

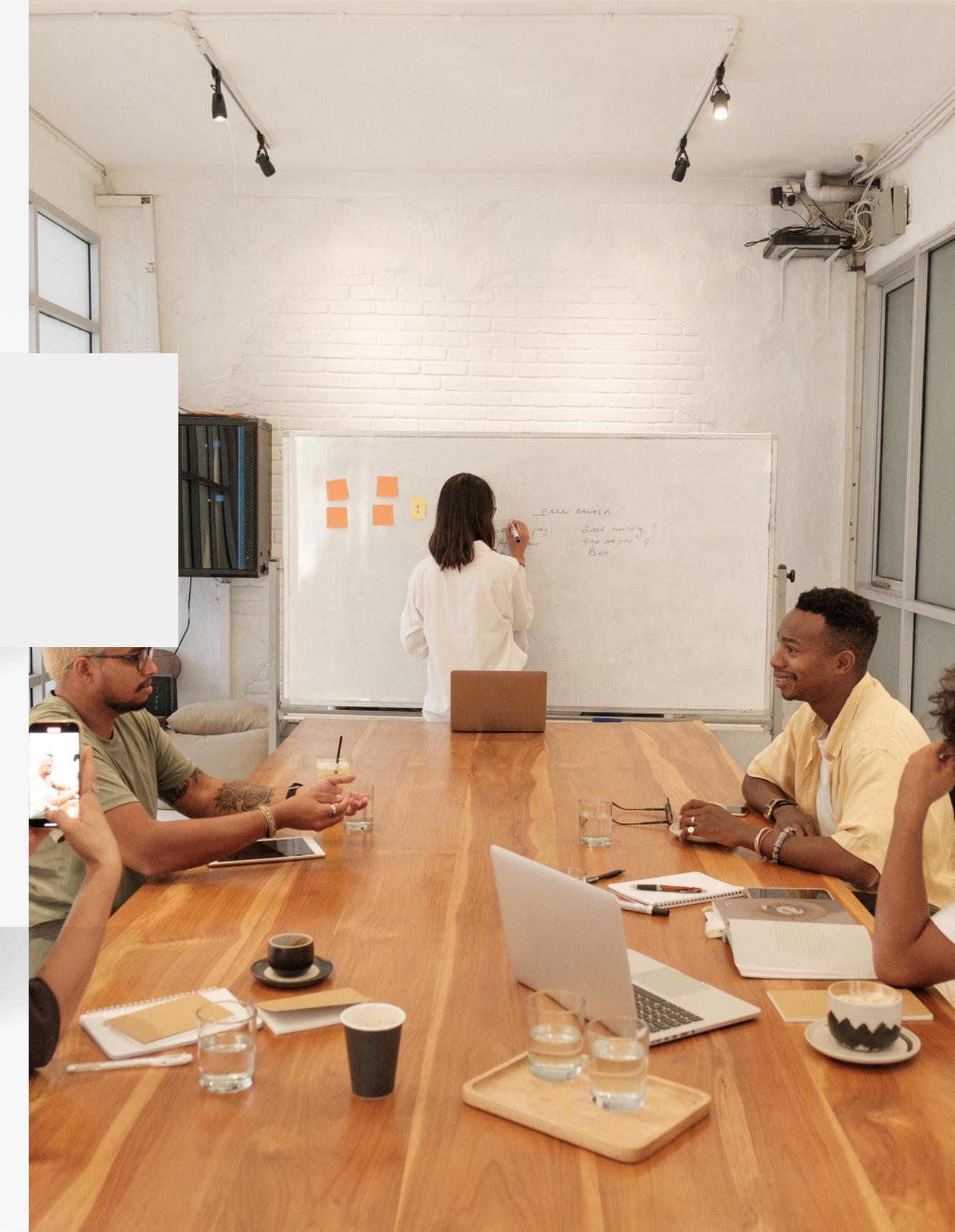


ALGORITMA

MACHINE LEARNING ADVANCE

ALGORITMA

Algoritma dalam machine learning adalah sekumpulan instruksi yang digunakan oleh sistem machine learning untuk belajar dari data dan menghasilkan model. Model machine learning adalah representasi dari data yang dapat digunakan untuk membuat prediksi, klasifikasi, atau clustering.



JENIS-JENIS ALGORITMA DALAM MACHINE LEARNING

Supervised

adalah algoritma yang menggunakan data berlabel untuk belajar. Data berlabel adalah data yang memiliki label yang sesuai dengan output yang diinginkan.

Unsupervised

adalah algoritma yang tidak menggunakan data berlabel.

SUPERVISED

Algoritma supervised learning dapat digunakan untuk berbagai macam tugas, seperti:

- Klasifikasi adalah tugas machine learning yang melibatkan pengkategorian data ke dalam kelompok. Misalnya, algoritma klasifikasi dapat digunakan untuk mengklasifikasikan email sebagai spam atau tidak spam.
- Regresi adalah tugas machine learning yang melibatkan memprediksi nilai numerik. Misalnya, algoritma regresi dapat digunakan untuk memprediksi harga rumah berdasarkan karakteristiknya.
- Prediksi adalah tugas machine learning yang melibatkan memprediksi kejadian di masa mendatang. Misalnya, algoritma prediksi dapat digunakan untuk memprediksi cuaca atau hasil pertandingan olahraga.

UNSUPERVISED

Algoritma unsupervised learning adalah algoritma yang belajar dari data yang tidak memiliki label. Data yang tidak memiliki label disebut data acak.

Algoritma unsupervised learning dapat digunakan untuk berbagai macam tugas, seperti:

- Clustering adalah tugas machine learning yang melibatkan pengelompokan data ke dalam kelompok berdasarkan kesamaannya. Misalnya, algoritma clustering dapat digunakan untuk mengelompokkan pelanggan berdasarkan perilaku pembelian mereka.

scikit-learn

Machine Learning in Python

[Getting Started](#)[Release Highlights for 1.3](#)[GitHub](#)

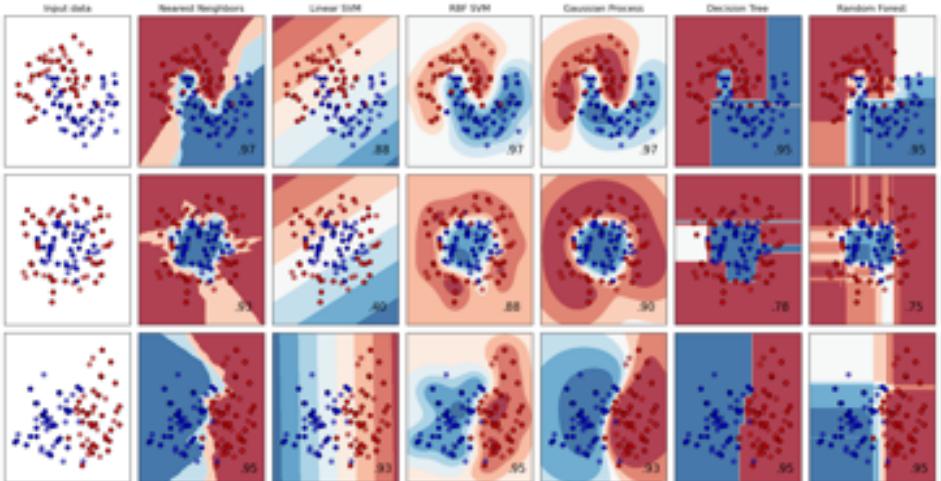
- Simple and efficient tools for predictive data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

Classification

Identifying which category an object belongs to.

Applications: Spam detection, image recognition.

Algorithms: Gradient boosting, nearest neighbors, random forest, logistic regression, and more...

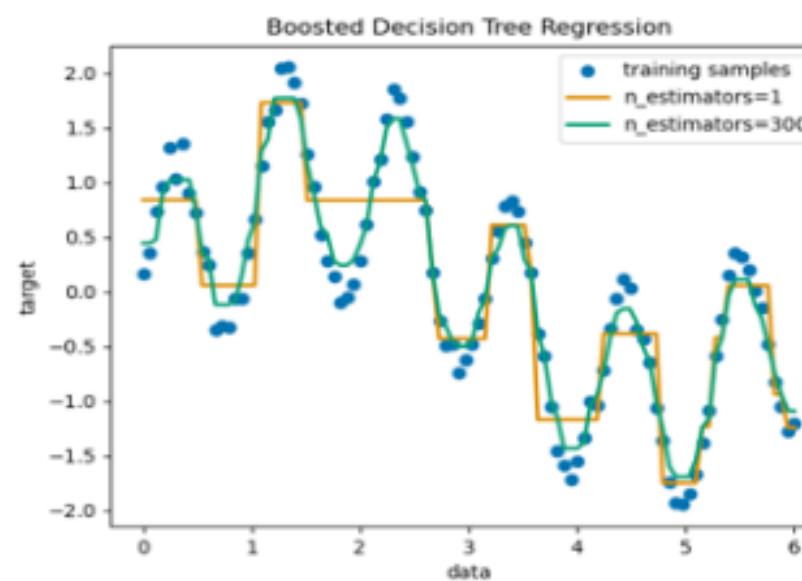
[Examples](#)

Regression

Predicting a continuous-valued attribute associated with an object.

Applications: Drug response, Stock prices.

Algorithms: Gradient boosting, nearest neighbors, random forest, ridge, and more...

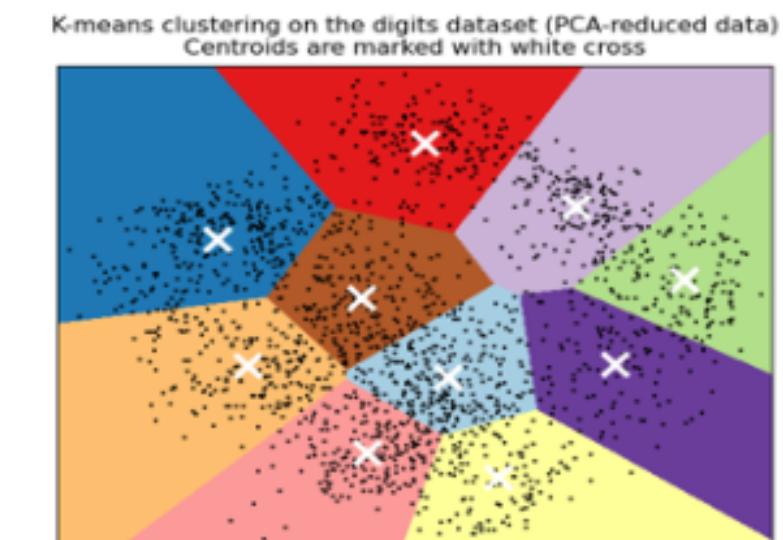
[Examples](#)

Clustering

Automatic grouping of similar objects into sets.

Applications: Customer segmentation, Grouping experiment outcomes

Algorithms: k-Means, HDBSCAN, hierarchical clustering, and more...

[Examples](#)

**ALWAYS CHECK
LIBRARY
DOCUMENTATION**



You

jelaskan ke saya algoritma tentang decision tree dan jelaskan algoritma tersebut biasa digunakan dalam modelling prediksi, klasifikasi atau klustering?

Penggunaan dalam Pemodelan:

1. **Prediksi:** Decision Tree dapat digunakan untuk memprediksi nilai variabel target berdasarkan fitur-fitur yang diberikan. Misalnya, memprediksi apakah suatu pelanggan akan membeli produk atau tidak.
2. **Klasifikasi:** Salah satu penggunaan paling umum dari Decision Tree adalah dalam klasifikasi, di mana tujuannya adalah untuk mengelompokkan data ke dalam kategori atau kelas yang berbeda. Contoh termasuk klasifikasi spam email atau identifikasi penyakit berdasarkan gejala.
3. **Klustering:** Meskipun Decision Tree biasanya tidak digunakan secara langsung untuk klustering, namun prinsip pemisahan data dalam Decision Tree dapat memberikan pandangan yang baik tentang bagaimana data dapat dikelompokkan atau di-kluster.



scikit
learn

Install



Prev

Up

Next

scikit-learn 1.3.2

[Other versions](#)

Please [cite us](#) if you use the
software.

[1.10. Decision Trees](#)

[1.10.1. Classification](#)

[1.10.2. Regression](#)

[1.10.3. Multi-output problems](#)

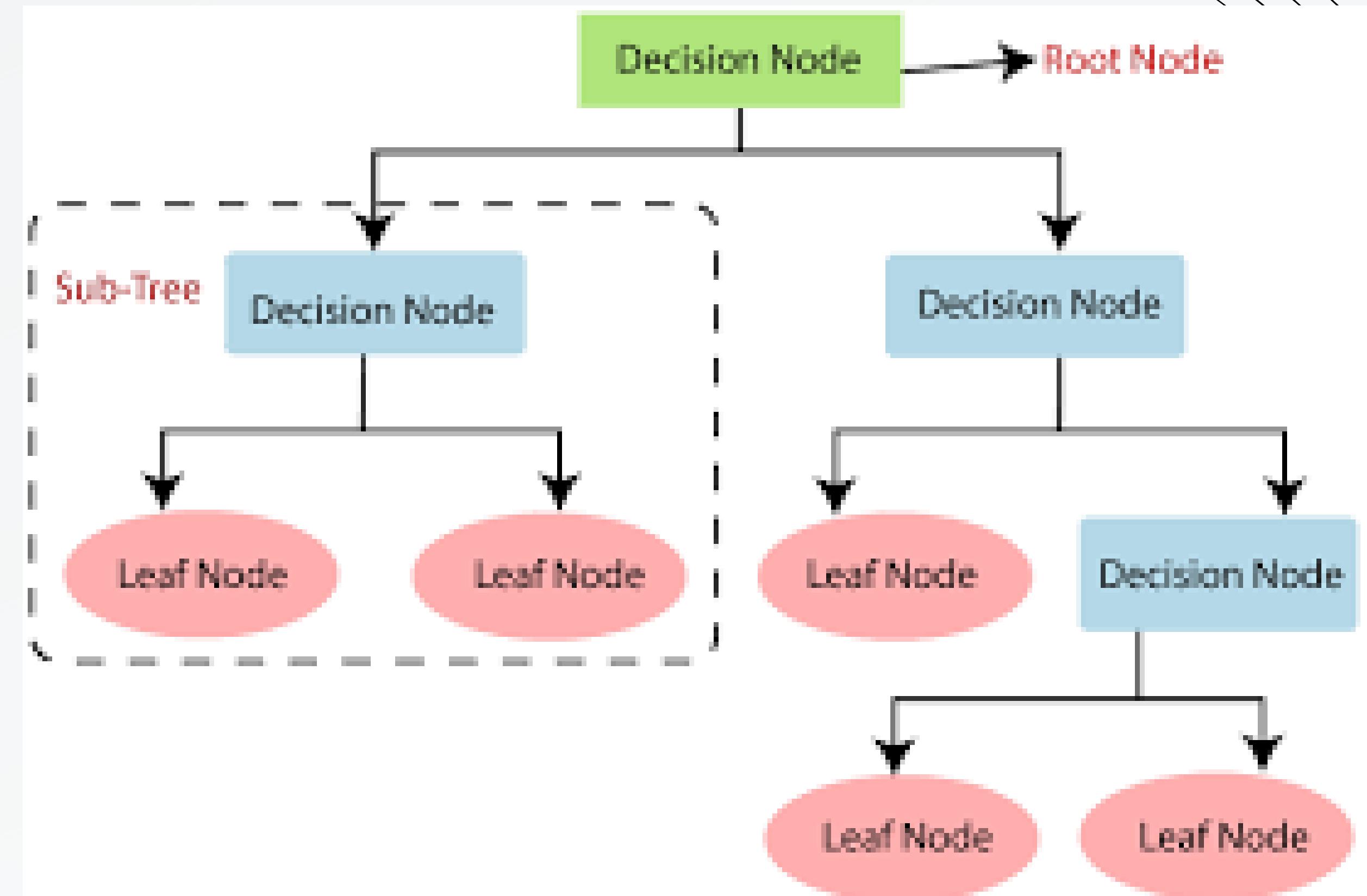
[1.10.4. Complexity](#)

[1.10.5. Tips on practical use](#)

4549

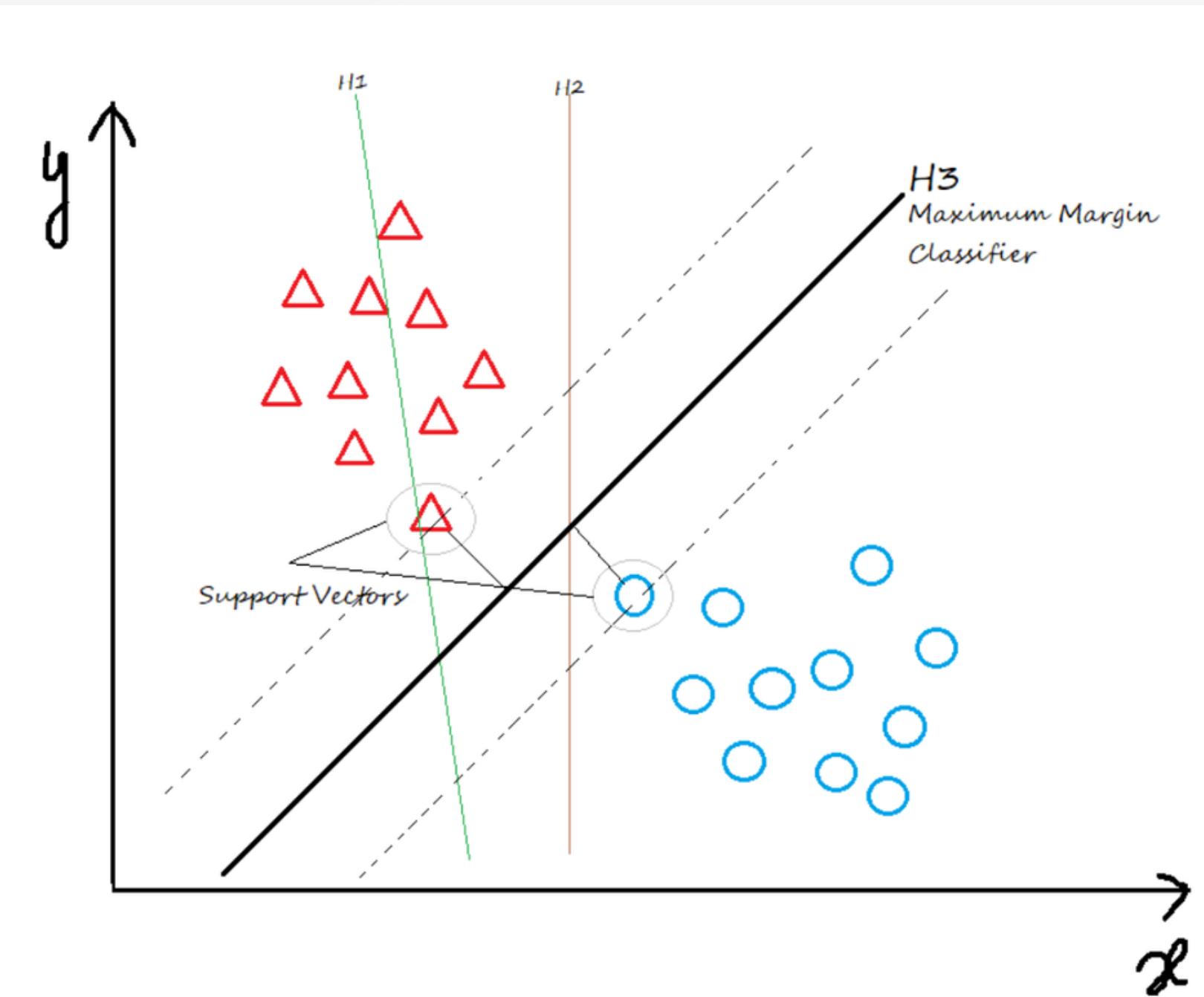
DECISION TREE

algoritma klasifikasi dan regresi yang menggunakan pohon keputusan untuk mengelompokkan data ke dalam kelas-kelas.



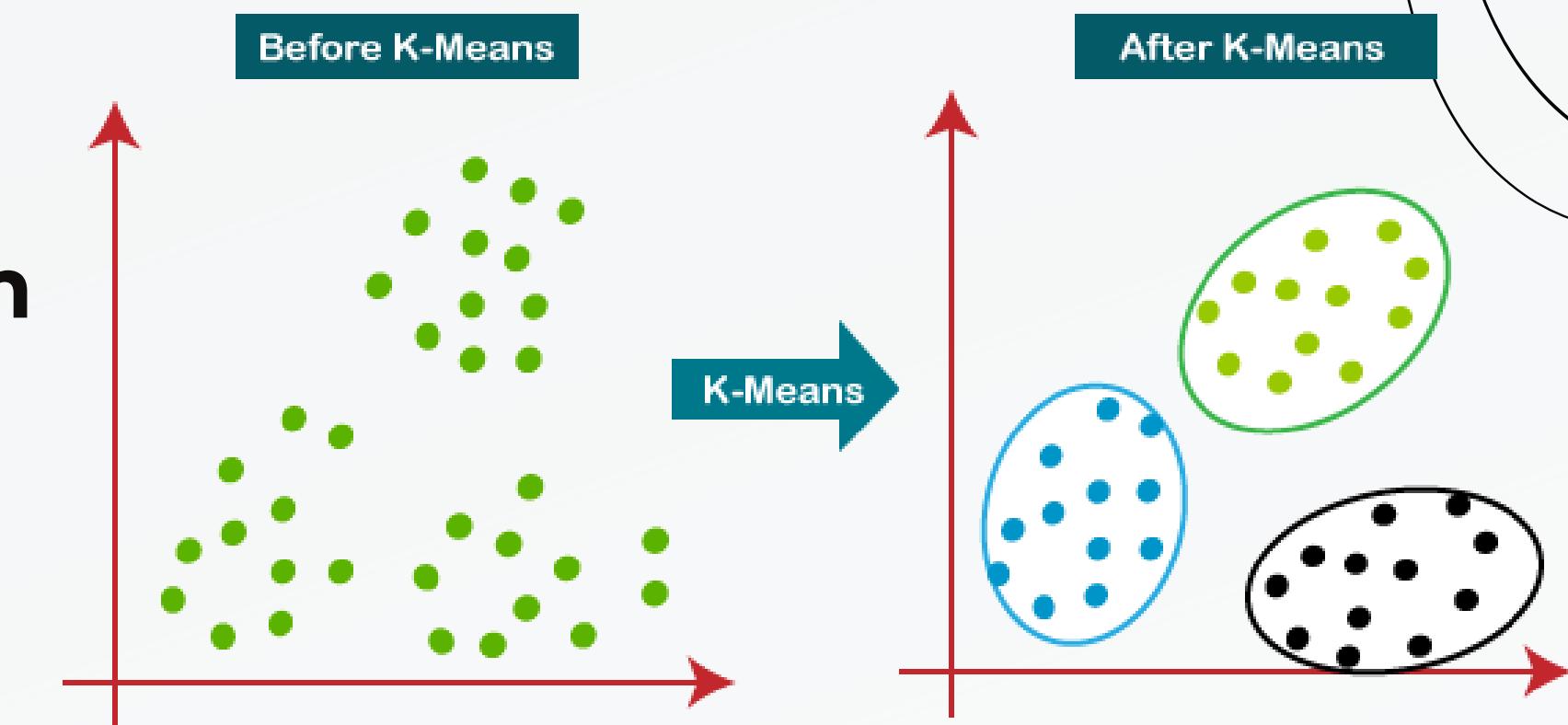
SVM

Support vector machine adalah sekumpulan metode pembelajaran terbimbing yang digunakan untuk klasifikasi, regresi, dan deteksi outlier. yang menggunakan hyperplane untuk memisahkan data ke dalam kelas-kelas.



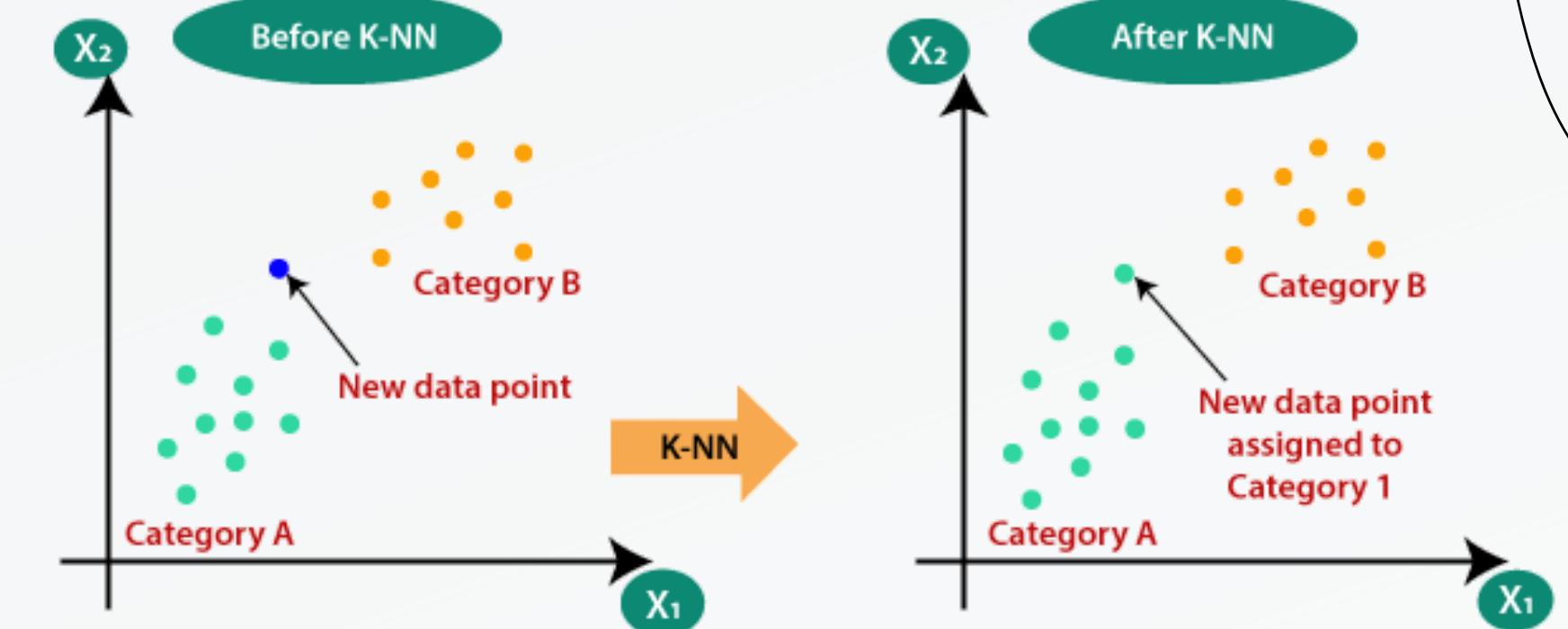
K-MEANS

Algoritma KMeans mengelompokkan data dengan mencoba memisahkan sampel dalam n kelompok dengan varians yang sama, meminimalkan kriteria yang dikenal sebagai kelembaman atau jumlah kuadrat dalam kluster (lihat di bawah). Algoritma ini membutuhkan jumlah cluster yang ditentukan. Ini menskalakan dengan baik untuk sejumlah besar sampel dan telah digunakan di berbagai bidang aplikasi di banyak bidang yang berbeda.



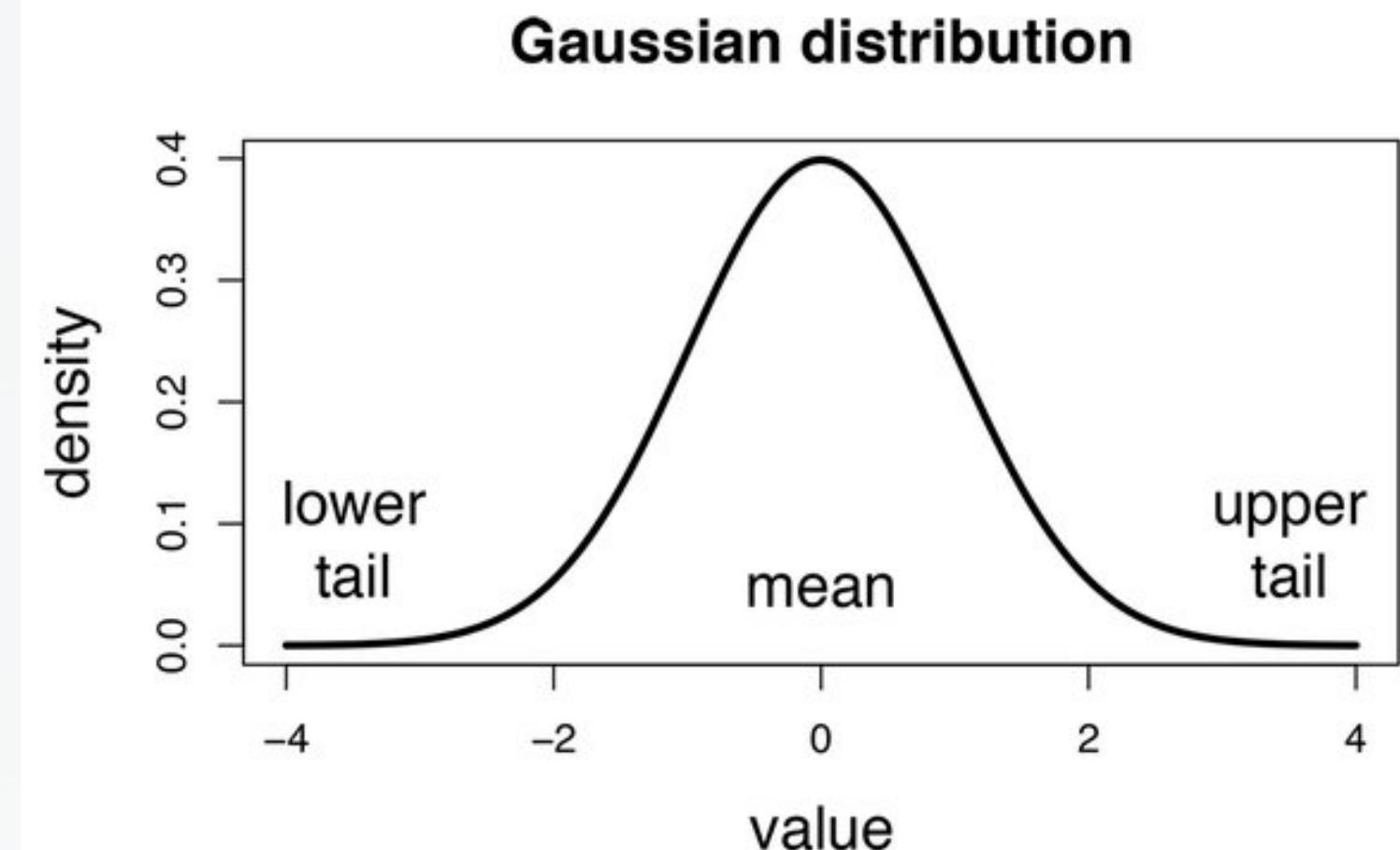
NEAREST NEIGHBOR

algoritma nearest neighbor dapat digunakan untuk mengelompokkan data point ke dalam cluster yang memiliki bentuk yang tidak jelas atau tidak diketahui jumlah clusternya.



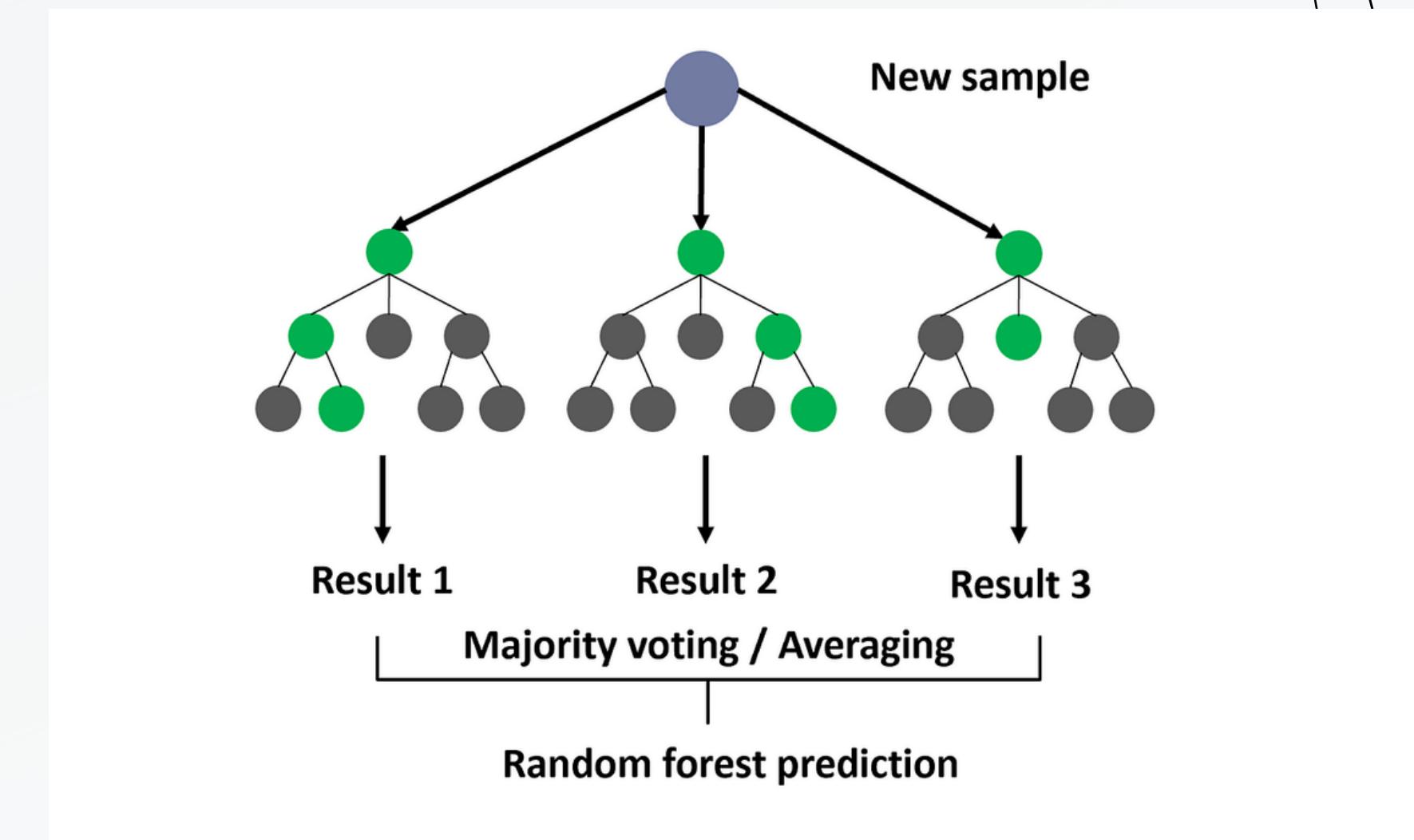
GAUSSIAN NAIVE BAYES

Algoritma Gaussian Naive Bayes adalah salah satu varian algoritma Naive Bayes yang menggunakan distribusi Gaussian untuk menggambarkan probabilitas bersyarat dari setiap fitur yang diberikan kelas.



RANDOM FOREST

Random forest adalah algoritma pembelajaran mesin yang menggabungkan keluaran dari beberapa pohon keputusan untuk mencapai satu hasil. Sesuai namanya, "hutan" dibentuk dari banyak pohon (pohon) yang diperoleh melalui proses bagging atau bootstrap aggregating. Setiap pohon pada Random Forest akan mengeluarkan prediksi kelas.



PRAKTEK RULES



LARANA, INC.

Aturan Praktik Data Science

- **Penggunaan ChatGPT:** ChatGPT boleh digunakan untuk klarifikasi konsep, pemecahan masalah, dan bimbingan umum. Waktu penggunaan ChatGPT dibatasi 30 menit agar fokus pada pemahaman mandiri. Hindari ketergantungan berlebihan pada ChatGPT; gunakan sebagai alat bantu, bukan pengganti pemahaman konsep.
- **Penggunaan Library:** Pemilihan library harus didasarkan pada kebutuhan proyek dan kompatibilitas dengan tujuan data science. Dilarang menyalin dan menempel kode dari dokumentasi tanpa pemahaman menyeluruh.
- Boleh membaca dokumentasi library; fokus pada pemahaman inti dan implementasi.

Aturan Praktik Data Science

- Dokumentasi: Setiap langkah dalam siklus hidup data science harus didokumentasikan dengan baik. Gunakan format dokumentasi yang jelas dan mudah dipahami oleh orang lain. Update dokumentasi saat proyek berkembang atau ada perubahan signifikan.**

**SELAMAT
MENGERJAKAN**



.S.

LARANA, INC.