

Predictive Model (in Python)

1. Jenis-Model Predictive dan teknik

Secara perbandingan antara model predictive.

Nama Model	Fungsi Utama	Teknik/Algoritma
Linear Regression	Meramalkan nilai berterusan (contoh: harga, jualan) berdasarkan hubungan linear antara ciri.	Algoritma Linear
Logistic Regression	Meramalkan hasil kategori (contoh: ya/tidak, 0/1) berdasarkan kebarangkalian.	Algoritma Linear
Decision Tree	Membuat keputusan berdasarkan pembahagian data kepada kumpulan-kumpulan kecil.	Algoritma Non-linear
Random Forest	Kombinasi pelbagai Decision Trees untuk ramalan yang lebih tepat dan stabil.	Ensemble (Non-linear)
Support Vector Machine (SVM)	Memisahkan data ke dalam kategori menggunakan hiperplan terbaik.	Algoritma Linear/Non-linear
K-Nearest Neighbors (KNN)	Mengklasifikasikan data berdasarkan jarak terdekat dengan data latih.	Algoritma Non-parametrik
Naive Bayes	Model klasifikasi berdasarkan teorem Bayes, sesuai untuk data teks atau kategori.	Probabilistik
XGBoost	Model berasaskan gradien boosting, sangat berkesan untuk data kompleks.	Ensemble (Boosting)
K-Means Clustering	Membahagikan data ke dalam kumpulan tanpa label berdasarkan kesamaan.	Clustering (Unsupervised)
Prophet	Ramalan deret masa (time series) dengan pola bermusim, trend, dan cuti.	Time Series (Additive)
ARIMA	Ramalan deret masa dengan pola auto-regressive dan moving average.	Time Series (Statistical)
Neural Networks	Meramalkan hasil kompleks melalui lapisan neuron tiruan.	Algoritma Deep Learning

Penjelasan Teknik atau Jenis Algoritma

1. Linear: Membentuk hubungan linear antara ciri input dan output.
2. Non-linear: Model lebih fleksibel untuk menangani hubungan yang kompleks.
3. Ensemble: Gabungan beberapa model untuk meningkatkan ketepatan.
4. Time Series: Direka untuk ramalan berasaskan masa dengan pola tertentu.
5. Probabilistik: Menggunakan kebarangkalian untuk membuat keputusan.
6. Clustering: Mengelompokkan data ke dalam kumpulan tanpa label.

7. Deep Learning: Menggunakan rangkaian neural untuk ramalan data berskala besar atau kompleks.

Apakah maksud linear?

Dalam konteks matematik dan pembelajaran mesin, linear merujuk kepada hubungan antara pembolehubah di mana perubahan dalam satu pembolehubah adalah berkadar terus (proportional) dengan perubahan dalam pembolehubah lain. Hubungan ini dapat digambarkan dengan menggunakan garis lurus pada graf.

2. Jenis model yang boleh digunakan dalam Python

Dalam python ada beberapa library yang kita boleh guna untuk buat analisis data gunakan Predictive Model:

Jenis Model Predictive	Penerangan	Library
Linear Regression	Meramalkan nilai berterusan berdasarkan hubungan linear antara ciri input dan output.	scikit-learn, statsmodels
Logistic Regression	Meramalkan hasil kategori binari (contoh: Ya/Tidak) berdasarkan kebarangkalian.	scikit-learn, statsmodels
Decision Tree	Membuat keputusan berdasarkan pembahagian data secara berulang menggunakan nod dan cabang.	scikit-learn
Random Forest	Gabungan beberapa Decision Trees untuk meningkatkan ketepatan dan mengurangkan overfitting.	scikit-learn
Support Vector Machine (SVM)	Memisahkan data ke dalam kategori menggunakan hiperplan terbaik untuk klasifikasi atau regresi.	scikit-learn
K-Nearest Neighbors (KNN)	Klasifikasi berdasarkan majoriti jiran terdekat dalam ruang ciri.	scikit-learn
Naive Bayes	Klasifikasi berdasarkan teorem Bayes, sesuai untuk data teks atau kategori.	scikit-learn
XGBoost	Model berasaskan gradien boosting yang sangat berkesan untuk ramalan data kompleks.	xgboost
Gradient Boosting	Menggabungkan model lemah secara berperingkat untuk meningkatkan ketepatan ramalan.	scikit-learn, xgboost, lightgbm, catboost
K-Means Clustering	Membahagikan data ke dalam kumpulan tanpa label berdasarkan kesamaan dalam ciri.	scikit-learn
Prophet	Ramalan deret masa dengan pola bermusim, trend, dan cuti.	prophet
ARIMA/SARIMA	Ramalan deret masa berdasarkan pola auto-regressive dan moving average.	statsmodels, pmdarima

Neural Networks (ANN)	Model berdasarkan rangkaian neuron tiruan untuk mempelajari hubungan kompleks.	tensorflow, keras, pytorch
LSTM (Long Short-Term Memory)	Model deep learning yang sesuai untuk deret masa atau data bergantung masa.	tensorflow, keras, pytorch
Time Series Clustering	Mengelompokkan pola dalam data deret masa untuk mencari persamaan.	tslearn, sklearn
AutoML	Pengoptimuman automatik untuk memilih model terbaik berdasarkan data.	h2o, autosklearn, tpot

Penjelasan Library

1. `scikit-learn` : Pustaka yang sangat popular untuk pelbagai model pembelajaran mesin (Machine Learning).
2. `statsmodels` : Digunakan untuk model statistik seperti regresi linear dan ARIMA.
3. `xgboost`, `lightgbm`, `catboost` : Model boosting yang sangat cekap dan digunakan dalam pertandingan data.
4. `prophet` : Direka oleh Facebook untuk ramalan deret masa dengan pola bermusim.
5. `tensorflow`, `keras`, `pytorch` : Pustaka untuk pembelajaran mendalam (Deep Learning).
6. `tslearn` : Digunakan untuk analisis deret masa, termasuk clustering dan klasifikasi.
7. `autosklearn`, `tpot`, `h2o` : Alat AutoML yang mempercepatkan proses mencari model terbaik.

Penggunaan library dan contoh:

Library	Kegunaan Utama	Contoh Nyata
scikit-learn	Regresi, klasifikasi, clustering	Ramalan harga rumah, churn detection
statsmodels	Model statistik, ARIMA	Kadar inflasi, hubungan harga & permintaan
xgboost, lightgbm, catboost	Boosting model, data kompleks	Ramalan pembelian promosi, kelulusan pinjaman
prophet	Deret masa, pola bermusim	Jualan bulanan, trafik laman web
tensorflow, keras, pytorch	Deep learning, imej, suara	Pengecaman wajah, NLP, teks ke suara
tslearn	Deret masa, clustering, klasifikasi	Pola perbelanjaan, anomali data sensor
autosklearn, tpot, h2o	AutoML, automasi model terbaik	Ramalan jualan, faktor kepuasan pelanggan

Penerangan lanjut: `scikit-learn` : Digunakan untuk tugas sehari-hari seperti analisis data perniagaan, ramalan harga, atau pengesanan pelanggan yang mungkin berhenti melanggan.

`statsmodels` : Sesuai untuk analisis regresi dan memahami hubungan statistik.
Contohnya, syarikat insurans menggunakan model ARIMA untuk meramal tuntutan insurans bulanan.

`xgboost`, `lightgbm`, `catboost` : Banyak digunakan dalam pertandingan sains data seperti Kaggle, tetapi juga sangat berkesan dalam perniagaan untuk mengenal pasti pelanggan yang mempunyai risiko rendah atau tinggi.

`prophet` : Ramalan masa depan untuk tujuan pengurusan, seperti mempersiapkan bekalan stok berdasarkan musim perayaan.

`tensorflow`, `keras`, `pytorch` : Dalam aplikasi teknologi tinggi seperti pengesanan wajah, ramalan harga saham, atau pemprosesan imej untuk aplikasi perubatan.

`tslearn` : Untuk aplikasi IoT dan analisis pola seperti mengenal pasti pola dalam penggunaan tenaga rumah sepanjang hari.

`autosklearn`, `tpot`, `h2o` : Digunakan untuk menjimatkan masa dalam membina model ramalan dengan membiarkan sistem memilih model terbaik secara automatik.