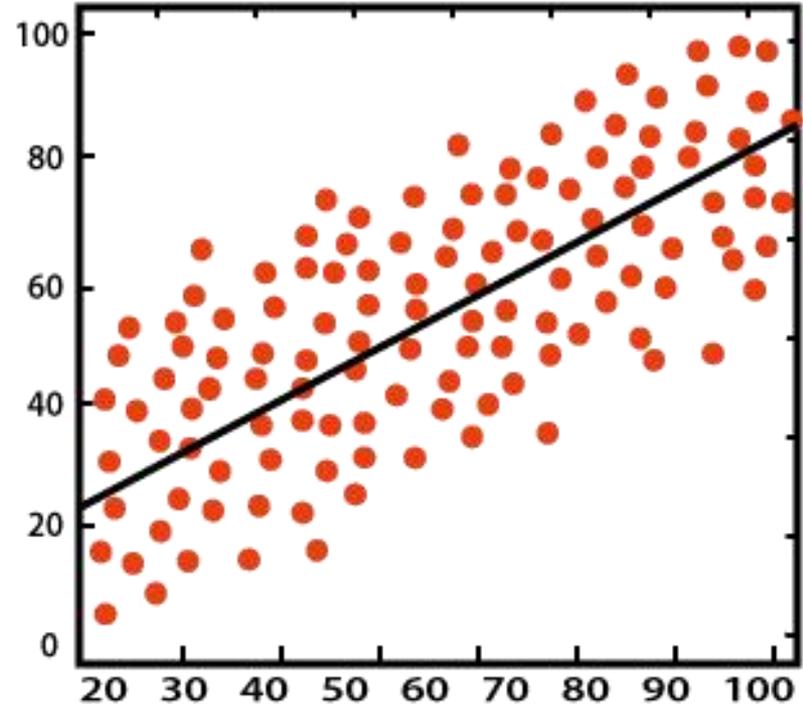
Pengelasan (Classification)

Masalah klasifikasi adalah ketika pemboleh ubah output adalah kategori, seperti "merah" atau "biru" atau "penyakit" dan "tidak ada penyakit". Model klasifikasi cuba membuat kesimpulan dari nilai yang diperhatikan. Diberi satu atau lebih input, model klasifikasi akan cuba meramalkan nilai satu atau lebih hasil.

Perbedaan Regresi dan Pengelasan

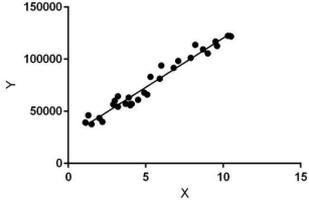
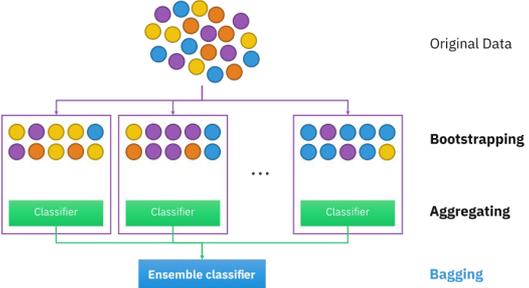
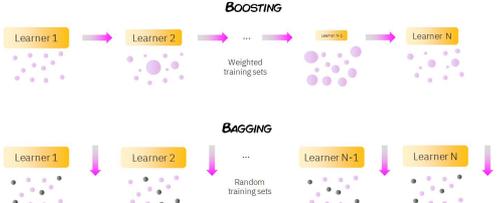


Classification



Regression

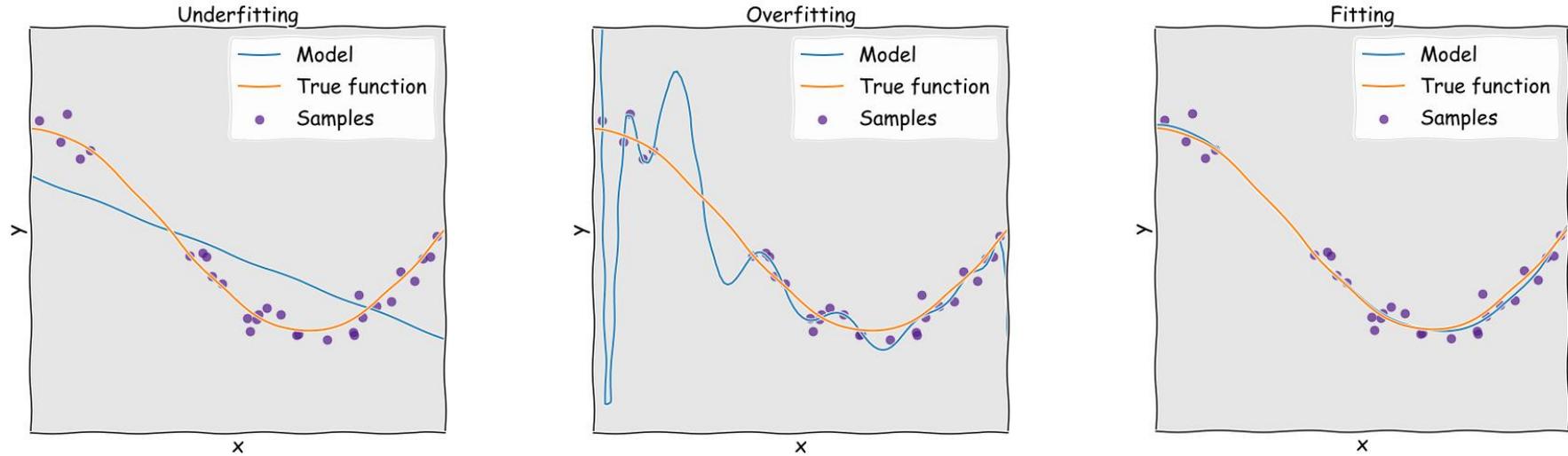
Jenis-jenis Algoritma

Jenis	Contoh	Penjelasan
Linear		<p>Regresi memodelkan nilai ramalan sasaran berdasarkan pemboleh ubah bebas. Ia digunakan untuk mengetahui hubungan antara pemboleh ubah dan ramalan.</p>
Bagging		<p>Pengumpulan bootstrap, juga disebut bagging (dari agregasi bootstrap), adalah meta-algoritma ensemble pembelajaran mesin yang direka untuk meningkatkan kestabilan dan ketepatan algoritma pembelajaran mesin.</p>
Bosting		<p>Boosting, yang awalnya bernama Hypothesis Boosting, terdiri dari idea menyaring atau menimbang data yang digunakan untuk melatih data yang lemah.</p>

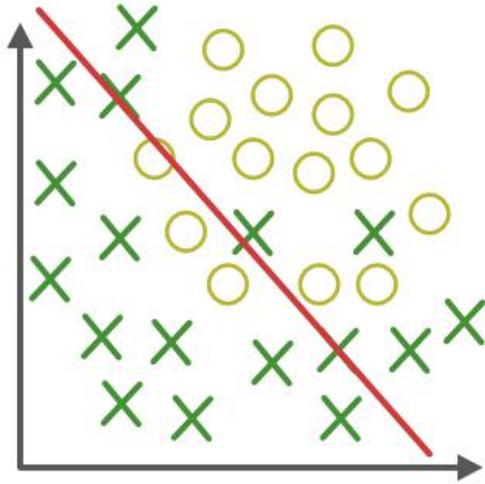
Menilai Model

Apakah fungsi penilaian model?

Model Evaluasi adalah bahagian yang tidak terpisahkan dari proses pengembangan model. Ia membantu mencari model terbaik yang mewakili data kami dan seberapa baik model yang dipilih akan berfungsi pada masa hadapan. Menilai prestasi model dengan data yang digunakan untuk latihan tidak dapat diterima dalam sains data kerana dapat dengan mudah menghasilkan model yang terlalu optimis dan berlebihan.

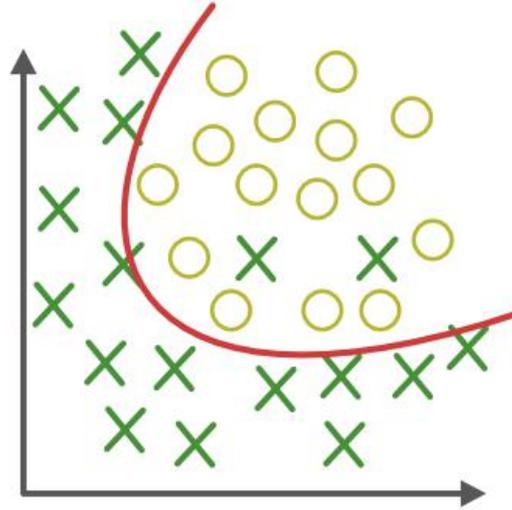


Contoh Model Pengelasan

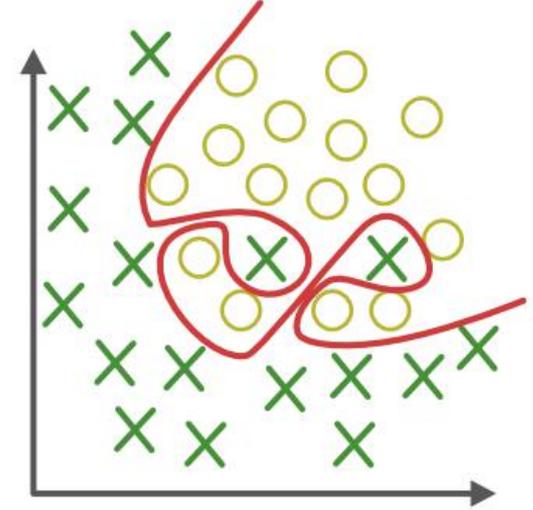


Under-fitting

(too simple to explain the variance)



Appropriate-fitting



Over-fitting

(forcefitting--too good to be true)

Penilaian Model Regresi

1. Root Mean Squared Error - RMSE adalah formula yang popular untuk mengukur kadar ralat model regresi. Namun, hanya dapat dibandingkan antara model yang kesalahannya diukur dalam unit yang sama.
2. Relative Squared Error - Tidak seperti RMSE, ralat kuasa dua relatif (RSE) dapat dibandingkan antara model yang kesalahannya diukur dalam unit yang berlainan.
3. Mean Absolute Error - Kesalahan mutlak min (MAE) mempunyai unit yang sama dengan data asal, dan hanya dapat dibandingkan antara model yang kesalahannya diukur dalam unit yang sama. Ia biasanya serupa dengan ukuran RMSE, tetapi sedikit lebih kecil.
4. R2 Score - R2 menerangkan bahagian varian pemboleh ubah bersandar yang dijelaskan oleh model regresi. Jika model regresi "sempurna", SSE adalah nol, dan R2 adalah 1. Jika model regresi adalah kegagalan total, SSE sama dengan SST, tidak ada variasi yang dijelaskan oleh regresi, dan R2 adalah nol.

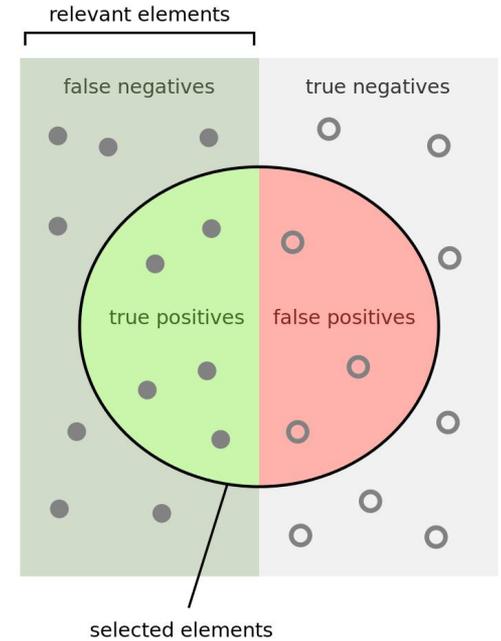
Penilaian Model Pengelasan

Confusion Matrix - Matriks kekeliruan menunjukkan bilangan ramalan yang betul dan salah yang dibuat oleh model klasifikasi berbanding dengan hasil sebenar (nilai sasaran) dalam data. Matriks adalah $N \times N$, di mana N adalah bilangan nilai sasaran (kelas). Prestasi model sedemikian biasanya dinilai menggunakan data dalam matriks.

- Precision - Precision adalah nisbah pemerhatian positif yang diramalkan dengan betul dan jumlah pemerhatian positif yang diramalkan.
- Recall - Recall adalah nisbah pemerhatian positif yang diramalkan dengan betul untuk semua pemerhatian dalam kelas sebenar
- F1-Score - F1-Score adalah purata wajaran Precision and Recall. Oleh itu, skor ini mengambil kira positif positif dan negatif palsu. Secara intuitif ia tidak semudah difahami ketepatan, tetapi F1 biasanya lebih berguna daripada ketepatan.
- Accuracy - Ketepatan adalah ukuran prestasi yang paling intuitif dan ia hanyalah nisbah pemerhatian yang diramalkan dengan betul dan jumlah pemerhatian. Seseorang mungkin berfikir bahawa, jika kita mempunyai ketepatan yang tinggi maka model kita adalah yang terbaik. Ketepatan adalah ukuran yang baik tetapi hanya apabila set data adalah simetri dimana nilai negatif positif dan negatif palsu hampir sama

Contoh Matriks Kekeliruan

	precision	recall	f1-score	support
0	0.90	0.99	0.94	67
1	0.98	0.85	0.91	47
accuracy			0.93	114
macro avg	0.94	0.92	0.93	114
weighted avg	0.93	0.93	0.93	114



How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

Q&A



TERIMA KASIH