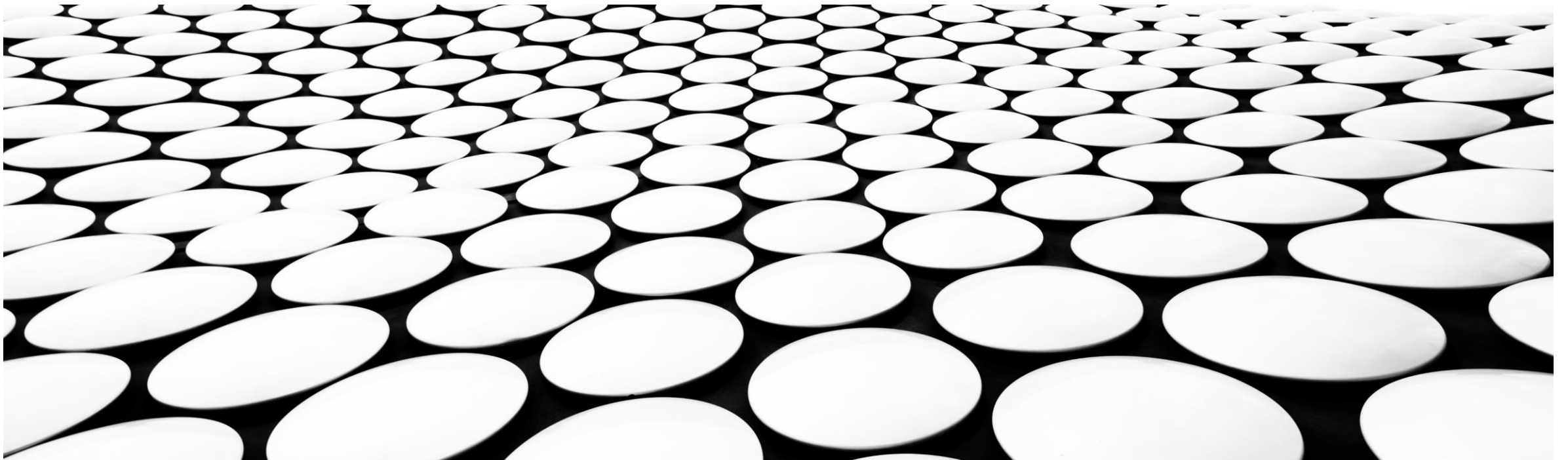

MACHINE LEARNING

MUHAMMAD NURIL HUDA



WHAT IS MACHINE LEARNING ?

- Machine Learning (ML) merupakan ilmu (dan seni) pemrograman komputer yang dapat belajar dari data.
- Bidang studi yang memberi komputer kemampuan untuk belajar tanpa diprogram secara eksplisit (Arthur Samuel, 1959)

Penjelasan yang lebih berorientasi kepada Teknik :

- Sebuah program komputer yang dikatakan belajar dari pengalaman E sehubungan dengan beberapa tugas T dan beberapa ukuran kinerja P. Kinerja pada T, yang diukur dengan P, meningkat dengan pengalaman E (Tom Mitchell, 1997)

WHY USE MACHINE LEARNING ?

ML sangat bagus untuk :

- Masalah yang solusinya membutuhkan banyak *hand-tuning*, satu algoritma ML dapat menyederhanakan kode dan mempunyai kinerja yang lebih baik.
- Masalah kompleks yang tidak ada solusi menggunakan pendekatan tradisional, teknik ML dapat menemukan solusi terbaik.
- Sistem yang fluktuatif, sistem Machine Learning dapat beradaptasi dengan data baru.
- Mendapatkan insight tentang masalah kompleks dan data dalam jumlah besar.

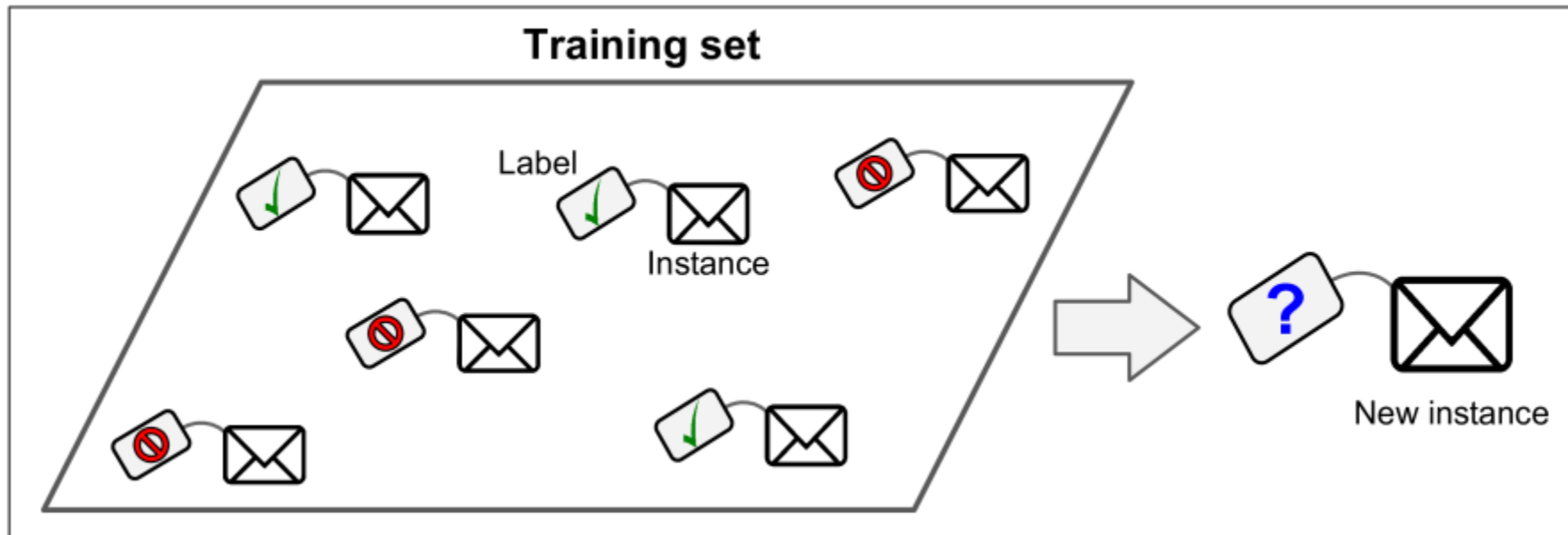
TYPES OF MACHINE LEARNING SYSTEMS


Ada banyak jenis algoritma ML, pengklasifikasiannya didasarkan pada :

1. Apakah dilatih terlebih dahulu atau tidak (supervised, unsupervised, semisupervised, and Reinforcement Learning)
2. Apakah dapat belajar secara bertahap dengan cepat (online vs batch learning)
3. Apakah bekerja hanya dengan membandingkan point data baru dengan point data yang sudah diketahui, atau mendeteksi pola dalam training data dan membangun model prediktif (instance-based vs model-based learning)

SUPERVISED LEARNING

Dalam Supervised Learning, data training yang di masukkan ke algoritma menyertakan solusi yang diinginkan, yang disebut label





Contoh Supervised Learning adalah klasifikasi (classification). Spam filter pada email merupakan contoh algoritma classification, spam filter dilatih dengan banyak contoh email beserta kelasnya (spam atau tidak), dan spam filter harus mempelajari cara mengklasifikasikan email baru.

Contoh lainnya yaitu memprediksi nilai yang bersifat numerik, seperti harga mobil, dengan serangkaian fitur (jarak tempuh, usia, merek, dll.) yang disebut prediktor. Algoritma ini disebut dengan regresi (regression). Untuk melatih sistem, kita perlu memberikan banyak contoh data, termasuk prediktor dan labelnya (harga).

Beberapa algoritma regresi dapat digunakan untuk klasifikasi juga, dan sebaliknya. Misalnya, Logistic Regression biasanya digunakan untuk klasifikasi, karena dapat menampilkan nilai yang sesuai dengan probabilitas (misalnya, 20% kemungkinan termasuk spam).

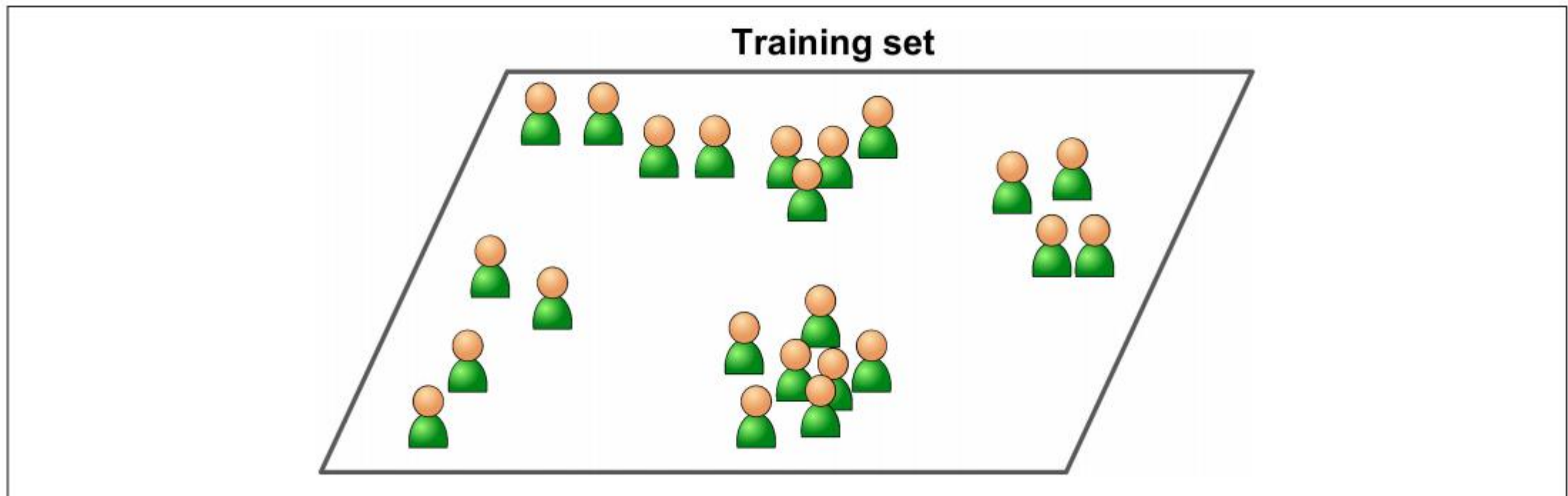


Contoh – contoh algoritma Supervised Learning :

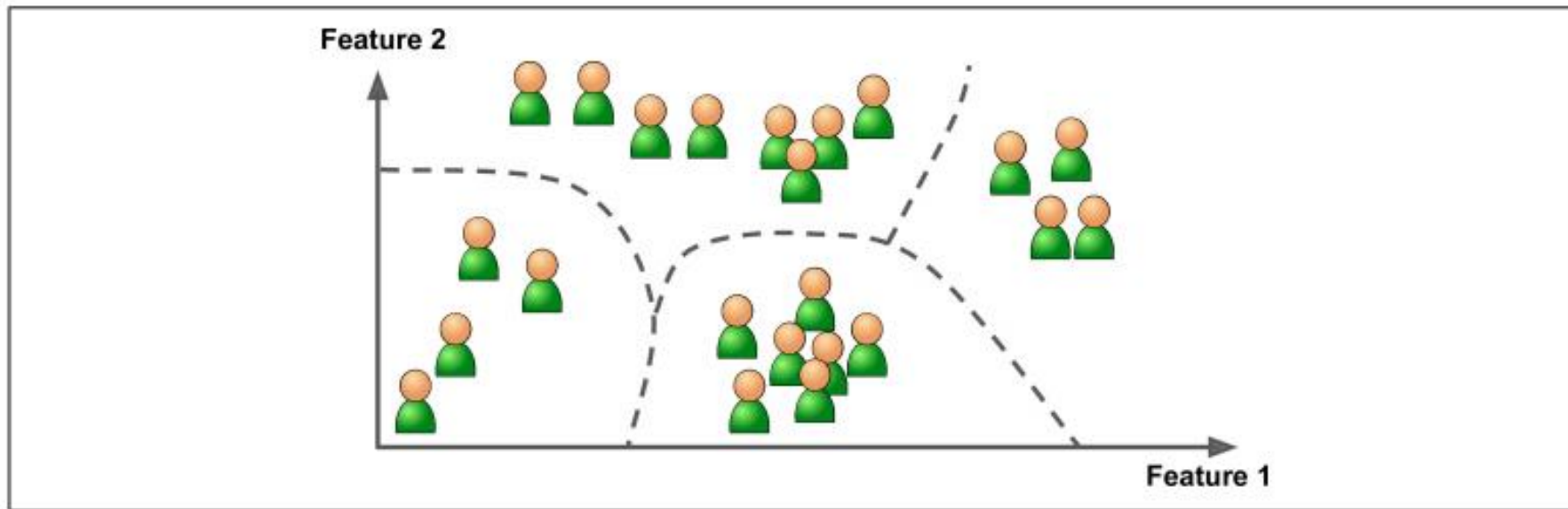
1. k-Nearest Neighbors
2. Linear Regression
3. Logistic Regression
4. Support Vector Machines (SVMs)
5. Decision Trees and Random Forests
6. Neural networks

UNSUPERVISED LEARNING

Dalam Unsupervised Learning, data training tidak berlabel. Sistem mencoba belajar tanpa guru.



Misalnya, data tentang pengunjung blog. Kita mungkin ingin melakukan pengelompokan untuk mencoba mendeteksi kelompok pengunjung yang sama. Kita tidak perlu memberi tahu system tentang kelompok pengunjung. Sistem akan menemukan kelompok pengunjung tanpa bantuan. Misalnya, mungkin 40% pengunjung blog merupakan pria yang menyukai komik dan biasanya membaca blog Anda di malam hari, sedangkan 20% adalah anak muda yang menyukai fiksi ilmiah yang berkunjung selama akhir pekan, dan seterusnya.

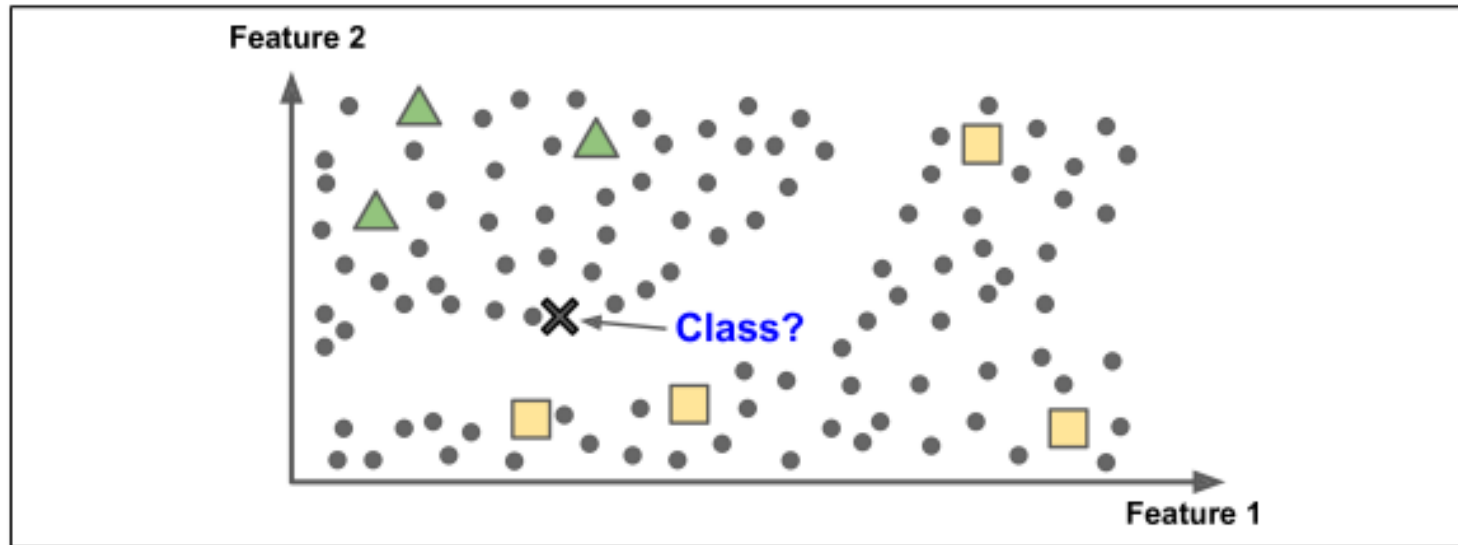



Contoh – contoh algoritma Unsupervised Learning :

1. Clustering : K-Means, DBSCAN, Hierarchical Cluster Analysis (HCA)
2. Anomaly detection and novelty detection : One-class SVM, Isolation Forest
3. Visualization and dimensionality reduction : Principal Component Analysis (PCA), Kernel PCA, Locally-Linear Embedding (LLE), t-distributed Stochastic Neighbor Embedding (t-SNE)
4. Association rule learning : Apriori, Eclat

SEMISUPERVISED LEARNING

Beberapa algoritma dapat menangani sebagian data training yang berlabel, biasanya banyak data tidak berlabel dan sedikit data berlabel. Hal ini disebut dengan Semisupervised Learning





Contohnya beberapa layanan hosting foto, seperti Google Photos, Setelah mengunggah semua foto ke Google Photos, Google Photos secara otomatis mengenali bahwa orang A yang sama muncul di foto 1, 5, dan 11, sementara orang B muncul di foto 2, 5, dan 7. Hal tersebut merupakan bagian dari algoritma unsupervised (clustering). Kemudian sistem meminta memberi tahu siapa nama orang-orang tersebut. Hanya memasukkan satu nama per orang, sistem dapat memberi nama semua orang di setiap foto, yang berguna untuk mencari foto.

Kebanyakan algoritma semisupervised learning adalah kombinasi dari algoritma unsupervised dan supervised. Misalnya, deep belief networks (DBNs) didasarkan pada komponen unsupervised yang disebut restricted Boltzmann machines (RBM) yang di-stack satu sama lain. RBM dilatih secara berurutan dengan cara unsupervised, kemudian seluruh sistem disesuaikan menggunakan supervised learning .