

# Data Mining

Teknik Dimensionality Reduction

# Dimensional Reduction

- Proses mengurangi jumlah variabel acak yang sedang dipertimbangkan dengan memperoleh satu set variabel utama.

# Bagaimana Menyederhanakannya?

- Feature Selection
- Feature Extraction

# Feature itu apa?

- Fitur atau ciri → yang bisa digunakan untuk menyimpulkan atau membedakan sesuatu.

# Feature Selection

- Disebut juga variable selection, attribute selection atau variable subset selection.
- Proses pemilihan subset fitur yang relevan (variabel, prediktor/penebak) untuk digunakan dalam konstruksi model.

# Tujuan Feature Selection

- Penyederhanaan model untuk membuatnya lebih mudah diinterpretasikan oleh peneliti / pengguna.
- Waktu pelatihan yang lebih singkat.
- Untuk menghindari kutukan dimensi.
- Peningkatan generalisasi dengan mengurangi overfitting (secara formal, pengurangan varian)

# Diskusi 1.1

- Bagaimana kalian membedakan :
  - Makanan enak dan tidak enak?
  - Ruangan yang luas dan sempit?
  - Suara berisik dan sunyi?
  - Musik yang kalian suka dan tidak?
  - Ucapan yang sopan dan tidak sopan?

# Diskusi 1.2

- Apa fitur dari :
  - Gambar / foto
  - Video
  - Suara
  - Musik
  - Kalimat

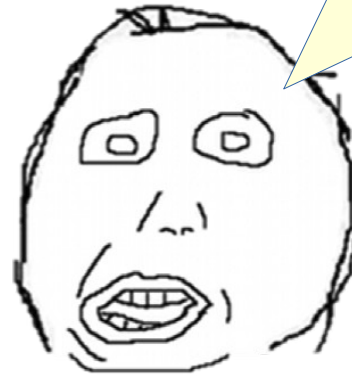


# Bagaimana Fitur si Herp?

Selain yang terakhir  
Mungkin perlu  
Di validasi  
Kebenarannya...



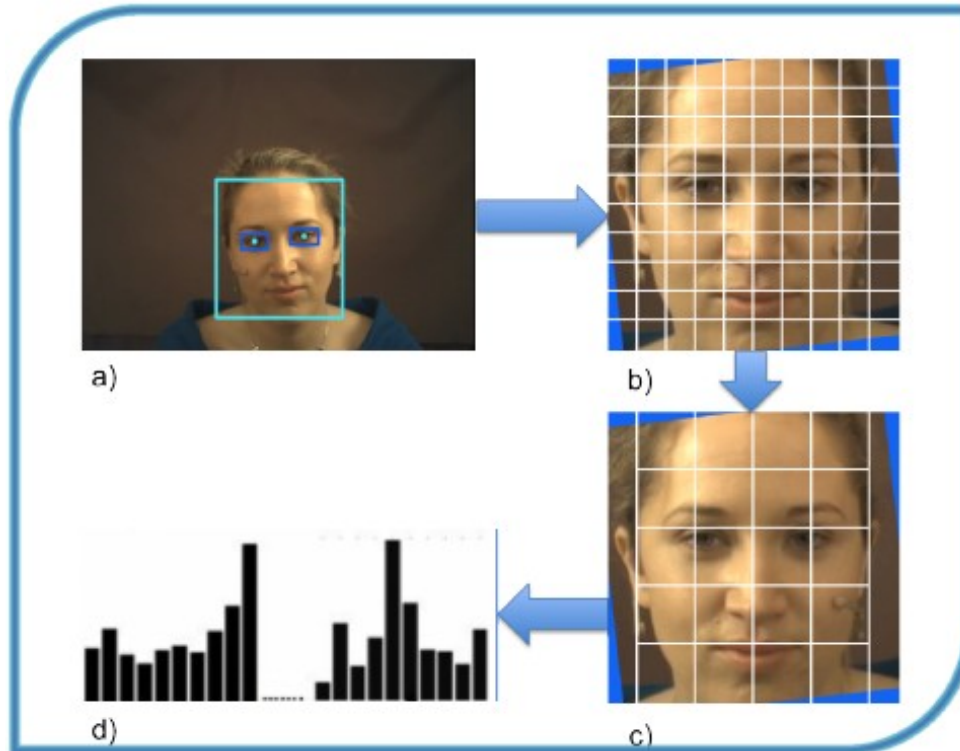
Wajah : Tamvan  
Bodi : Macho  
Moral : Baik  
Kecerdasan : Jenius  
Prilaku : Sopan santun  
Status : Jomblo



# Fitur Gambar / Foto

- Ukuran (Dimensi)
- Warna (Proporsi)
  - RGB, CMYK, Greyscale, Monochrome
- Edge (Batasan garis)
- Skala (dot / cm<sup>2</sup>)

# Fitur Wajah



# Fitur Video

- Ukuran (Dimensi)
  - 4k, full HD, HD, VGA
- Frame per second
  - 240fps, 120fps, 60fps, 30fps
- Durasi

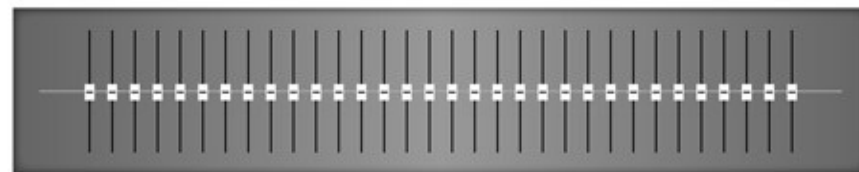
# Fitur suara

- Amplitudo (Volume)
- Frekuensi (Tinggi rendah suara)
- Pitch (umumnya hanya untuk suara manusia)
- Durasi

# Fitur Suara



Parametric Equalizer



Graphic Equalizer

Figure 42: Parametric and Graphic Equalizers

# Fitur Musik

- Chord
- Chords progression
- Notation
- BPM (Beats per minutes)





# Fitur Kalimat

- Subjek
- Predikat
- Objek
- Keterangan

# Feature Extraction

- Proses reduksi dimensi, di mana set awal variabel mentah direduksi menjadi kelompok yang lebih mudah dikelola (fitur) untuk diproses, sementara masih secara akurat dan lengkap menggambarkan set data asli.

# Contoh Feature Extraction

- Mengubah kontras warna dan menjadikan gambar monochrome untuk memudahkan deteksi edge fitur wajah.
- Menyamakan pitch suara dan durasi untuk memudahkan deteksi ucapan tertentu.

# Berkat Feature Extraction

Setelah ekstraksi  
Fitur wajah pada  
rekaman CCTV,  
anda terbukti tidak bayar  
Di kantin tanggal xx  
Pukul yy

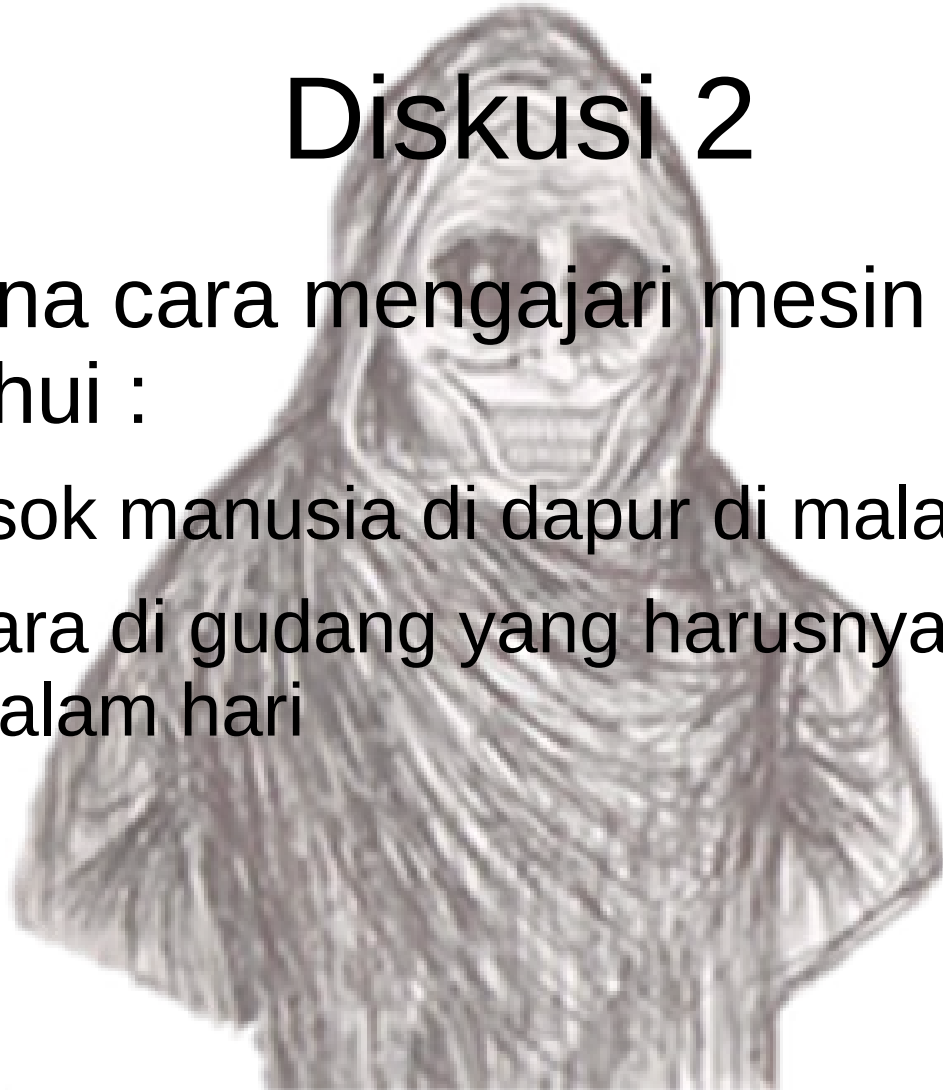


Waduh canggih amat  
Kantin jaman now  
Tobat dah saya  
Mulai sekarang...



# Diskusi 2

- Bagaimana cara mengajari mesin untuk mengetahui :
  - Ada sosok manusia di dapur di malam hari
  - Ada suara di gudang yang harusnya tidak dihuni pada malam hari



# Konsep ML

- Jika ingin mendeteksi penyusup perlu dilatih kondisi :
  - Citra ruangan dapur kosong
  - Citra ruangan dapur dengan sosok manusia
  - Suara di gudang saat sepi
  - Suara yang dianggap tidak wajar

# Kenapa perlu Training Data?

- Kita tidak bisa menggambarkan bagaimana bentuk manusia dengan **bahasa mesin** TETAPI kita menunjuk foto mana yang berisi manusia dan tidak.
- Kita tidak bisa mengatakan tidak akan ada suara sama sekali di gudang ( $\text{amplitudo}=0$ ) TETAPI kita bisa memberi contoh suara yang tidak kita inginkan seperti pintu terbuka, suara manusia, barang jatuh dll.

# Diskusi 3.1

- Pilih salah satu dari topik berikut :
  - Klasifikasi kesuburan tanaman (IoT)
  - Deteksi suara alat musik band umum : gitar, bass, piano, drum (Suara)
  - Pengenalan umur berdasarkan citra wajah (Citra)
  - Klasifikasi bahasa jurnalisme dan percakapan (Bahasa)



# Diskusi 3.2

- Dari topik yang kalian pilih pertimbangkan :
  - Data apa yang harus dipersiapkan?
  - Bagaimana kalian membedakan data tersebut sebagai manusia?
  - Fitur apa yang kalian pilih dari data tersebut?
  - Bagaimana kalian melakukan ekstraksi fitur dari data tersebut?