



Institut Informatika & Bisnis  
**DARMAJAYA**  
Yayasan Alfian Husin



**Kampus  
Merdeka**  
INDONESIA JAYA

**MERDEKA  
BELAJAR**

DATA SCIENCE DARMAJAYA  
“YOUR BEST FUTURE IN DATA”

PERTEMUAN KE: 5

# **SUPERVISED LEARNING**

# **KULIAH**

OLEH: NURJOKO



# POKOK BAHASAN

- Pengenalan *Supervised Learning*
- Tipe-Tipe Masalah *Supervised Learning*
- Pemahaman Tentang Data (*training* dan *testing*)
- Fitur dan Label



# Learning Objectives

- Memahami Konsep *Supervised Learning*
- Mampu menjelaskan Tipe-Tipe Masalah *Supervised Learning*
- Memahami Tentang Data (*training* dan *testing*)
- Mampu Menjelaskan Fitur dan Label



## Pengenalan Supervised

- **Supervised Learning adalah salah satu paradigma dalam Machine Learning di mana algoritma mesin (model) "dilatih" menggunakan data yang telah diketahui labelnya.**
- **Dalam metode ini, algoritma belajar untuk membuat prediksi atau mengambil tindakan berdasarkan masukan (input) dan keluaran (output) yang telah ada dalam data pelatihan.**

## Ilustrasi.....



# Konsep Dasar Supervised Learning:

## 1. Data Label

Dalam Supervised Learning, kita bekerja dengan dataset yang terdiri dari dua komponen utama: **data input** dan **label output**.

Data input adalah informasi atau fitur-fitur yang digunakan sebagai dasar untuk membuat prediksi, **sedangkan** label output adalah hasil atau target yang ingin diprediksi.

**Misalnya**, dalam klasifikasi gambar, data input adalah gambar, dan label output adalah kategori yang mengidentifikasi apa yang ada dalam gambar tersebut, seperti "kucing" atau "anjing."

- Untuk melatih model dengan teknik supervised learning, kita membutuhkan **data** berukuran besar beserta **label** yang sesuai.
- Kedua hal tersebut akan digunakan oleh algoritma pembelajaran mesin untuk membuat sebuah rules berdasarkan pola yang terdapat pada data dan label tersebut.
- Rules inilah yang akan digunakan oleh mesin untuk membuat sebuah prediksi.



# Konsep Dasar Supervised Learning:

**2. Pelatihan Model:** Proses pelatihan melibatkan penggunaan dataset pelatihan untuk mengajarkan model (algoritma) bagaimana cara melakukan prediksi. Model tersebut menggali pola atau hubungan antara data input dan label output dalam dataset pelatihan.

**3. Fungsi Tujuan:** Model diarahkan untuk mengoptimalkan suatu fungsi tujuan (objective function) selama proses pelatihan. **Fungsi tujuan** ini mencerminkan sejauh mana prediksi model cocok dengan label yang sebenarnya dalam dataset pelatihan.

**Tujuan utama** adalah untuk meminimalkan selisih antara prediksi dan label.

## Konsep Dasar Supervised Learning:

**4. Generalisasi:** Setelah model dilatih dengan data pelatihan, tujuannya adalah untuk memiliki kemampuan generalisasi. Ini berarti model harus mampu melakukan prediksi yang akurat pada data yang belum pernah dilihat sebelumnya. Dalam pengujian model, kami menguji kemampuan model untuk membuat prediksi yang baik pada dataset yang tidak digunakan dalam pelatihan.

**5. Perbandingan Hasil:** Evaluasi model melibatkan perbandingan prediksi model dengan label yang sebenarnya pada dataset pengujian. Hasil ini digunakan untuk mengukur seberapa baik model mampu memprediksi, dan berbagai metrik evaluasi digunakan untuk mengukur performa model, seperti akurasi, presisi, recall, atau Mean Squared Error (MSE) dalam masalah regresi.

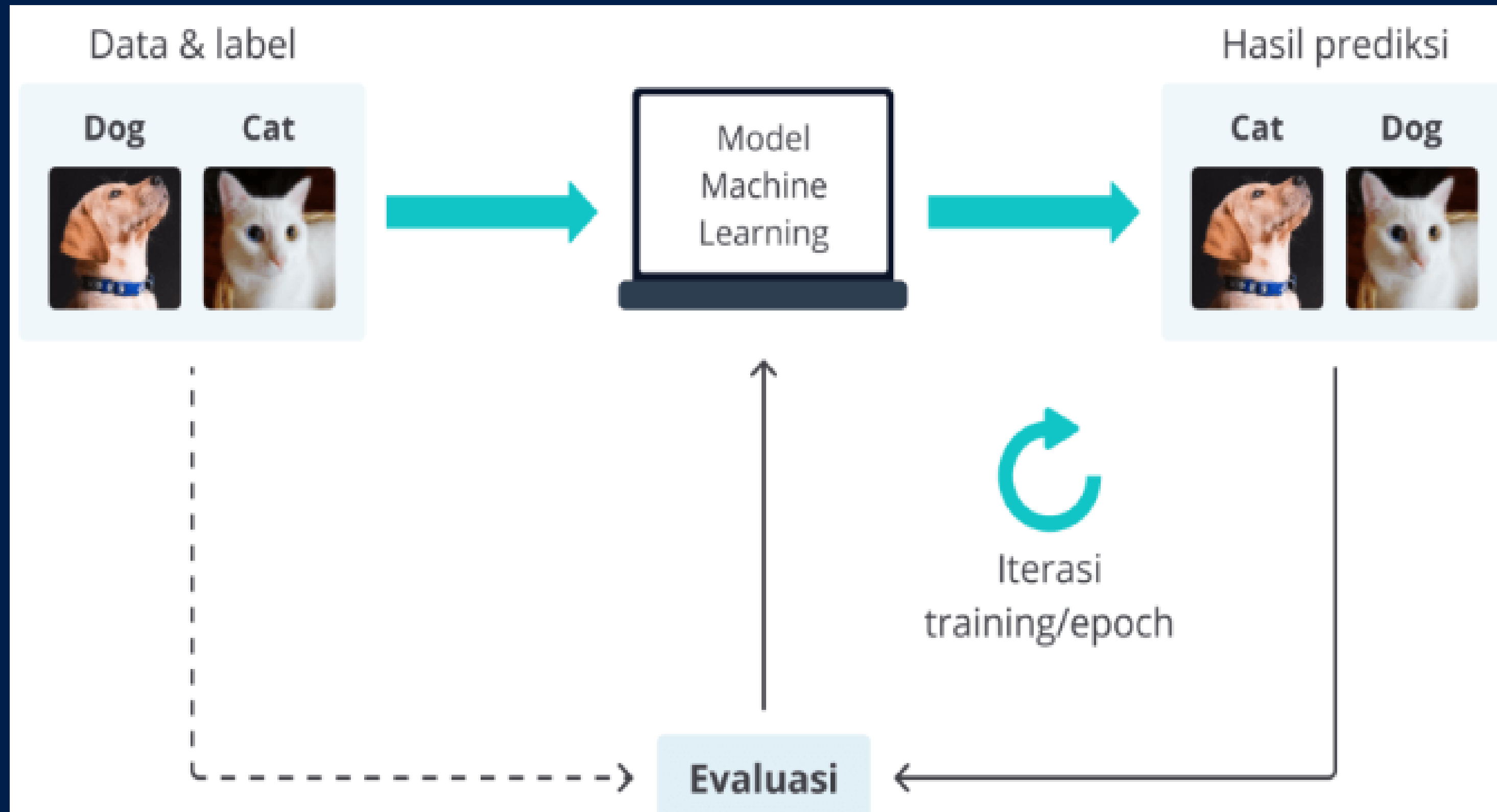


## **Tugas Supervised Learning:**

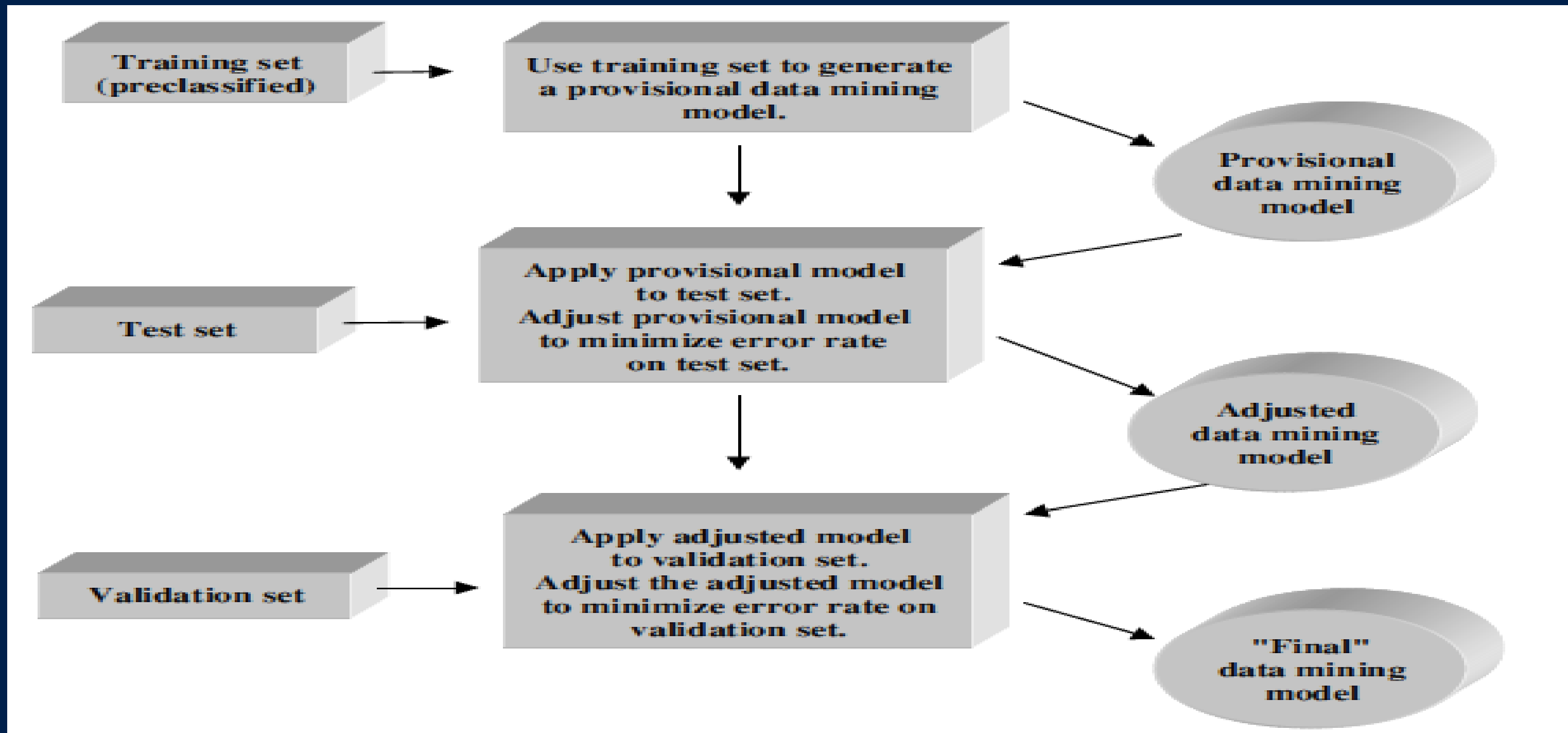
Supervised Learning digunakan dalam berbagai tugas seperti klasifikasi, regresi, pemahaman bahasa alami, pemrosesan citra, dan banyak lagi. Contoh tugas termasuk membedakan email spam dari email non-spam (klasifikasi), memprediksi harga rumah berdasarkan fitur-fiturnya (regresi), atau menerjemahkan teks dari satu bahasa ke bahasa lain (pemahaman bahasa alami).

## **Pentingnya Data Pelatihan:**

Kualitas dan kuantitas data pelatihan sangat penting dalam Supervised Learning. Semakin banyak data pelatihan yang baik, semakin baik performa model dalam membuat prediksi yang akurat.



# Methodology for Supervised Modelling



# Supervised Learning dapat diterapkan pada berbagai tipe masalah seperti:

## 1. Klasifikasi (Classification):

Tipe: Klasifikasi adalah masalah ketika kita mencoba untuk memprediksi kategori atau label dari data input. Misalnya, mengklasifikasikan email menjadi spam atau bukan spam, atau mengidentifikasi jenis hewan berdasarkan gambar.

## 2. Regresi (Regression):

Tipe: Regresi adalah masalah ketika kita mencoba untuk memprediksi nilai kontinu atau numerik sebagai output. Misalnya, memprediksi harga rumah berdasarkan fitur-fitur tertentu, atau memprediksi suhu berdasarkan data cuaca.

### **3. Pemahaman Bahasa Alami (Natural Language Processing, NLP):**

Tipe: Dalam pemahaman bahasa alami, algoritma belajar untuk memahami, menganalisis, dan menghasilkan teks dalam bahasa manusia. Ini mencakup tugas seperti pemodelan bahasa, analisis sentimen, penerjemahan mesin, dan banyak lagi.

### **4. Pemrosesan Citra (Computer Vision):**

Tipe: Pemrosesan citra melibatkan analisis dan interpretasi gambar dan video. Ini dapat mencakup tugas seperti deteksi wajah, pengenalan objek, segmentasi gambar, dan banyak lagi.

## **5. Pengenalan Suara (Speech Recognition):**

Tipe: Dalam pengenalan suara, algoritma belajar untuk mengubah sinyal suara menjadi teks atau perintah yang dapat dipahami oleh komputer. Ini digunakan dalam aplikasi seperti asisten suara (misalnya, Siri atau Google Assistant) dan pengenalan ucapan.

## **6. Rekomendasi (Recommendation):**

Tipe: Rekomendasi melibatkan penyediaan saran atau rekomendasi kepada pengguna berdasarkan preferensi mereka. Ini dapat mencakup sistem rekomendasi film, produk, musik, dan lainnya.

## **7. Anomali Deteksi (Anomaly Detection):**

Tipe: Anomali deteksi adalah identifikasi data yang tidak sesuai dengan pola yang diharapkan atau perilaku aneh. Ini digunakan dalam deteksi fraud, keamanan jaringan, dan pemeliharaan mesin.



## 8. **Biologi Komputasional (Computational Biology):**

Tipe: Dalam biologi komputasional, Supervised Learning digunakan untuk memahami struktur protein, analisis sekuens DNA, prediksi interaksi obat-protein, dan banyak lagi.

## 9. **Finansial (Financial):**

Tipe: Dalam industri keuangan, Supervised Learning digunakan untuk peramalan harga saham, analisis kredit, pengawasan transaksi keuangan, dan lainnya.

## 10. **Medis (Medical):**

Tipe: Dalam bidang medis, Supervised Learning digunakan untuk diagnosis penyakit, penggalian data medis, dan penentuan perawatan berdasarkan data pasien.



## Data Pelatihan (Training Data)

**Definisi:** Data pelatihan adalah sebagian besar data yang digunakan untuk melatih model atau algoritma mesin. Data ini telah diketahui labelnya, yang berarti kita tahu keluaran yang benar atau target yang harus diprediksi untuk setiap sampel data.

**Fungsi:** Data pelatihan digunakan untuk mengajarkan model atau algoritma bagaimana melakukan prediksi yang benar. Model belajar dari data ini untuk menemukan pola dan hubungan antara data input dan label output.

**Proporsi:** Biasanya, sebagian besar data (sekitar 70-80%) digunakan sebagai data pelatihan.



## Data Pengujian (Testing Data):

**Definisi:** Data pengujian adalah bagian dari dataset yang tidak digunakan selama pelatihan model. Seperti data pelatihan, data pengujian juga memiliki label, tetapi label ini tidak digunakan selama pelatihan model.

• **Fungsi:** Data pengujian digunakan untuk menguji sejauh mana model yang telah dilatih mampu melakukan prediksi yang akurat pada data yang belum pernah dilihat sebelumnya.

**Proporsi:** Sebagian kecil data (sekitar 20-30%) dialokasikan sebagai data pengujian.

# Fitur dan Label

"fitur" (features) dan "label" adalah dua komponen penting dalam dataset. Mereka memiliki peran yang berbeda dan membentuk dasar untuk melatih dan menguji model machine learning

## 1. Fitur (Features):

**Definisi:** Fitur adalah atribut atau variabel yang digunakan sebagai input untuk model machine learning. Fitur adalah representasi dari data yang membantu model memahami dan membuat prediksi. Fitur dapat berupa angka, teks, gambar, atau jenis data lainnya tergantung pada jenis masalah yang diselesaikan.

**Contoh:** Dalam masalah klasifikasi gambar kucing atau anjing, fitur-fitur dapat berupa piksel-piksel dalam gambar. Dalam prediksi harga rumah, fitur-fitur dapat mencakup jumlah kamar, luas tanah, lokasi, dan sebagainya.

## 2. Label:

### Definisi:

- Label adalah nilai target atau keluaran yang ingin diprediksi oleh model. Label adalah hasil yang diharapkan atau yang ingin ditebak oleh model machine learning. Dalam supervised learning, dataset selalu berisi label yang sesuai dengan setiap contoh data.

### Contoh:

- Dalam masalah klasifikasi gambar kucing atau anjing, label-labelnya mungkin adalah "kucing" atau "anjing." Dalam prediksi harga rumah, labelnya adalah harga sebenarnya dari rumah-rumah dalam dataset.

## Hubungan antara Fitur dan Label:

- Fitur dan label bekerja sama dalam supervised learning. Model machine learning memahami pola dan hubungan antara fitur dan label dalam data pelatihan.
- Fitur digunakan sebagai input model untuk membuat prediksi.
- Label digunakan sebagai hasil yang diharapkan, dan model berusaha untuk meminimalkan selisih antara prediksi dan label dalam proses pelatihan.
- **Contoh Penggunaan Fitur dan Label:** Misalkan kita ingin membangun model untuk mengklasifikasikan email sebagai "spam" atau "bukan spam."

- **Fitur:** Fitur-fitur dalam kasus ini mungkin mencakup kata-kata dalam email, frekuensi kata tertentu, atau atribut lain yang dapat membantu model memahami konten email.
- **Label:** Label dalam kasus ini adalah kategori "spam" atau "bukan spam" yang telah ditentukan untuk setiap email dalam dataset.
- Model belajar dari data pelatihan yang berisi contoh-contoh email dengan fitur-fitur dan label-labelnya. Setelah melatih model, kita dapat menggunakan fitur-fitur email baru yang belum pernah dilihat sebelumnya oleh model untuk memprediksi apakah email tersebut "spam" atau "bukan spam."



# CONCLUSION

Fill in IEEE Style



# REFERENCES

Fill in IEEE Style



Institut Informatika & Bisnis  
**DARMAJAYA**  
Yayasan Alfian Husin



**Kampus  
Merdeka**  
INDONESIA JAYA

**MERDEKA  
BELAJAR**

# THANK YOU!!

DATA SCIENCE DARMAJAYA "YOUR BEST FUTURE IN DATA"