IF5181 Pengenalan Pola

Pendahuluan

Masayu Leylia Khodra

Outline

- Pengantar kuliah IF5181
- Pengenalan Pola:
    - Definisi pola, kategori pola, pengenalan pola
    - Aplikasi pengenalan pola
    - Pendekatan pengenalan pola

Pengantar Kuliah IF5181 (3 sks)

- Pengajar:
  - Mgg1 sd UTS - Masayu Leylia K (masayu@informatika.org)
  - Mgg9 sd UAS - Nugraha Utama (utama@informatika.org)
- Tatap muka: Selasa 10.00-12.00; Rabu 10.00-11.00
- Mailing-list: if5181@students.if.itb.ac.id
- Situs Kuliah: TBD
- Materi: https://bit.ly/2ZhiC10
  https://www.dropbox.com/sh/fa0ahdosyjntify/AADH3hwWrG6ILc5uc-jmZBL8a?dl=0

Luaran Kuliah

1. Mahasiswa mampu mendeskripsikan dan membangun modul untuk membangun suatu sistem pengenalan pola.
2. Mahasiswa mampu mendefinisikan metode/ teknik pengenalan pola yang sesuai untuk kasus tertentu.
3. Mahasiswa mampu membangun secara utuh suatu sistem pengenalan pola untuk suatu kasus.
4. Mahasiswa mampu mengevaluasi kinerja suatu sistem pengenalan pola.
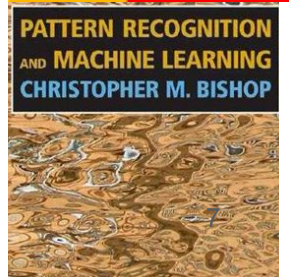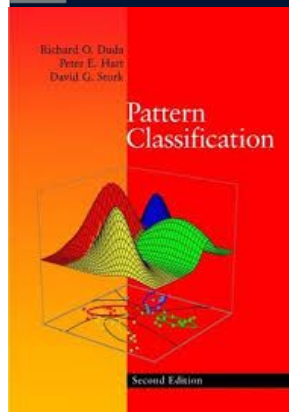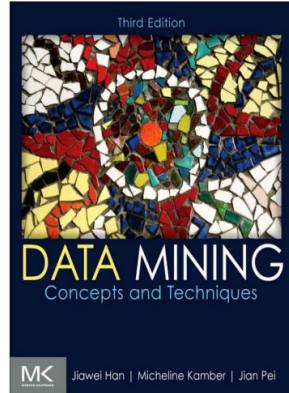
Materi Kuliah Mgg1-7

1. Pendahuluan
2. Classification
3. Clustering
4. Outlier Detection
5. Mining Frequent Patterns, Associations, and Correlations
6. Mining Stream, Time-Series, and Sequence Data
7. Pengenalan Pola pada Teks

Penilaian

- Komponen: UTS, UAS, Kuis, PR & latihan, Tugas
- Syarat lulus IF5181: nilai setiap komponen > 0
- Segala bentuk kecurangan akan mengakibatkan nilai E
- Susulan UTS UAS hanya untuk mahasiswa yang dirawat di rumah sakit atau terkena musibah. Susulan hanya untuk UTS atau UAS (tidak keduanya)
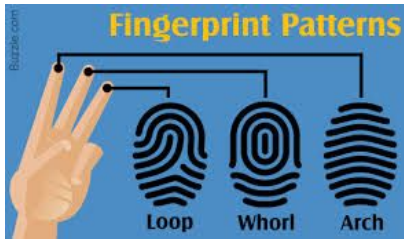
Referensi

- Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.
- R. Duda, P. Hart, and D. Stork, Pattern Classification, 2nd ed. John Wiley & Sons, 2001
- C. Bishop, Pattern Recognition and Machine Learning. Springer, 2006
- Digital library / http://scholar.google.com/
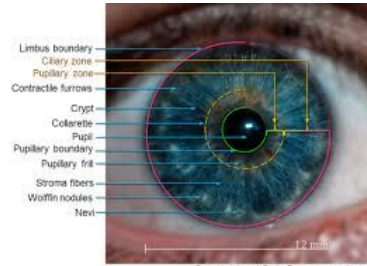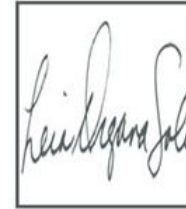
Pola / Pattern

Pola biometrik

Fingerprint Patterns

Loop   Whorl   Arch

https://sciencestruck.com/identifying-types-of-fingerprints-patterns

Palmprint

Limbus boundary
Ciliary zone
Pupillary zone
Contractile furrows
Crypt
Collarette
Pupil
Pupillary boundary
Pupillary frill
Stroma fibers
Wolffin nodules
Nevi

12 mm

Source (eye image): Dr. Jan Drewes www.jandrewes.de

https://indranilsinharoy.com/2014/12/05/dissertation_series/
https://www.bayometric.com/biometric-system-architecture/

Signature

Pola pada teks

```python
import re
email = re.compile('\w+@\w+\.[a-z]{3}')
text = "To email Guido, try guido@python.org or the older address guido@google.com."
email.findall(text)
```

```
['guido@python.org', 'guido@google.com']
```

Pola: Definisi

- Gabungan fitur yang merupakan karakteristik dari suatu individu
- Pasangan kumpulan observasi (features) and konsepnya (labelnya).
  - Feature: any distinctive aspect, quality or characteristic which, may be symbolic (i.e., color) or numeric (i.e., height).
  - feature vector: combination of d features, that is represented as a d-dimensional column vector.
  - feature space: d-dimensional space defined by the feature vector

Kategori / Kelas Pola

● Kumpulan objek yang "mirip" (tidak harus identik)
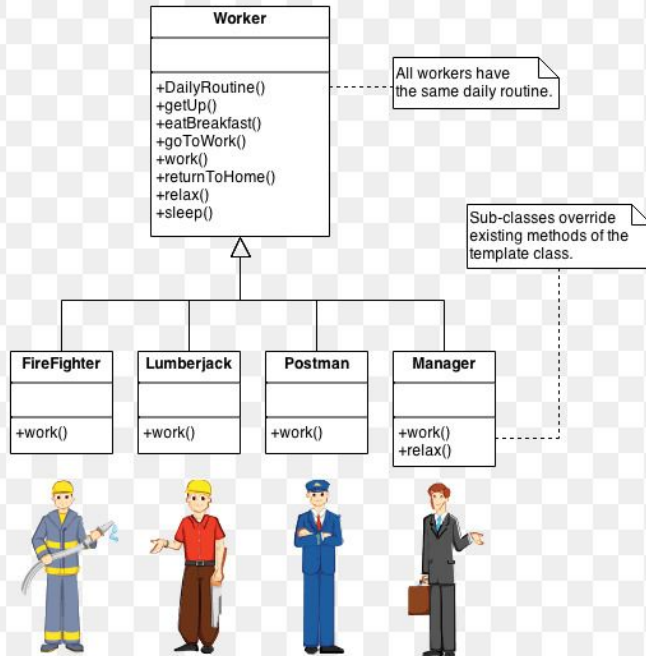
Manusia: Konsep Semantik Tingkat Tinggi

Citra	Label	Deskripsi label
	Apel	*a round fruit with firm, white flesh and a green, red, or yellow skin*
	Lemon	*a yellow, oval citrus fruit with thick skin and fragrant, acidic juice*
	Pisang	*a long curved fruit that grows in clusters and has soft pulpy flesh and yellow skin when rip*
	Belimbing	*a golden-yellow juicy fruit with a star-shaped cross section*

Variasi Intra Class

Penampakan yang berbeda dari sejumlah objek berbeda dari kategori yang sama
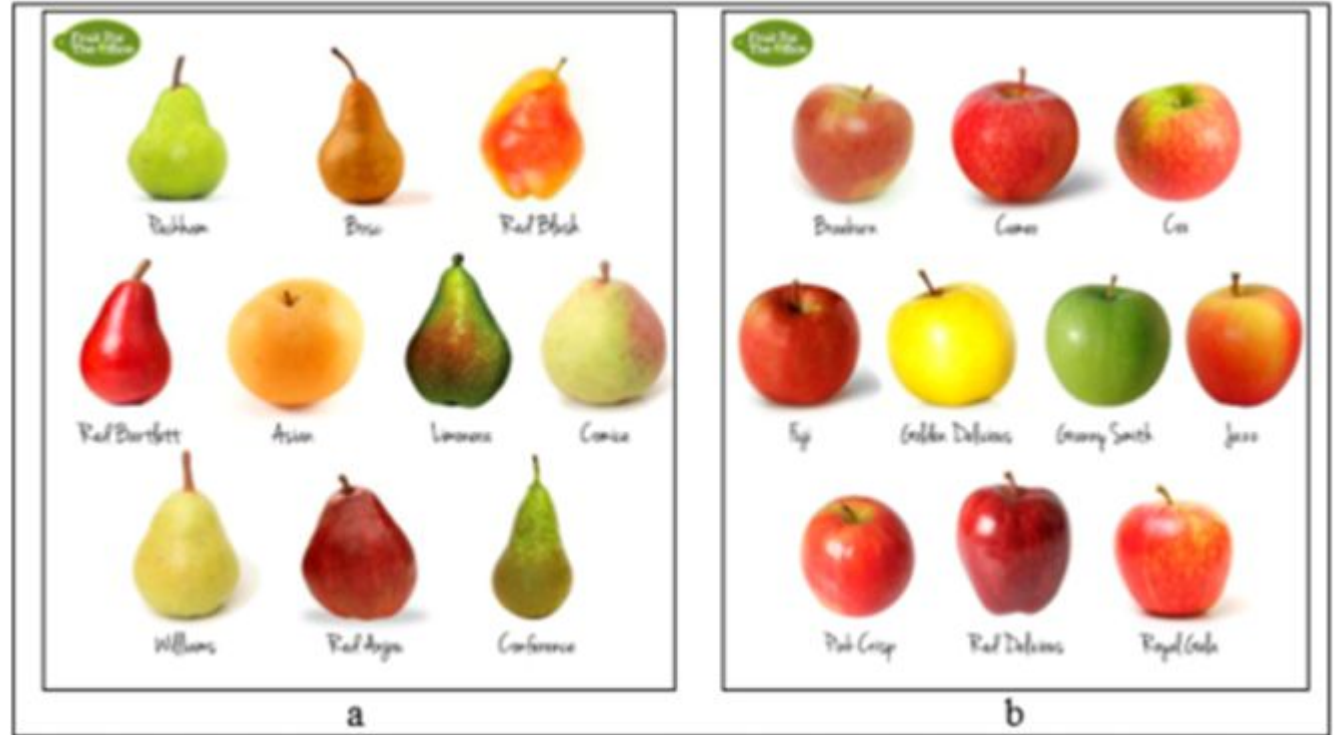
Variasi Inter Class

Penampakan yang mirip pada objek yang berbeda kategori
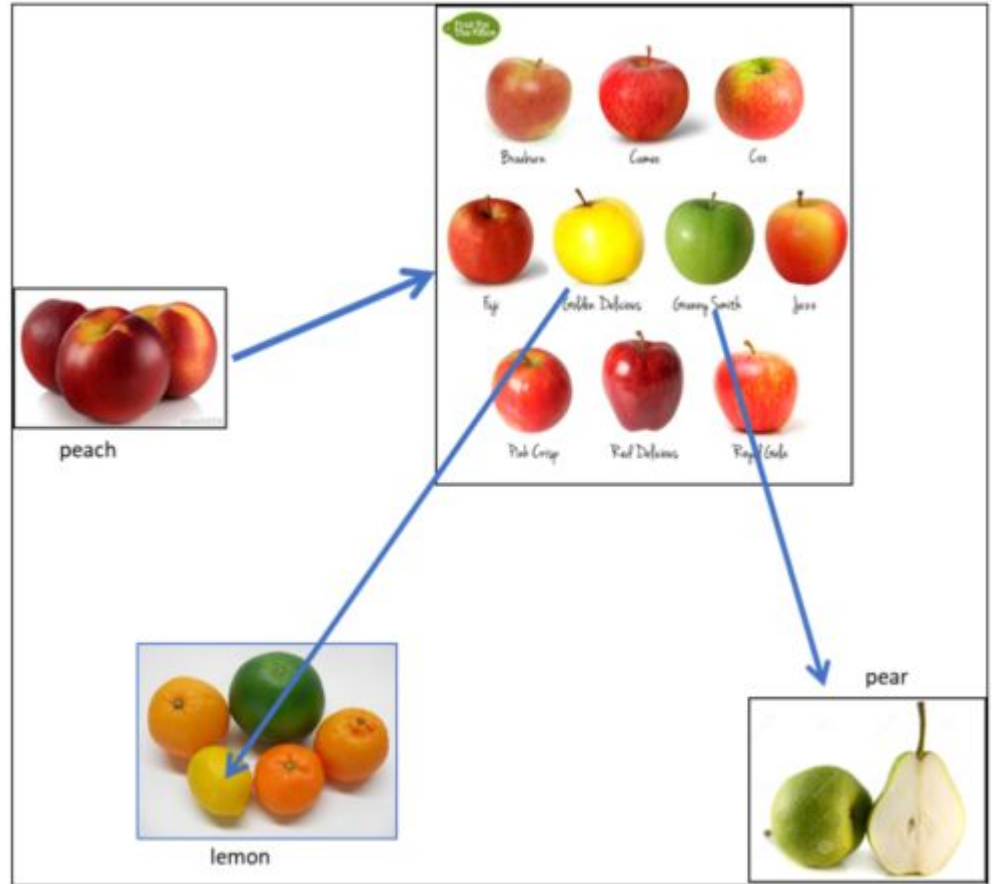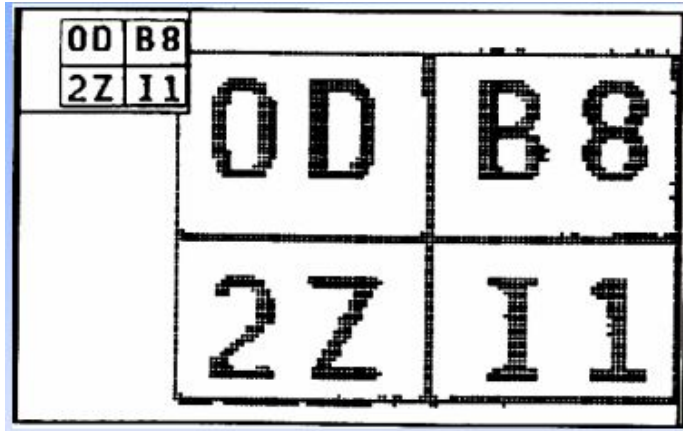
Pengenalan Pola: Definisi

● Penemuan keteraturan dalam data secara otomatis (Bishop, 2006)

● Aksi mengambil data mentah dan melakukan aksi berdasarkan kategori polanya (Duda et al., 2001)

● Persoalan menemukan properti penting dari suatu kategori (Duda et al., 2001)

Pengenalan Pola: Definisi (2)

- Bidang yang mempelajari bagaimana mesin mengobservasi lingkungannya, belajar untuk membedakan pola yang menarik, dan menentukan kategori pola tersebut (Jain, 1999).
- Generalisasi dari sejumlah contoh (CSE 802 MSU, 2008)

Pengenalan Pola: Mengapa ?

- AI: komputer berperilaku seperti manusia (acting humanly)
  - "The real power of human thinking is based on recognizing patterns. The better computers get at pattern recognition, the more human-like they will become". Ray Kurzweil, NY Times, Nov 24, 2003

-

What is AI? 8 Definitions, 4 Approaches

Thinking humanly

Acting humanly

Thinking rationally

Acting rationally

Thinking Humanly

"The exciting new effort to make computers think ... *machines with minds*, in the full and literal sense." (Haugeland, 1985)

"[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning ..." (Bellman, 1978)

Thinking Rationally

"The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985)

"The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)

Acting Humanly

"The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990)

"The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)

Acting Rationally

"Computational Intelligence is the study of the design of intelligent agents." (Poole et al., 1998)

"AI ... is concerned with intelligent behavior in artifacts." (Nilsson, 1998)

Figure 1.1 Some definitions of artificial intelligence, organized into four categories.

Acting humanly: Turing Test Approach
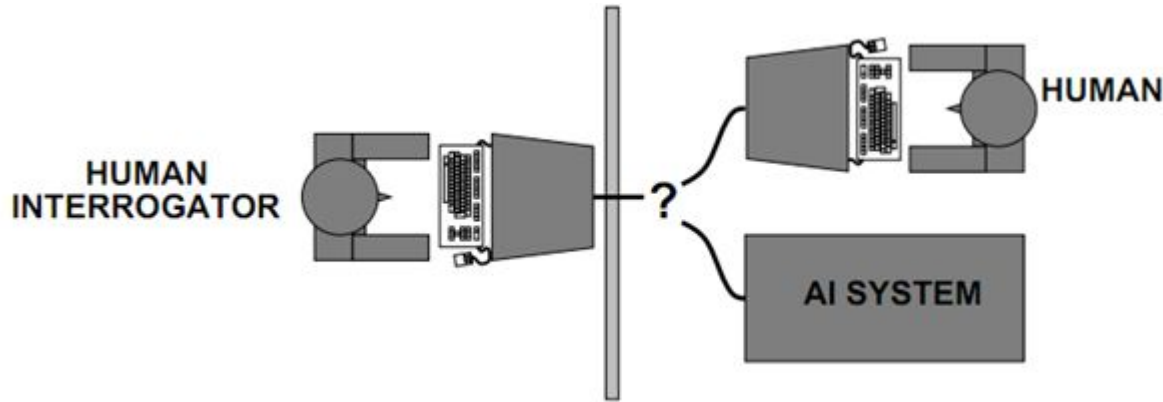
- A satisfactory operational definition of intelligence
- Pass: human interrogator cannot tell whether the responses come from a person or from a computer
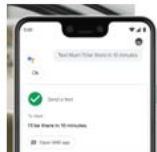
Human inside Computer

Putting a human inside a computer is not Artificial Intelligence

http://www.computervisionblog.com/2015/03/deep-learning-vs-machine-learning-vs.html

Aplikasi Pengenalan Pola: Speech Recognition

Speech → Text

Personal Assistant (word cloud: Aido, Ubi, now, uHound, top, Jibo, Mycroft, Cubic, Google, Intelligent, Siri, MD.com, Lucida, Maluuba, Automated, BlackBerry, SILVIA, Amazon, Echo, Vlingo, Nina, Cortana, Braina)

"Hey Google, text Mum I'll be there in 10 minutes"

Debat Capres #5

Ekonomi dan Kesejahteraan Sosial, Keuangan dan Investasi, serta Perdagangan dan Industri
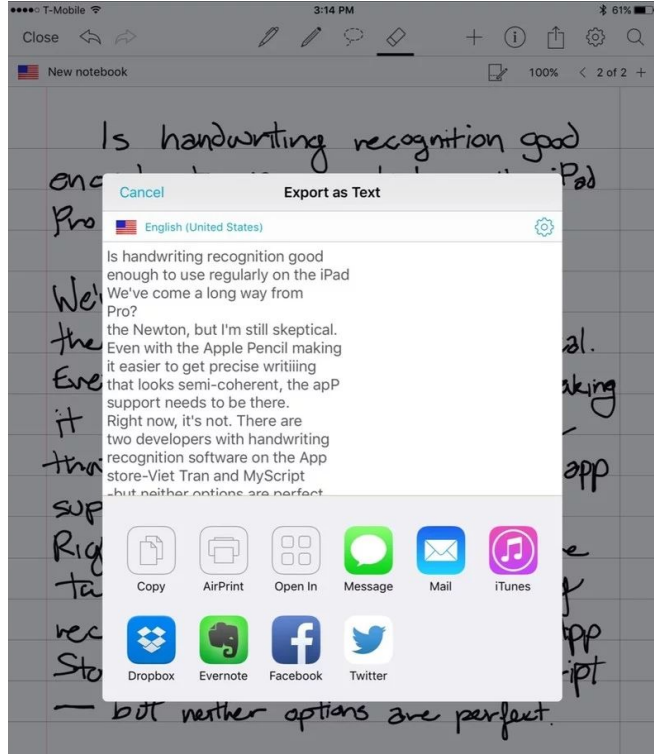
13 April 2019 | Status Pengolahan Data : Selesai

Transkrip Lengkap

Search

No	Pembicara	Waktu (WIB)	Transkrip
1	Balques Manisang	20:04	baik kami meminta hadirin baik kami meminta hadirin berdiri karena kita akan bersama sama menyanyikan lagu kebangsaan Indonesia raya
2	pembicara	20:04	(Lagu Kebangsaan Indonesia Raya)
3	Balques Manisang	20:06	pak Maruf Amin untuk bersalaman juga dengan Prabowo Subianto dan bapak Salahudin iya silakan baik tidak sabar rasanya Pak silakan Pak monggo ditempati tempatnya satu-satu
4	Balques	20:07	baik sebelum memulai acara karena sudah siap semua tapi ada yang juga sangat penting untuk memulai acara yang besar ini kita akan membacakan doa menundukan sejenak kepala kita dan yang akan memimpin doa adalah Profesor Kiai Haji Nasaruddin Umar MA Ph D Imam Besar Masjid Istiqlal

Aplikasi Pengenalan Pola: Handwriting Recognition

Aplikasi Pengenalan Pola: Text Recognition

Google's Word Lens
https://mashable.com/2017/01/26/word-lens-japanese/

https://apps.apple.com/us/app/scan-translate-text-grabber/id845139175

Aplikasi Pengenalan Pola: Face Recognition

https://www.nec.com/en/global/solutions/safety/face_recognition/NeoFaceWatch.html

https://www.grab.com/sg/driver-2/selfieverification/

http://www.vvip.co/news/facebook-accused-of-using-10-year-challenge-meme-to-improve-facial-recognition/

23

Aplikasi Pengenalan Pola: Pengenalan Emosi
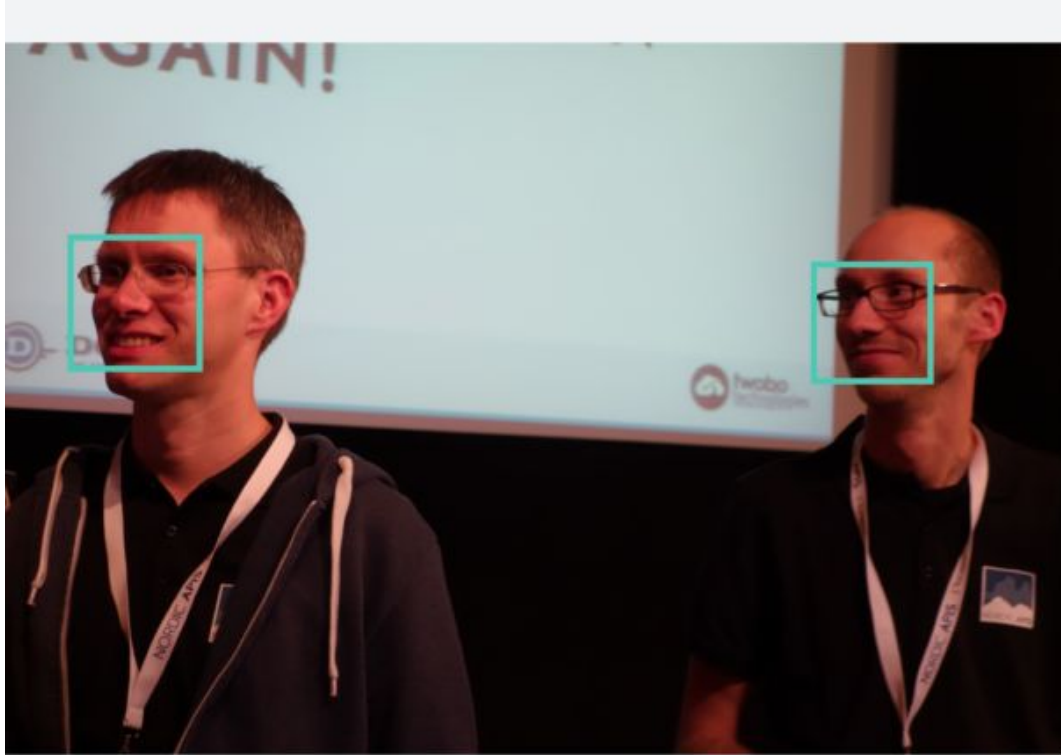
Detection Result:
2 faces detected

JSON:
[
  {
    "faceRectangle": {
      "left": 120,
      "top": 362,
      "width": 255,
      "height": 255
    },
    "scores": {
      "anger": 6.506412e-7,
      "contempt": 0.00000107357334,
      "disgust": 0.0000137053685,
      "fear": 2.51182275e-9,
      "happiness": 0.9994379,
      "neutral": 0.000546224066,
      "sadness": 1.46409562e-7,
      "surprise": 2.88747827e-7

https://nordicapis.com/20-emotion-recognition-apis-that-will-leave-you-impressed-and-concerned/

Aplikasi Pengenalan Pola: Analisis Teks

In fact, the **Chinese** `NORP` market has the **three** `CARDINAL` most influential names of the retail and tech space – **Alibaba** `GPE`, **Baidu** `ORG`, and **Tencent** `PERSON` (collectively touted as **BAT** `ORG`), and is betting big in the global **AI** `GPE` in retail industry space. The **three** `CARDINAL` giants which are claimed to have a cut-throat competition with the **U.S.** `GPE` (in terms of resources and capital) are positioning themselves to become the 'future **AI** `PERSON` platforms'. The trio is also expanding in other **Asian** `NORP` countries and investing heavily in the **U.S.** `GPE` based **AI** `GPE` startups to leverage the power of **AI** `GPE`. Backed by such powerful initiatives and presence of these conglomerates, the market in APAC AI is forecast to be the fastest-growing **one** `CARDINAL`, with an anticipated **CAGR** `PERSON` of **45%** `PERCENT` over

To further elaborate on the geographical trends, **North America** `LOC` has procured in **2017** `DATE` and has been leading the regional landscape of **AI** `GPE` in the retail credit in the regional trends with **over 65%** `PERCENT` of investments (including M&As artificial intelligence technology. Additionally, the region is a huge hub for startups in such as **Google** `ORG`, **IBM** `ORG`, and **Microsoft** `ORG`.

https://www.depends-on-the-definition.com/text-analysis-with-named-entities/

Polaritas	Aspek	Ekspresi

Keterangan warna: Positif Negatif Netral

Tempatnya bagus banget terlebih ada view kota bandungnya . Cuma sayang banget kemaren pesen makanan di restauran cave nya lama banget datengnya ampe setengah jam baru dateng , rasanya lumayan , penyajiannya lumayan , harga rata rata di atas 30 . Porsi makanannya besar , cukup untuk dua orang . Lokasi strategis dekat Lembang .
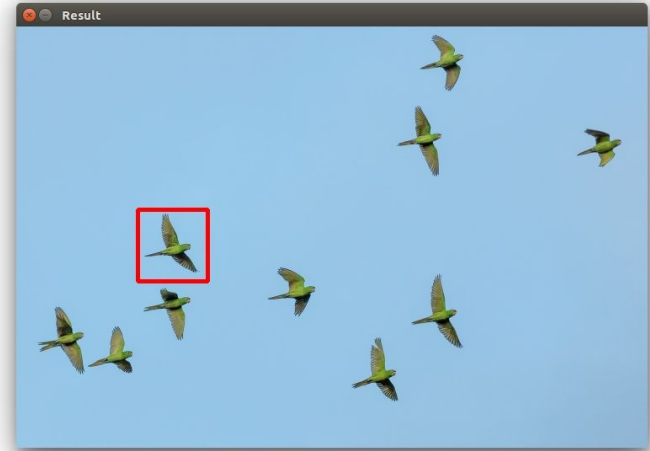
http://nlp.yuliadi.pro/

25

Pengenalan Pola: Pendekatan (Jain, 1999)

- Template matching
  - Matching: operator untuk menentukan kemiripan antar entitas dengan tipe yang sama
- Statistical decision theoretic
  - Quantitative features, and decision boundaries (classifier)
- Structural/syntactic
  - Morphological features, and parsing relational graphs

Pendekatan Template Matching

TEST INSTANCE

"J" TEMPLATE   "T" TEMPLATE
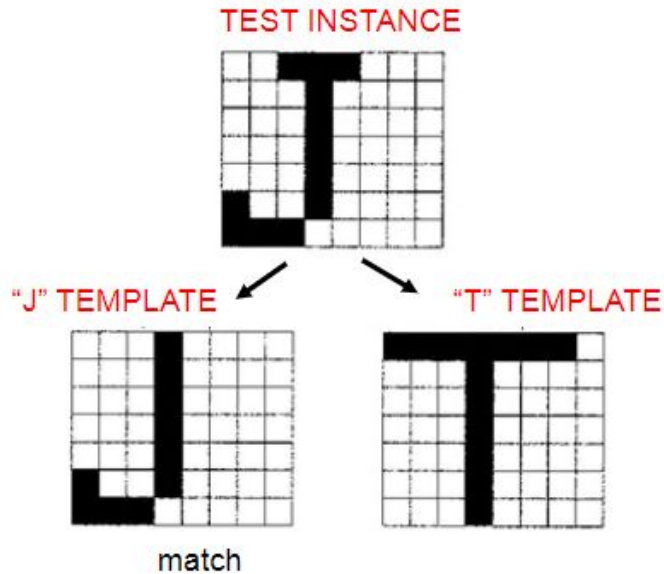
match

Template matching with OpenCV and Python.
Template (left), result image (right)

https://pythonspot.com/object-detection-with-templates/
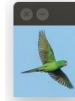
Pendekatan Template Matching untuk Teks

```
import re
email = re.compile('\w+@\w+\.[a-z]{3}')
text = "To email Guido, try guido@python.org or the older address guido@google.com."
email.findall(text)
```

```
['guido@python.org', 'guido@google.com']
```

https://www.nzini.com/lessons/NLP2+-+Template+Matching.html

```
Got input: How cold is it in the kitchen?
cleanedInput: how cold is it in the kitchen
We found a match!
        Regex: how.*cold.*kitchen.*
        Logical Form: ['kitchen', 'temperature']
The temperature in the kitchen is 81
```

```
EXPRESSIONS = [("what.*temp.*kitchen.*",   ["kitchen", "temperature"]),
               ("how.*hot.*kitchen.*",     ["kitchen", "temperature"]),
               ("how.*cold.*kitchen.*",    ["kitchen", "temperature"]),
               ("what.*temp.*bath.*",      ["bathroom", "temperature"]),
               ("how.*hot.*bath.*",        ["bathroom", "temperature"]),
               ("how.*cold.*bath.*",       ["bathroom", "temperature"]),
               ("what.*temp.*living.*",    ["livingroom", "temperature"]),
               ("how.*hot.*living.*",      ["livingroom", "temperature"]),
               ("how.*cold.*living.*",     ["livingroom", "temperature"]),
               ("what.*temp.*family.*",    ["livingroom", "temperature"]),
               ("how.*hot.*family.*",      ["livingroom", "temperature"]),
               ("how.*cold.*family.*",     ["livingroom", "temperature"]),
               ("what.*temp.*bed.*",       ["bedroom", "temperature"]),
               ("how.*hot.*bed.*",         ["bedroom", "temperature"]),
               ("how.*cold.*bed.*",        ["bedroom", "temperature"]),
               ("what.*temp.*dining.*",    ["diningroom", "temperature"]),
               ("how.*hot.*dining.*",      ["diningroom", "temperature"]),
               ("how.*cold.*dining.*",     ["diningroom", "temperature"])]
```

rule

28
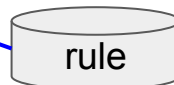
Pendekatan Statistical Decision Theoritic

- Pola direpresentasikan sebagai d-features, dan dapat dilihat sebagai titik pada ruang d-dimensi
- Tujuan: memilih pola sehingga vektor pola milik kategori yang berbeda berada pada wilayah yang disjoint dalam ruang d-dimensi
- Efektifitas representasi (set fitur) diukur dengan seberapa baik pola dari kategori yang berbeda dipisahkan
- Decision boundaries ditentukan oleh distribusi peluang setiap pola milik setiap kategori

Example: Differentiate between square and triangle

Number of segments: 4
Number of horizontal segments : 2
Number of vertical segments: 2
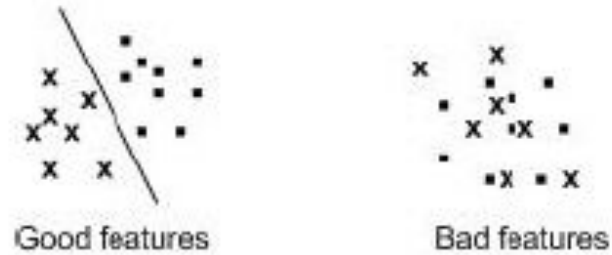Number of diagonal segments : 0
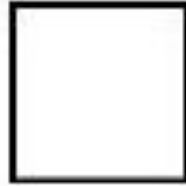
Number of segments: 3
Number of horizontal segments : 1
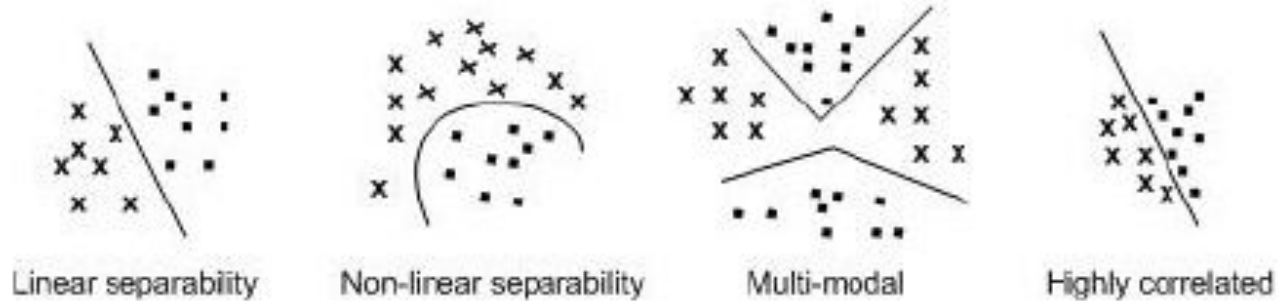Number of vertical segments: 0
Number of diagonal segments : 2

A statistical approach extracts quantitative features such as the number of horizontal, vertical, and diagonal segments which are then passed to a decision-theoretic classifier.

https://www.byclb.com/TR/Tutorials/neural_networks/ch1_1.htm

Good vs Bad Features

Good features   Bad features

(a)

Linear separability   Non-linear separability   Multi-modal   Highly correlated

(b)

Pendekatan Structural/Syntactic

- Pola ~ sentence in formal language theory
- Structural / syntactic features ~ primitives
- Feature vectors ~ morphological features
- Classification is performed by parsing the relational graphs with syntactic grammars or by rules.

# Example: Differentiate between square and triangle

horizontal
vertical   vertical
horizontal

diagonal ↔ diagonal
horizontal
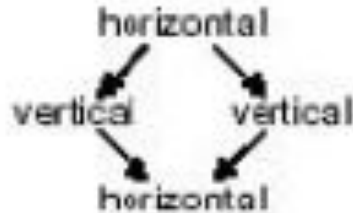
A structural approach extracts morphological features and their interrelationships, encoding them in relational graphs; classification is performed by parsing the relational graphs with syntactic grammars.
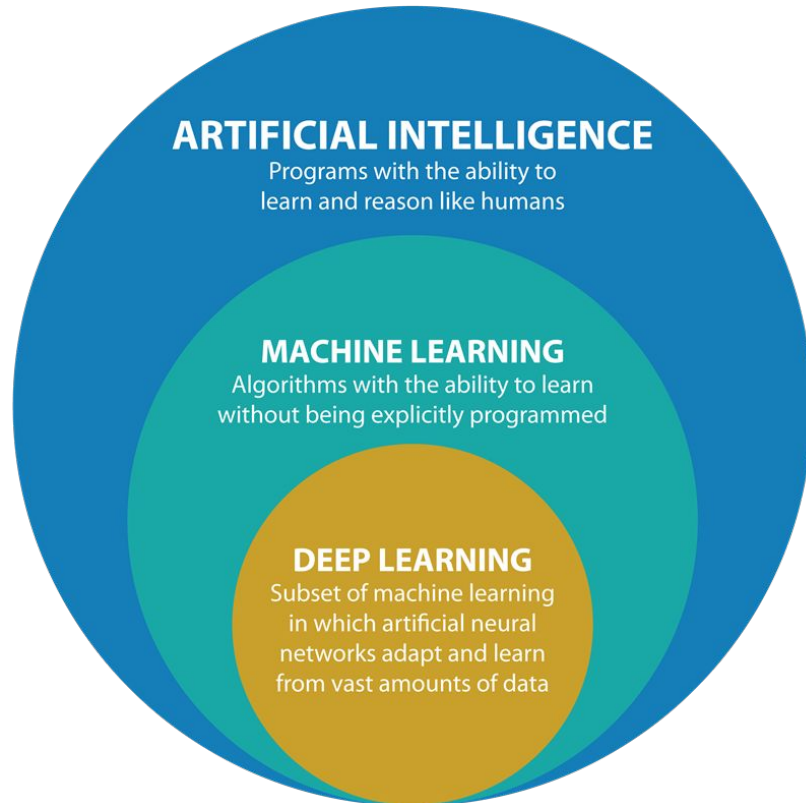
Learning Buzz Words: PR - ML - DL

Interest over time

Worldwide. 2004 - present. Web Search.

● machine learning ● pattern recognition ● deep learning
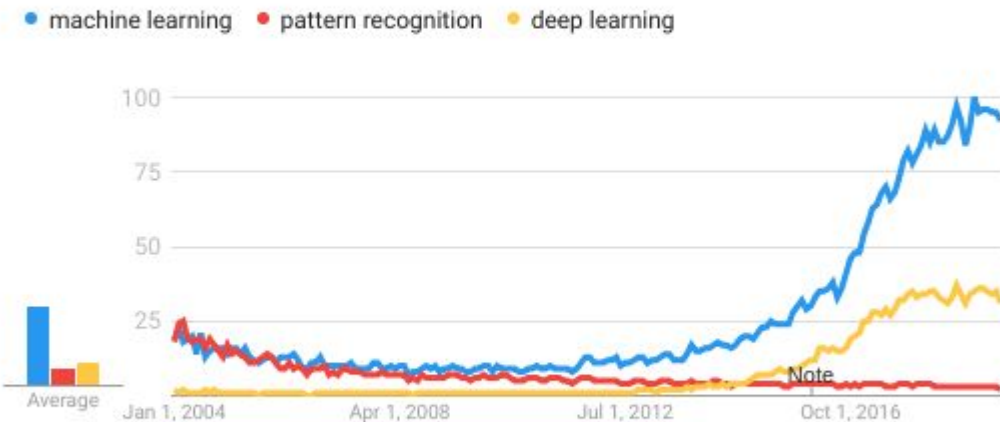
Google Trends

a) Machine Learning is rising like a true champion, b) Pattern Recognition started as synonymous with Machine Learning, c) Pattern Recognition is dying, d) Deep Learning is new and rising fast.

AI - ML - DL

ARTIFICIAL INTELLIGENCE
Programs with the ability to
learn and reason like humans

MACHINE LEARNING
Algorithms with the ability to learn
without being explicitly programmed

DEEP LEARNING
Subset of machine learning
in which artificial neural
networks adapt and learn
from vast amounts of data

Machine Learning: Definition

Arthur Samuel (1959): Machine Learning is the field of study that gives the computer the ability to learn without being explicitly programmed. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.368.2254&rep=rep1&type=pdf

Tom Mitchell (1998): a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

Machine Learning Workflow

http://nkonst.com/machine-learning-explained-simple-words/
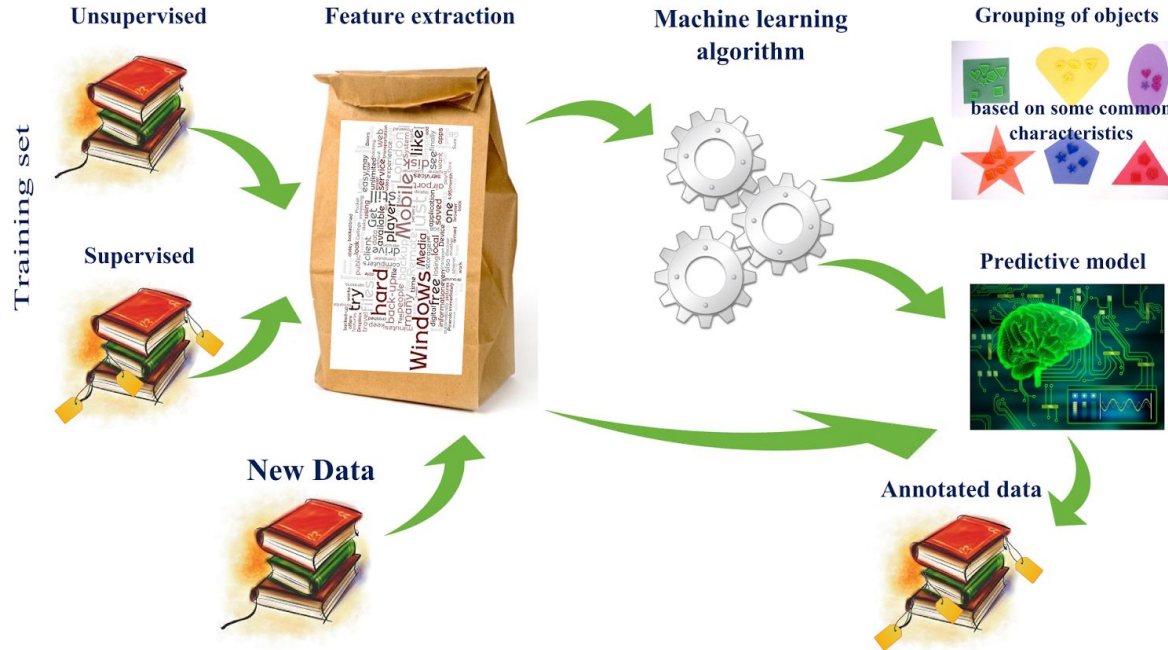
Machine Learning: Why
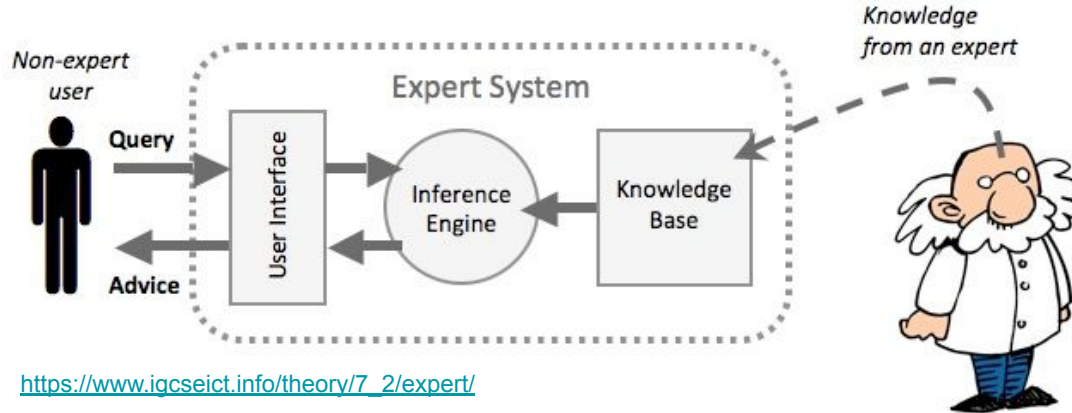
https://www.igcseict.info/theory/7_2/expert/

Dalam sistem berbasis pengetahuan, sebagai teknik akuisisi pengetahuan otomatis untuk menangani knowledge acquisition bottleneck

Develop software applications that we can't program by hand (poorly understood domains). Humans might not have the knowledge to develop effective algorithms

face recognition

https://www.nec.com/en/global/solutions/safety/face_recognition/NeoFaceWatch.html

self driving car

https://www.thestar.com.my/tech/tech-news/2019/07/08/waymo-taxis-on-the-way-california-green-lights-test-programme/

38

Pertanyaan ?

PR

- Berikanlah satu contoh persoalan yang diselesaikan dengan 3 pendekatan template matching, statistical, dan sintaksis.