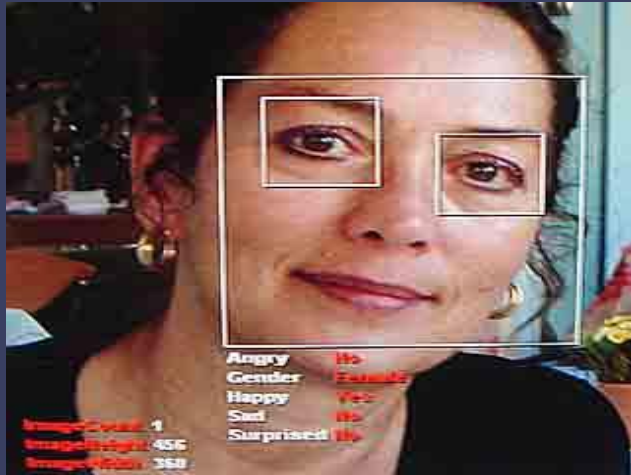


# Machine Learning



## Pengantar Machine Learning

Entin Martiana

Knowledge Engineering Research Group

Soft Computing Laboratory

Department of Information and Computer Engineering

Politeknik Elektronika Negeri Surabaya

Politeknik Elektronika Negeri Surabaya  
Departemen Teknik Informatika dan Komputer



# Konten

- Latar Belakang Machine Learning
- Definisi Machine Learning
- Penerapan Machine Learning
- Jenis Proses Learning
- Pokok Bahasan dalam Kuliah Machine Learning

# Tujuan Instruksi Umum

Mahasiswa mampu menyelesaikan masalah – masalah menggunakan metode mesin pembelajaran yang tepat berdasarkan supervised, unsupervised dan reinforcement learning, baik secara individu maupun berkelompok/kerjasama tim.

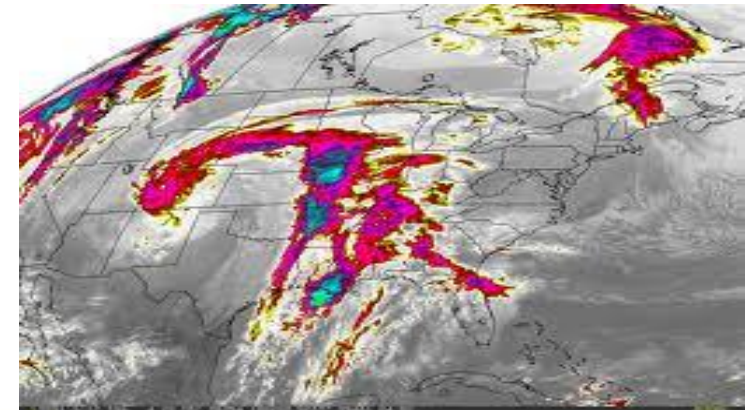
# Tujuan Instruksi Khusus

- Mengetahui definisi Mesin Pembelajaran
- Mengetahui jenis-jenis Pembelajaran
- Mengetahui data training dan data test dan kegunaannya dalam mesin pembelajaran

# Learning from Data

*The world is driven by data.*

- Germany's climate research centre generates 10 petabytes per year
- Google processes 24 petabytes per day
- The Large Hadron Collider produces 60 gigabytes per minute (~12 DVDs)
- There are over 50m credit card transactions a day in the US alone.



# The World of Data





1 **NEW** DEFINITION IS ADDED ON **URBAN**

1,600+ **READS ON Scribd.**

13,000+ HOURS **MUSIC** STREAMING ON **PANDORA**

12,000+ **NEW ADS** POSTED ON **craigslist**

370,000+ MINUTES **VOICE CALLS** ON **skype**

98,000+ **TWEETS**



20,000+ **NEW** POSTS ON **tumblr.**

THE **LARGEST** SOCIAL READING PUBLISHING COMPANY

320+ **NEW** **twitter** ACCOUNTS

100+ **NEW** **LinkedIn** ACCOUNTS



13,000+ **iPhone** APPLICATIONS DOWNLOADED

1 **associated content** **NEW** ARTICLE IS PUBLISHED

THE **WORLD'S LARGEST** COMMUNITY CREATED CONTENT!!

**QUESTIONS** ASKED ON THE INTERNET...

100+ **Answers.com** 40+ **YAHOO! ANSWERS**

6,600+ **NEW** PICTURES ARE UPLOADED ON **flickr**



600+ **NEW** **VIDEOS**



50+ **WORDPRESS** DOWNLOADS

695,000+ **facebook** STATUS UPDATES



125+ **PLUGIN** DOWNLOADS

25+ HOURS **TOTAL** DURATION

70+ **DOMAINS** REGISTERED

60+ **NEW** **BLOGS**

168 MILLION **EMAILS** ARE SENT

694,445 **SEARCH** QUERIES

1,700+ **Firefox** DOWNLOADS

79,364 **WALL** POSTS

510,040 **COMMENTS**



1,500+ **BLOG** POSTS



**Google**

Google Search



# Learning from Data

**Data** is recorded from some real-world phenomenon.

What might we want to do with that data?

## Prediction

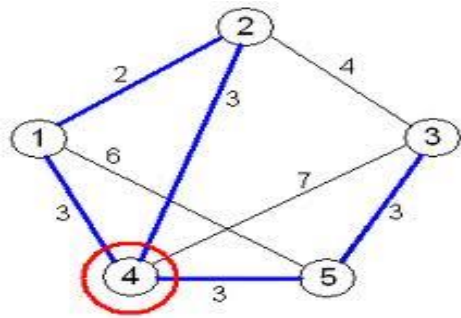
- what can we **predict** about this phenomenon?

## Description

- how can we **describe/understand** this phenomenon in a new way?



AHRQ Prevention Quality Indicators						
Dehydration Admission Rate (PQI 10)						
Counties highlighted in GREEN are significantly lower than the National Average.						
Counties highlighted in RED are significantly higher than the National Average.						
County Name	Cases	Population	Crude Rate	Risk Adj. Rate	Risk Adjusted LCL	Risk Adjusted UCL
Adair	79	13,774	5.74	4.62	5.19	5.76
Allen	28	14,299	1.96	1.41	2.00	2.59
Anderson	12	16,453	0.78	0.25	0.84	1.42
Ballard	8	6,538	1.22	0.24	1.03	1.83
Barren	102	31,112	3.28	2.56	2.93	3.31
Bath	15	8,943	1.68	0.84	1.55	2.26
Bell	122	23,056	5.29	4.52	4.96	5.41
Boone	68	78,500	0.87	0.85	1.14	1.42
Bourbon	20	16,245	1.31	0.70	1.26	1.81
Boyd	32	39,393	0.81	0.39	0.72	1.06
Boyle	32	22,387	1.43	0.88	1.34	1.79
Bracken	18	6,700	2.69	1.78	2.63	3.47
Breathitt	40	12,261	3.23	2.84	3.59	4.16
Breckinridge	23	15,006	1.53	0.94	1.50	2.07
Bullitt	23	52,112	0.44	0.23	0.58	0.93
Butler	9	10,366	0.87	0.18	0.96	1.54
Caldwell	13	10,281	1.26	0.39	1.00	1.61
Calloway	28	29,186	0.96	0.50	0.90	1.30
Campbell	54	66,477	0.81	0.53	0.89	1.07
Carlisle	5	4,215	1.19	0.00	0.93	1.89
Carroll	20	7,950	2.52	1.77	2.56	3.35
Carter	18	21,160	0.85	0.57	0.85	1.34
Casey	47	12,646	3.72	2.72	3.30	3.89





# Learning from Data

*How can we extract knowledge from data to help humans take decisions?*

*How can we automate decisions from data?*

*How can we adapt systems dynamically to enable better user experiences?*

---

Write code to explicitly  
do the above tasks



Write code to make the computer  
**learn** how to do the tasks



# Apa itu Machine Learning?

- Machine Learning adalah salah satu disiplin ilmu dari Computer Science yang mempelajari bagaimana membuat komputer/mesin itu mempunyai suatu kecerdasan
- Agar mempunyai suatu kecerdasan, komputer/mesin harus dapat belajar.
- Dengan kata lain, Machine Learning adalah suatu bidang keilmuan yang berisi tentang pembelajaran komputer/mesin untuk menjadi cerdas

# Learning process

- Supervised learning
- Unsupervised learning
- Reinforcement learning



# Supervised Learning

Diibaratkan sebagai proses belajar dari seorang murid yang berada dalam sebuah kelas

- Si murid diperbolehkan bertanya kepada guru yang telah mengetahui aturan jawabannya, dan kemudian si dosen menjawab pertanyaan tersebut.
- Dari hasil tanya-jawab berkali-kali, si murid akan bisa memahami rule dari permasalahan, sehingga jika ada permasalahan lain, si murid akan membandingkan dengan rule yang ia simpulkan sebelumnya, sehingga ia bisa memberikan jawaban.
- Memerlukan training. Semakin lama training, semakin pintar pula si murid memecahkan masalah.



# Unsupervised Learning

- Ketika si murid menjumpai masalah, ia harus dapat menjawab masalah tersebut dengan sendirinya.
- Semakin banyak ia berusaha menjawab sendiri, ia akan semakin pandai dalam menemukan rule yang dapat digunakan untuk memecahkan permasalahan di kemudian hari.
- Sangat bermanfaat jika memang permasalahan yang dihadapi relatif tidak bisa atau sulit sekali dijawab oleh sang guru.

# Reinforcement Learning

Contoh adalah bagaimana sikap yang diambil oleh seorang siswa di dalam kelas.

- Asumsikan bahwa sang guru sudah menjelaskan seperangkap aturan yang harus ditaati oleh siswa di dalam kelas.
- Suatu ketika, seorang siswa berteriak di dalam kelas. Maka sang guru langsung memberikan hukuman kepada siswa tersebut.
- Dari hukuman itu, siswa tadi akan merubah sikapnya untuk tidak berteriak lagi.
- Juga demikian, kepada siswa yang tekun mengikuti pelajaran di dalam kelas, maka sang guru memberikan kepada mereka semacam hadiah atau penghargaan.
- Jika sistem ini berjalan dalam jangka waktu tertentu, maka keadaan siswa tadi pasti akan konvergen untuk mengambil sikap yang baik di dalam kelas.

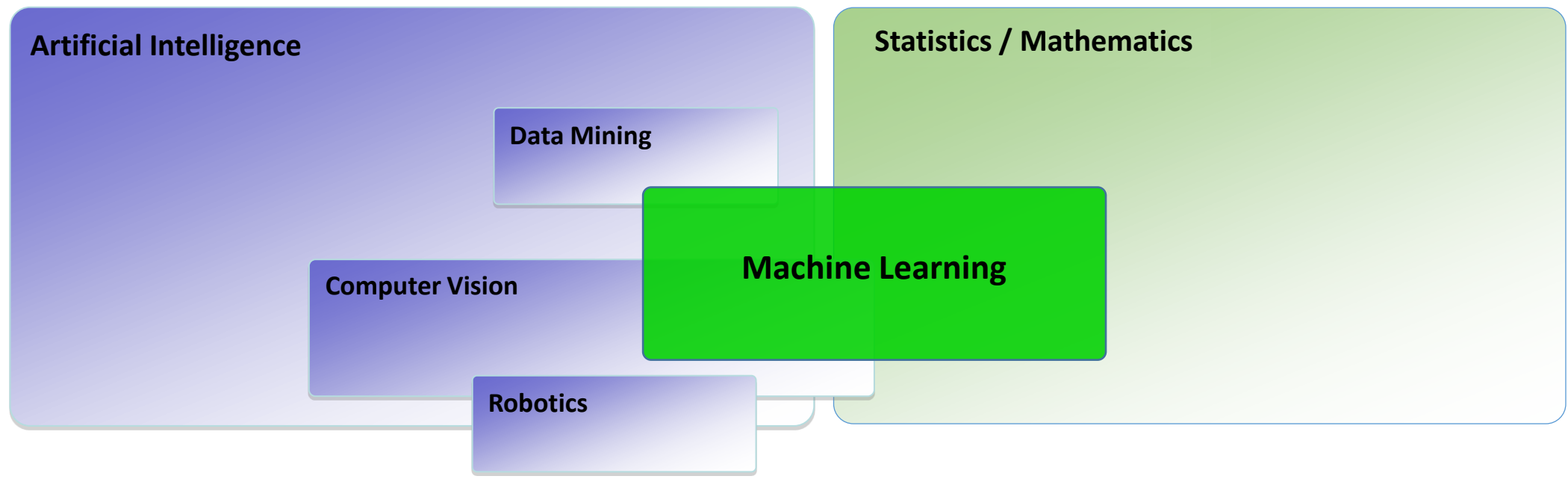


# Reinforcement Learning

- Seorang siswa yang bersikap baik di dalam kelas, ia akan mendapatkan reward. Dengan reward itu, ia akan bersikap lebih baik lagi. Jika ia bersikap lebih baik lagi, ia akan mendapatkan reward lagi. Demikian seterusnya yang terjadi sehingga ia pasti akan semakin konvergen dalam bersikap baik di dalam kelas.
- Sebaliknya, jika ia bersikap buruk, maka ia akan menerima punishment. Dengan punishment itu, ia akan merubah sikapnya. Jika punishment itu tidak cukup untuk membuatnya berubah, maka ia akan mendapatkan punishment lagi, sehingga dalam batasan tertentu, ia pasti akan berubah sikap yang hasilnya adalah ia akan mendapatkan reward. Demikian seterusnya, sehingga pada suatu saat nanti, ia akan konvergen bersikap baik di dalam kelas.

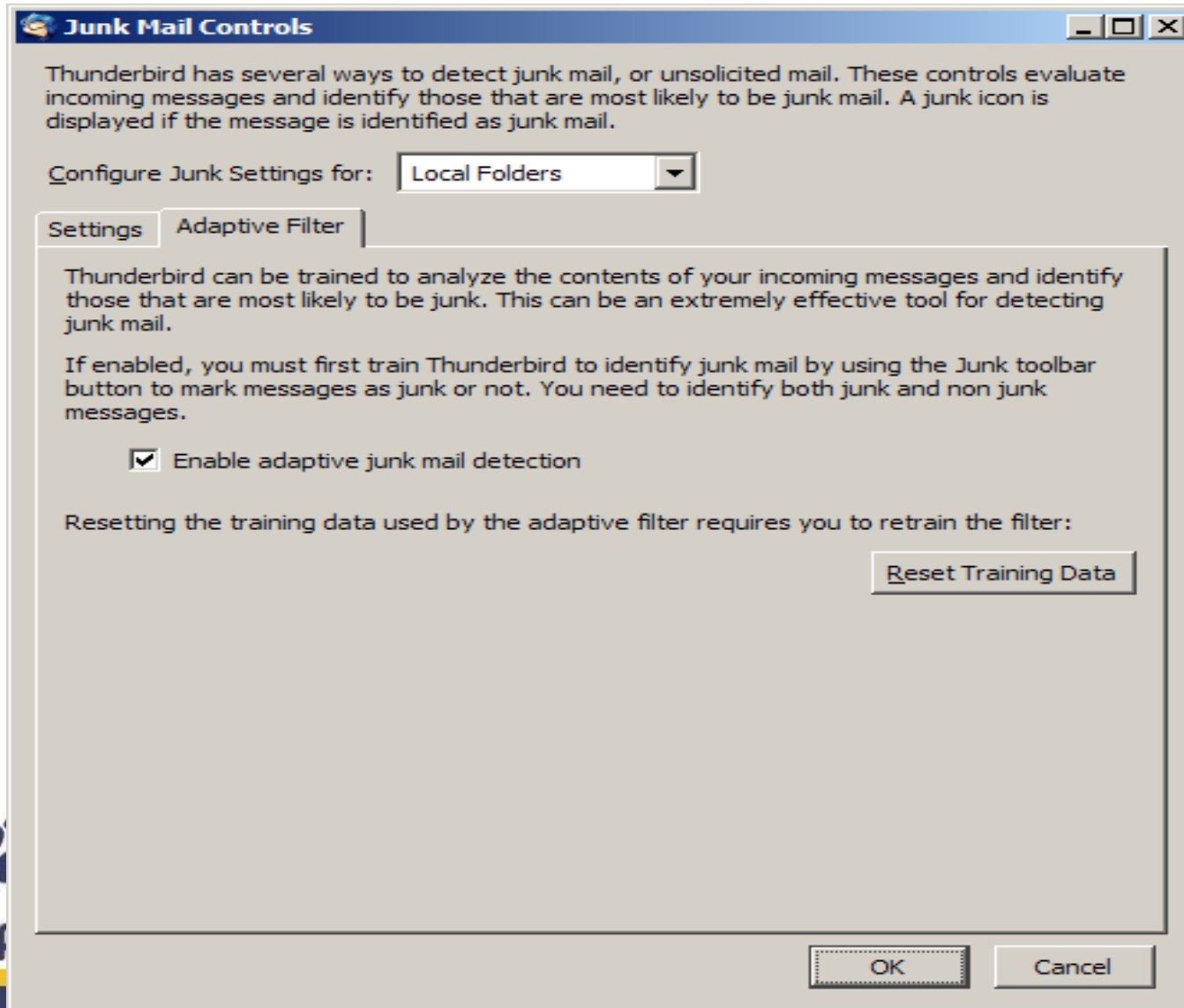
# Machine Learning

Where does it fit? What is it **not**?





- Using machine learning to detect spam emails.



To: you@gmail.com

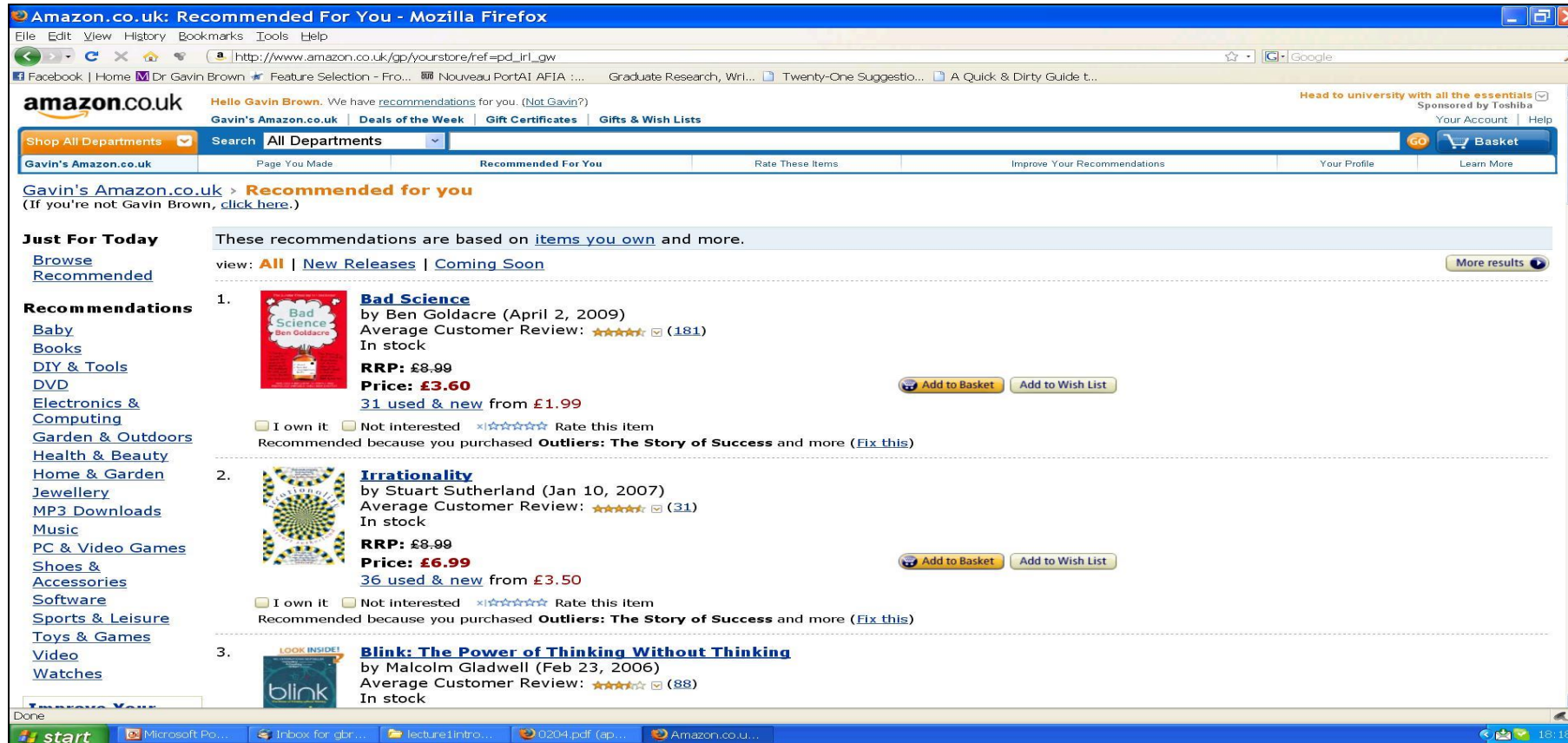
GET YOUR DIPLOMA TODAY!

If you are looking for a fast and cheap way to get a diploma, this is the best way out for you. Choose the desired field and degree and call us right now: For US: 1.845.709.8044 Outside US: +1.845.709.8044 "Just leave your NAME & PHONE NO. (with CountryCode)" in the voicemail. Our staff will get back to you in next few days!

## ALGORITHM

Naïve Bayes  
Rule mining

- Using machine learning to recommend books.

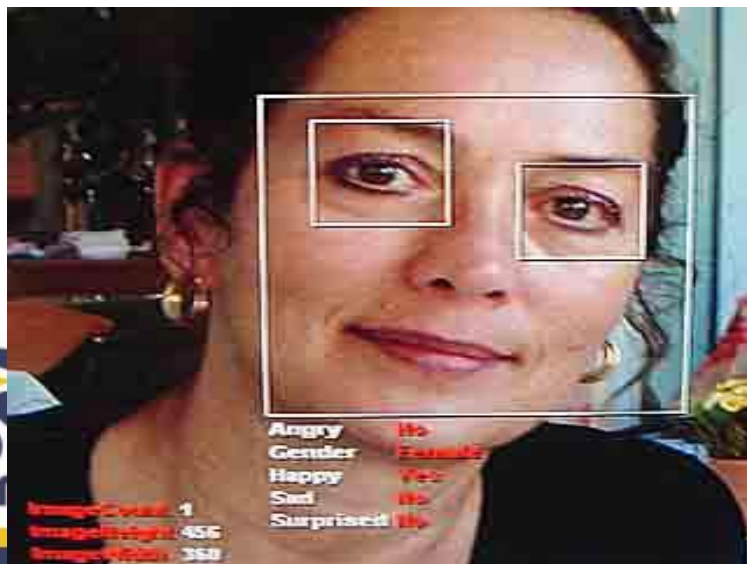


## ALGORITHMS

- Collaborative Filtering
- Nearest Neighbour
- Clustering



- Using machine learning to identify faces and expressions.

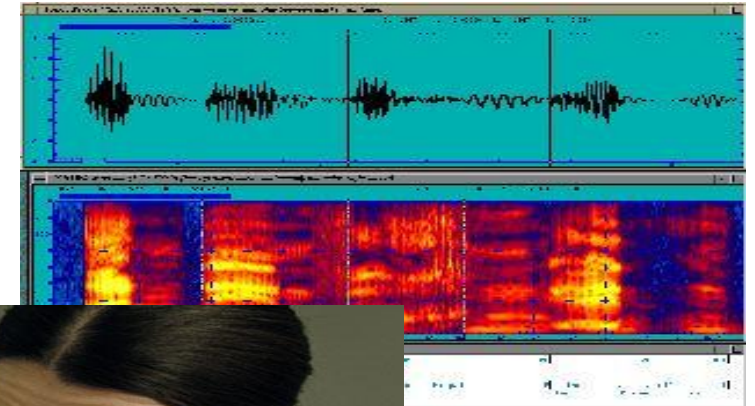
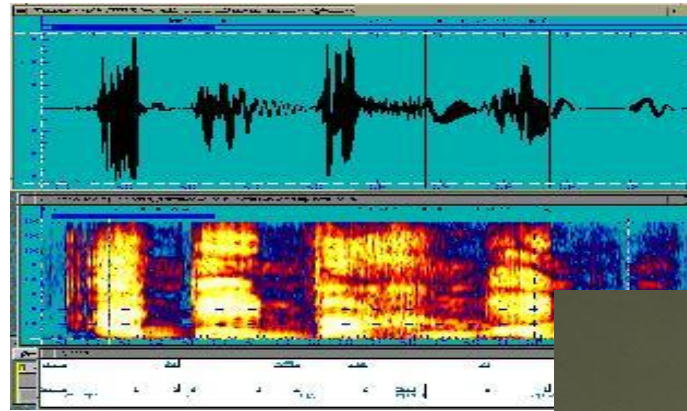
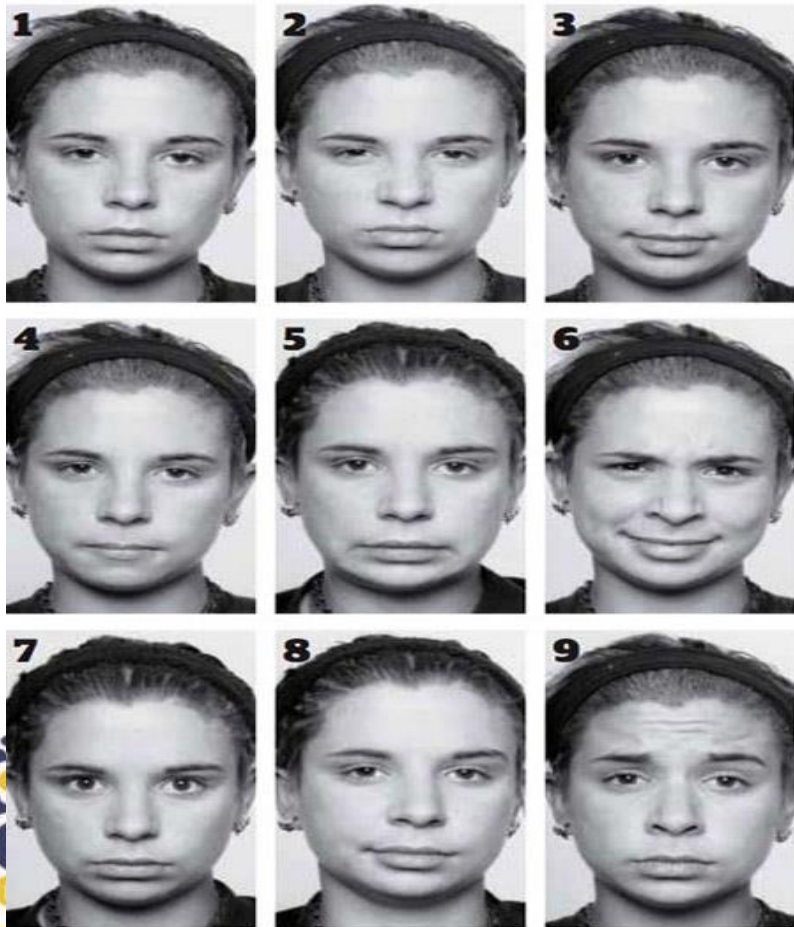


## ALGORITHMS

Decision Trees

Adaboost

- Using machine learning to identify vocal patterns



## ALGORITHMS

- Feature Extraction
- Probabilistic Classifiers
- Support Vector Machines
- + many more...

- ML for working with social network data: detecting fraud, predicting click-thru patterns, targeted advertising, etc etc etc .



## ALGORITHMS

Support Vector Machines  
Collaborative filtering  
Rule mining algorithms  
Many many more....

**Driving a car**

**Recognising spam emails**

**Recommending books**

**Reading handwriting**

**Recognising speech, faces, etc**

How would you write these programs?

Would you want to?!?!?!?



# Pembahasan di dalam ML

- Concept learning
- Bayesian learning
- Instance based learning (clustering)
- Neural Networks
- Genetic Algorithm
- Reinforcement Learning
- Dan lain-lain



# Concept learning

- Learning from examples
- General to specific ordering of hypotheses
- Uses only the positive data → Find-S
- Uses both positive and negative data → Candidate-Elimination
- Data harus konsisten
- Jawaban berada dalam 2 kemungkinan, ada atau tidak ada



# Contoh kasus

Data	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

# Bayesian Learning

- Learning from examples
- Termasuk ke dalam supervised learning
- Didasari pada Bayes Theorem
- Uses both positive and negative data
- Tidak mengharuskan data harus konsisten
- Jawaban ditunjukkan oleh nilai probabilitas
- Biasanya dipakai untuk fungsi-fungsi klasifikasi

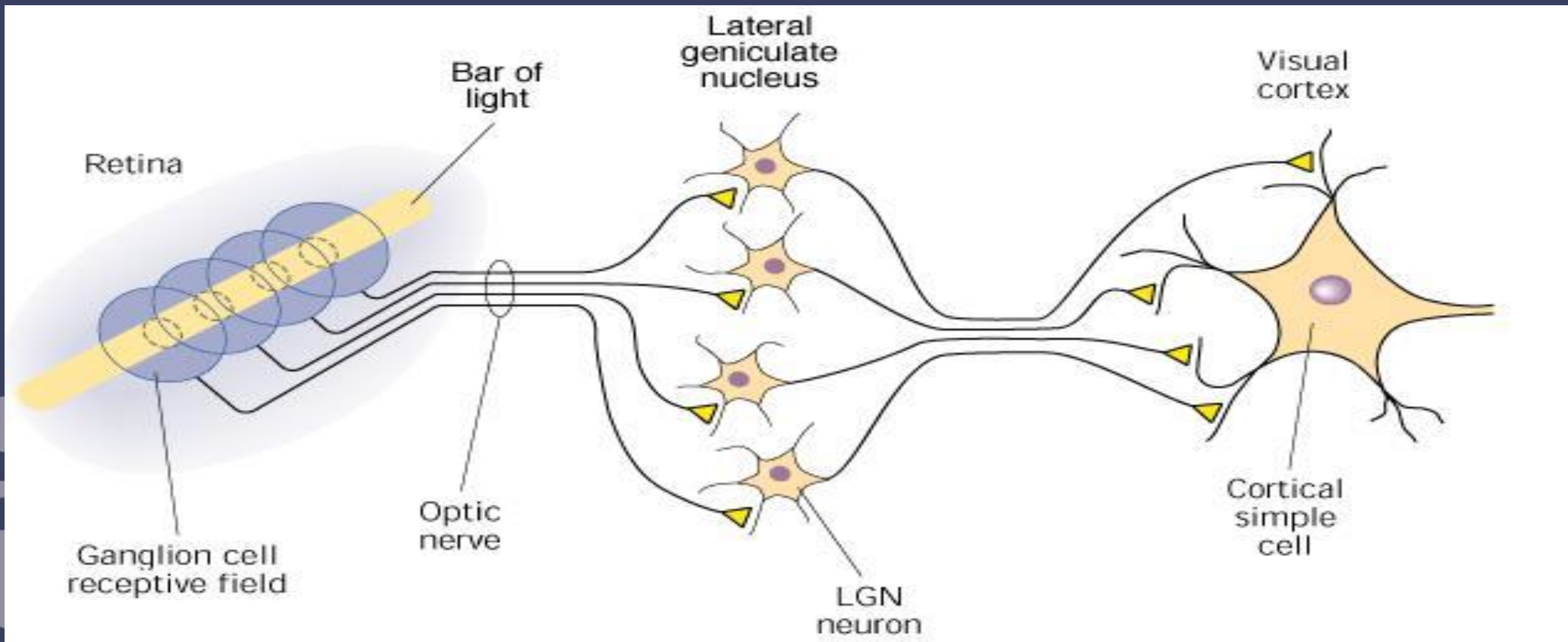


# Instance based learning (Clustering)

- Tidak melibatkan jawaban dalam data
- Termasuk ke dalam unsupervised learning
- Hanya membuat suatu klasifikasi tanpa label/jawaban
- Labelisasi (pemberian jawaban) menjadi tanggung jawab user

# Neural networks

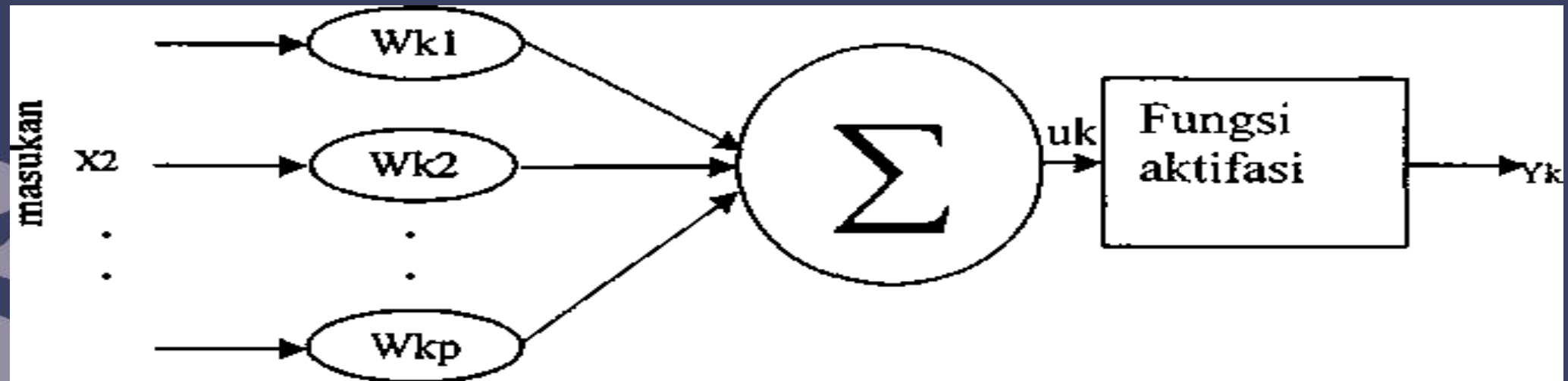
- Mensimulasikan kerja otak manusia
- *Neuron* adalah satuan unit pemroses terkecil pada otak
- Bentuk standard ini mungkin dikemudian hari akan berubah
- Jaringan otak manusia tersusun tidak kurang dari  $10^{13}$  buah neuron yang masing-masing terhubung oleh sekitar  $10^{15}$  buah *dendrite*
- Fungsi dendrite adalah sebagai penyampai sinyal dari neuron tersebut ke neuron yang terhubung dengannya
- Sebagai keluaran, setiap neuron memiliki *axon*, sedangkan bagian penerima sinyal disebut *synapse*
- Penjelasan lebih rinci tentang hal ini dapat diperoleh pada disiplin ilmu *biology molecular*
- Secara umum jaringan saraf terbentuk dari jutaan (bahkan lebih) struktur dasar neuron yang terinterkoneksi dan terintegrasi antara satu dengan yang lain sehingga dapat melaksanakan aktifitas secara teratur dan terus menerus sesuai dengan kebutuhan



Penimbang

penjumlahan

Keluaran

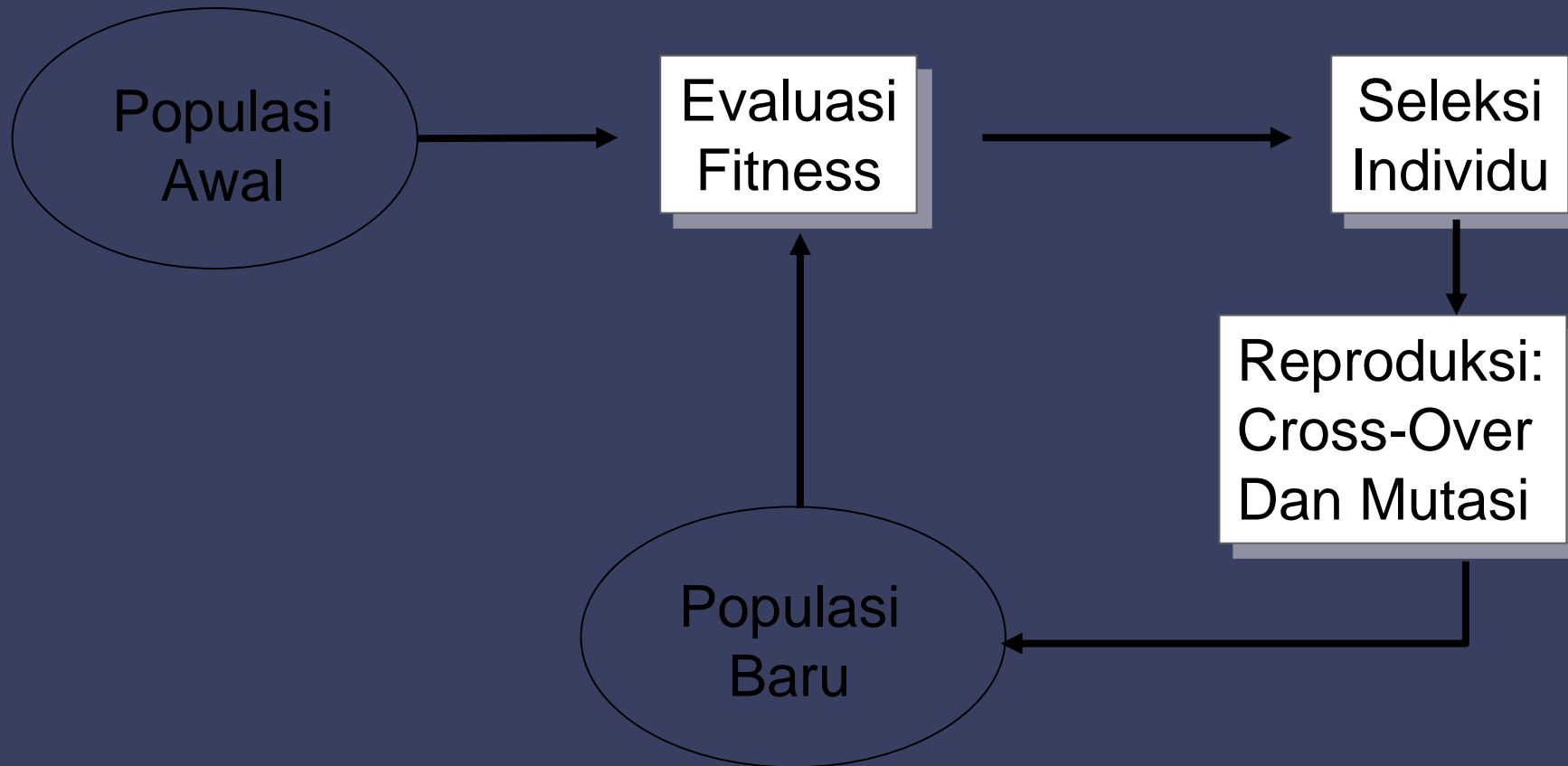


# Genetic Algorithm

- ◆ Algoritma Genetika adalah algoritma yang memanfaatkan proses seleksi alamiah yang dikenal dengan proses evolusi.
- ◆ Dalam proses evolusi, individu secara terus-menerus mengalami perubahan gen untuk menyesuaikan dengan lingkungan hidupnya. **“Hanya individu-individu yang kuat yang mampu bertahan”**.
- ◆ Proses seleksi alamiah ini melibatkan perubahan gen yang terjadi pada individu melalui proses perkembang-biakan. Dalam algoritma genetika ini, proses perkembang-biakan ini menjadi proses dasar yang menjadi perhatian utama, dengan dasar berpikir: **“Bagaimana mendapatkan keturunan yang lebih baik”**.



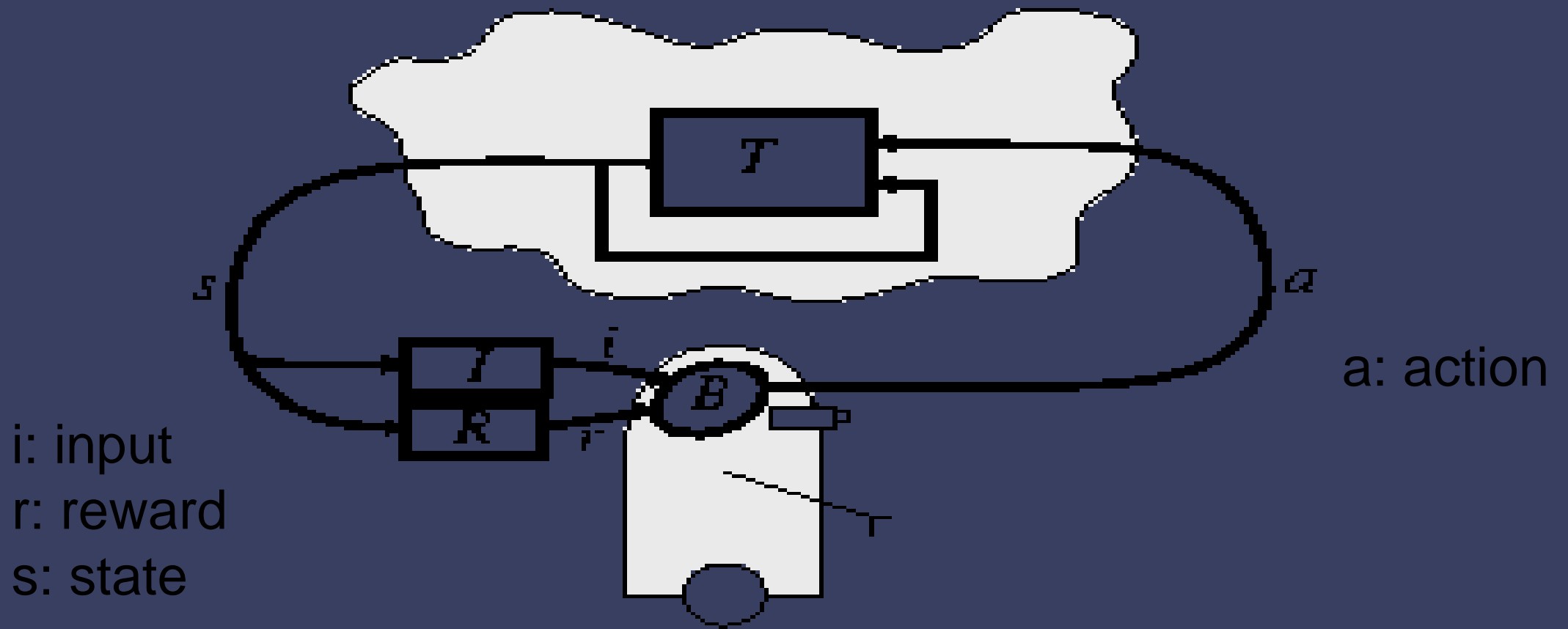
# Siklus Genetic Algorithm





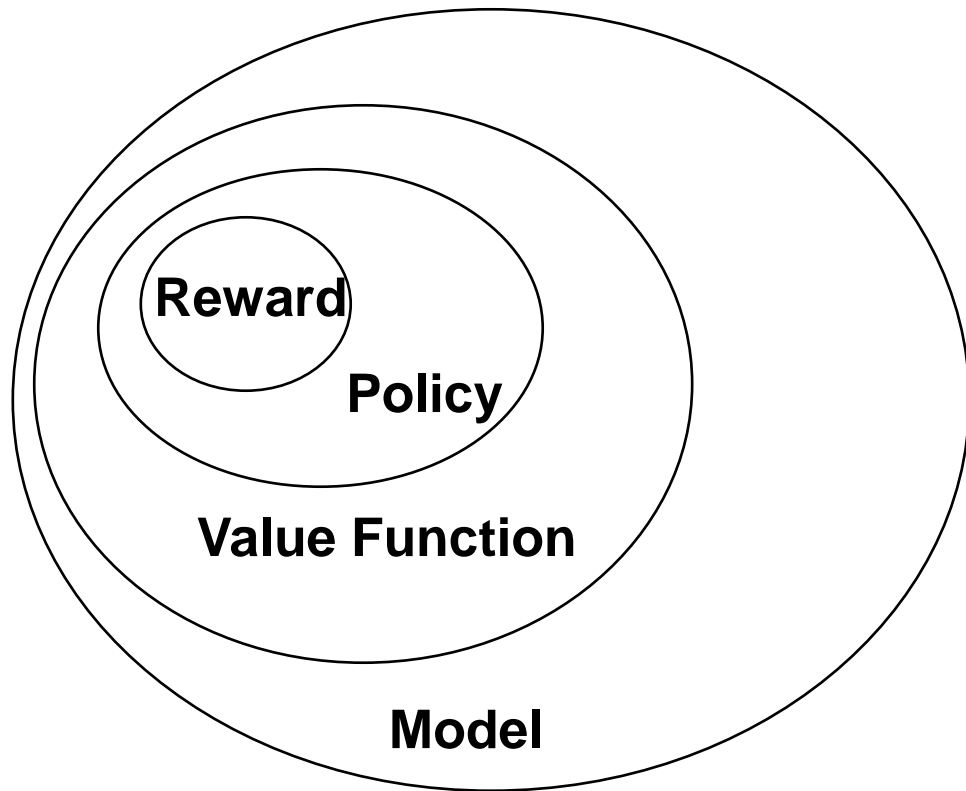
# Reinforcement Learning

- Learning from experiences
- Memakai konsep reward dan punishment dalam proses learning



The standard reinforcement-learning model

# Components of Reinforcement Learning



Reward: How good is this action?

Policy: what do I do now?

Value function: how good is this state?

Model: what happens if I do this action?

# Latihan Soal

1. Carilah artikel tentang Pengenalan Machine Learning!
2. Carilah aplikasi Machine Learning selain yang sudah dijelaskan pada modul ini!
3. Carilah tema-tema dari PA PENS yang merupakan implementasi dari Machine Learning!

# Referensi

- Modul Ajar Machine Learning, Entin Martiana, Ali Ridho Barakbah, Nur Rosyid Mubtadaí, Politeknik Elektronika Negeri Surabaya, 2013.
- Introduction to Machine Learning  
<http://www.cs.manchester.ac.uk/ugt/COMP24111>
- Machine Learning, Tom Mitchell, McGraw-Hill. 2008.



**bridge to the future**

<http://www.eepis-its.edu>