



EduAPM

PSE LINGKUP  
PRIVAT

1/10

# Data adalah Bahan Bakar AI

## Mengenal Fungsi Data dalam Pembelajaran Mesin

Tanpa data yang tepat,  
model AI tidak bisa belajar  
dengan baik.



Sumber Data



Kualitas



Label



Pelatihan



Evaluasi



Data bukan sekadar banyak—  
data harus **relevan, bersih,**  
dan **mewakili kondisi nyata.**



Kumpulkan Data



Latih Model



Hasilkan Insight



Geser untuk memahami bagaimana  
data membuat AI belajar.





# Mengapa Data Disebut Bahan Bakar AI?

Model AI belajar dari contoh. Semakin relevan dan berkualitas datanya, semakin baik hasil belajarnya.



Data Masuk

01



Pola Terdeteksi

02



Model Belajar

03



Prediksi

04



Insight

05



Algoritma adalah mesin, tetapi **data adalah energi** yang membuatnya bekerja.



Contoh



Pola



Keputusan



Berikutnya: kenali **jenis-jenis data** yang dipakai AI.





EduAPM

PSE

LINGKUP  
PRIVAT

3/10

# Jenis Data dalam AI

AI bisa belajar dari angka, teks, gambar, suara, dan sinyal lainnya.



**Tabel/Angka**  
transaksi



**Teks**  
chat/email



**Gambar**  
foto/video



**Audio**  
suara



**Sensor/Log**  
aktivitas sistem



Bentuk data yang berbeda membutuhkan **cara olah yang berbeda** pula.



Data Terstruktur



Semi-terstruktur



Tidak Terstruktur



Berikutnya: lihat alur data sebelum melatih model.





# Alur Data Sebelum Model Belajar

Data yang baik lahir dari proses yang rapi, bukan dari kebetulan.



Kumpulkan  
source

01



Bersihkan  
cleaning

02



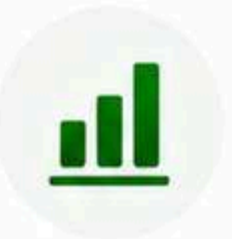
Labeli  
annotate

03



Latih  
training

04



Evaluasi  
testing

05



Di dunia nyata, tahap **pembersihan data** sering memakan waktu paling besar.



Sumber



Proses



Hasil



Berikutnya:  
kualitas data lebih penting dari sekadar jumlah.





# Kualitas Data > Kuantitas Data

Data banyak belum tentu berguna jika salah, kotor, atau tidak relevan.



## Data Berkualitas

- ✓ Akurat
- ✓ Bersih
- ✓ Relevan
- ✓ Konsisten
- ✓ Lengkap



## Data Tidak Berkualitas

- ✗ Salah
- ✗ Duplikat
- ✗ Tidak Relevan
- ✗ Tidak Konsisten
- ✗ Tidak Lengkap



### Relevan

Sesuai dengan tujuan dan konteks.



### Bersih

Bebas duplikasi, noise, dan error.



### Lengkap

Tidak ada data penting yang hilang.



### Konsisten

Format, nilai, dan definisi seragam.



### Beragam

Mewakili variasi kondisi nyata.



**10.000 data yang tepat** sering lebih bernilai daripada **1 juta data** yang kacau.



Benar



Rapi



Mewakili



Berikutnya: apa fungsi label dan anotasi data?





# Label & Anotasi Data

Model **supervised learning** belajar dari contoh yang sudah diberi jawaban.

Anotasi yang akurat membuat model memahami konteks dan mengambil keputusan dengan tepat.

## Tagging

#pantai #senja #alam



## Labeling

☒ Mobil  
☐ Motor  
☐ Bus



## Sentimen



Positif



### Spam / Bukan

Menandai email sebagai spam atau bukan spam.



### Kucing / Bukan

Menentukan apakah gambar berisi kucing.



### Sentimen

Menilai apakah teks bernuansa positif, netral, atau negatif.



### Objek Gambar

Menandai dan memberi label objek di dalam gambar.



### Kategori Produk

Mengelompokkan produk ke dalam kategori tertentu.



**Jika label salah,** model juga berisiko belajar dengan cara yang salah.



Data



Label



Belajar



Berikutnya: bagaimana data dibagi untuk **training** dan **testing**?





# Membagi Data untuk Belajar

Agar model tidak hanya menghafal, data perlu dipisah untuk training, validasi, dan testing.



Training Set

01



Validation

02



Testing

03



Cek Akurasi

04



Siap Pakai

05

- Pisahkan data sejak awal agar model belajar, disetel, dan diuji secara adil.



Pemakaian data uji membantu kita menilai apakah model benar-benar **siap menghadapi data baru**.



Belajar



Tuning



Uji



Berikutnya:  
kenali risiko bias, imbalance, dan privasi.





# Risiko dalam Data AI

Bias, ketidakseimbangan, dan privasi adalah isu penting yang tidak boleh diabaikan.



## Bias

Data bias dapat menghasilkan model yang tidak adil.



## Imbalance

Data tidak seimbang membuat model condong.



## Duplikasi

Data ganda dapat memengaruhi hasil pelatihan.



## Noise

Data bising menurunkan performa dan akurasi model.



## Privasi

Kebocoran data dapat melanggar privasi dan kepercayaan.



Data yang tidak mewakili kondisi nyata dapat menghasilkan AI yang tidak **adil** atau **tidak akurat**.



Adil



Aman



Bertanggung Jawab



Berikutnya: data juga bisa berubah seiring waktu.





EduAPM

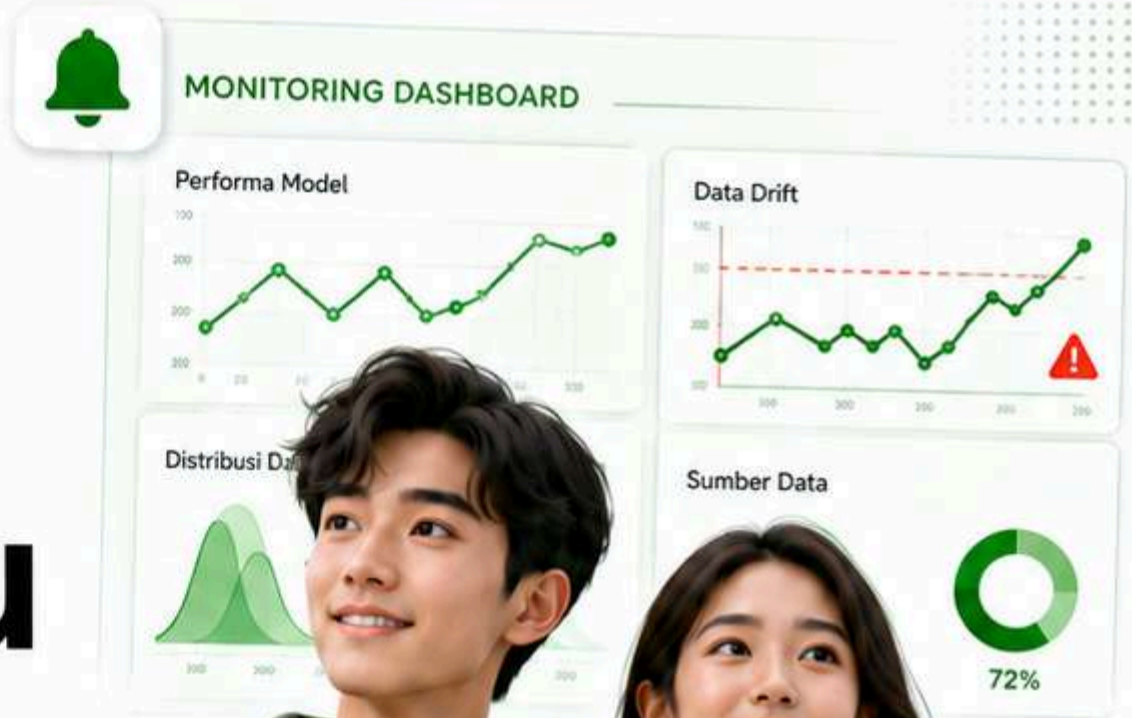
PSE

LINGKUP  
PRIVAT

9/10

# Saat Data Berubah, Model Perlu Dipantau

Perilaku pengguna, bahasa, pasar, dan lingkungan dapat berubah dari waktu ke waktu.



Data Baru

01



Bandingkan

02



Pantau  
Performa

03



Deteksi Drift

04



Update Model

05



Model AI bukan proyek sekali jadi—ia perlu **evaluasi** dan **pembaruan berkala**.



Monitor



Analisis



Perbarui



Terakhir: simpulkan mengapa data adalah fondasi AI.





EduAPM

PSE

LINGKUP  
PRIVAT

10/10

# Kesimpulan: Data adalah Fondasi AI

AI yang baik lahir dari data yang relevan, bersih, aman, dan terus dipantau.



## Relevan

Sesuai tujuan dan konteks masalah.



## Bersih

Akurat, lengkap, dan bebas noise.



## Berlabel

Informasi jelas, mudah dipahami.



## Teruji

Kualitas data tervalidasi.



## Dipantau

Terus dievaluasi dan diperbarui.



Mulailah dari **memahami data**, lalu bangun **keterampilan AI** secara bertahap.



Pahami Data



Latih Model



Bangun Portofolio



Simpan carousel ini dan ikuti EduAPM untuk belajar AI lebih dalam.

