



Introduction to Data Mining

Profile



 [linkedin.com/in/romansyasetyo/](https://www.linkedin.com/in/romansyasetyo/)



**You can have data without
information, but you cannot have
information without data**

- Daniel Keys Moran

Table of Content

What will We Learn Today?

1. What is Data Mining?
2. Data Mining vs Machine Learning
3. Gain Insight through Data





What is Data Mining?





WEB CONNECTION
AGGREGATION INNOVATION BUSINESS
SOFTWARE INFORMATION CODE SOLUTION COLLECT BUG ERROR STORAGE
CONCEPT TRANSFORMATION EXTRACTION NETWORK ANALYTICS
ELABORATION BINARY INTERNET ANALYSIS
PATTERN PROCESS
DATA MINING
RECOGNITION
SYSTEM CALCULATION
MANAGEMENT DECISION SECURITY COMPUTER TECHNOLOGY
CONNECTIVITY
MACHINE LEARNING
CUSTOMER TROUBLESHOOT



Total Data Produced

2005 – 130 ExaBytes

2010 – 1,200 ExaBytes

2015 – 7,900 ExaBytes

2020 – 40,900 ExaBytes

Now
what?

IDC's Digital Universe Study 2012

*1 Exabyte = 1.000.000.000 GB



Data Mining

Data Mining is the process of making use of the huge volume of data to extract the interesting patterns or information (non-trivial, implicit, previously unknown, and potentially useful)



Data Mining

Previously known as “**Knowledge Discovery in Databases**”
(KDD)

Other names :

knowledge extraction, data/pattern analysis, data archeology, data dredging,
information harvesting



Motivation in Data Mining

- ✓ Perkembangan dalam teknologi basis data dan tool terotomasi untuk pengumpulan data telah mengakibatkan menumpuknya data dalam basis data, *data warehouses* dan tempat penyimpanan data lainnya.

Kaya akan data, tapi miskin akan pengetahuan?

- ✓ Informasi ‘tersembunyi’ dalam data
- ✓ Analisis secara manual membutuhkan waktu yang cukup lama untuk mencari informasi yang menarik
- ✓ Kebanyakan data tidak pernah dianalisis setelah dikumpulkan



Data Mining

Not Data Mining

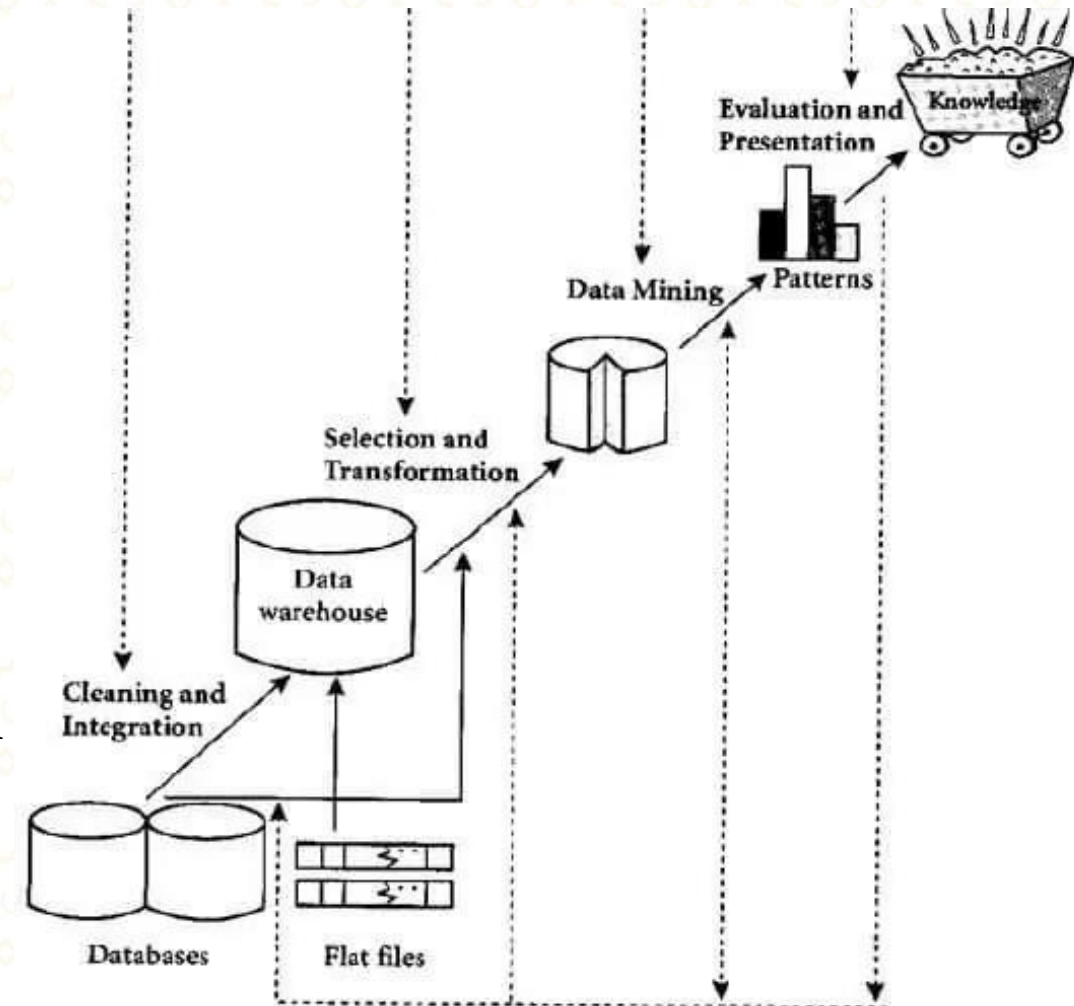
- Mencari nomor telpon dalam direktori telpon
- Melakukan query pada *Web search engine* untuk mencari informasi tentang “Amazon”

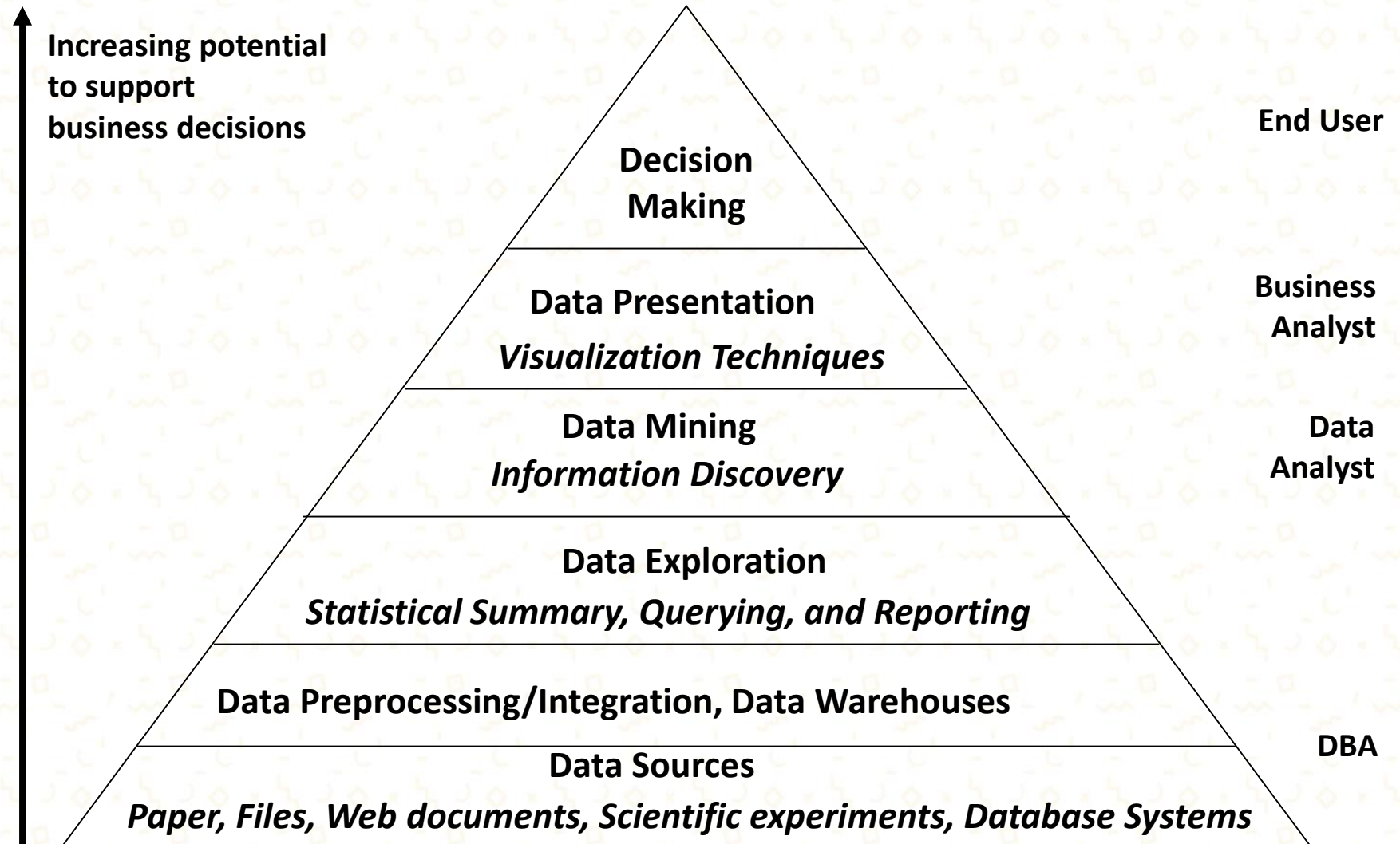
Data mining task

- Mencari nama tertentu yang lazim di wilayah/daerah tertentu. (Contoh: O’Brien, O’Rourke, O’Reilly... nama yang lazim di wilayah Boston)
- Mengelompokkan dokumen-dokumen yang mirip yang dikembalikan oleh *search engine* berdasarkan konteksnya (misalkan Amazon rainforest, Amazon.com)

Data Mining

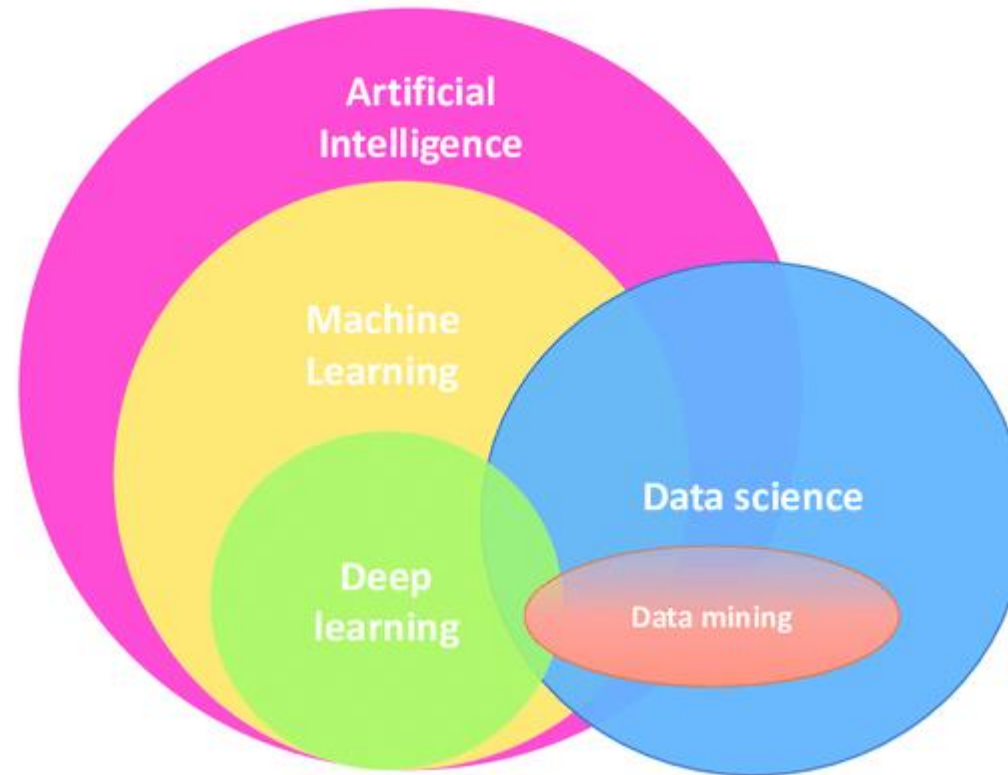
1. **Data Cleaning:** menghilangkan **noise** dan data yang tidak konsisten. Serta mengatasi **missing value**.
2. **Data Integration:** data digabungkan dari berbagai sumber.
3. **Data Selection:** data yang relevan dengan proses analisis diambil dari basis data.
4. **Data Transformation:** data ditransformasikan dengan cara dilakukan peringkasan atau operasi agregasi.
5. **Data mining:** beberapa macam metode diaplikasikan untuk mengekstrak pola-pola data.
6. **Pattern Evaluation:** melakukan evaluasi serta interpretasi atas pola-pola menarik yang ditemukan.
7. **Knowledge Presentation:** mempresentasikan pengetahuan yang telah digali kepada user.







Data Mining & Data Science





Data Mining & Data Science

Data Mining



A Technique.

Data Science



An Area



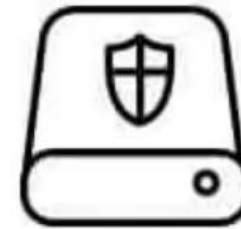
Data Mining & Data Science

Data Mining



Business Process.

Data Science

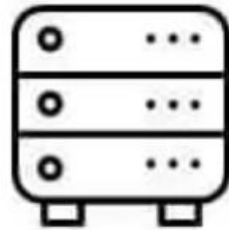


Scientific Study.



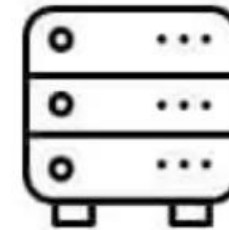
Data Mining & Data Science

Data Mining



Make
data more usable.

Data Science

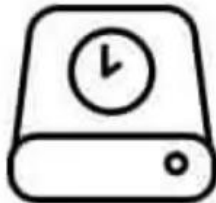


Building
Data-centric products for an
organization.



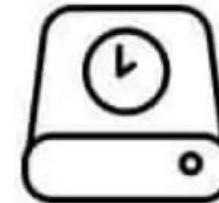
Data Mining & Data Science

Data Mining



Data mining can be a subset of Data Science as Mining activities are part of Data Science pipeline.

Data Science



Multidisciplinary
– Data Science consists of Data Visualizations, Computational Social Sciences, Statistics, Data Mining, Natural Language Processing, et cetera



Data Mining & Data Science

Data Mining



Mostly
structured.

Data Science



All
forms of data – structured, semi-
structured and unstructured.



Data Mining vs Machine Learning



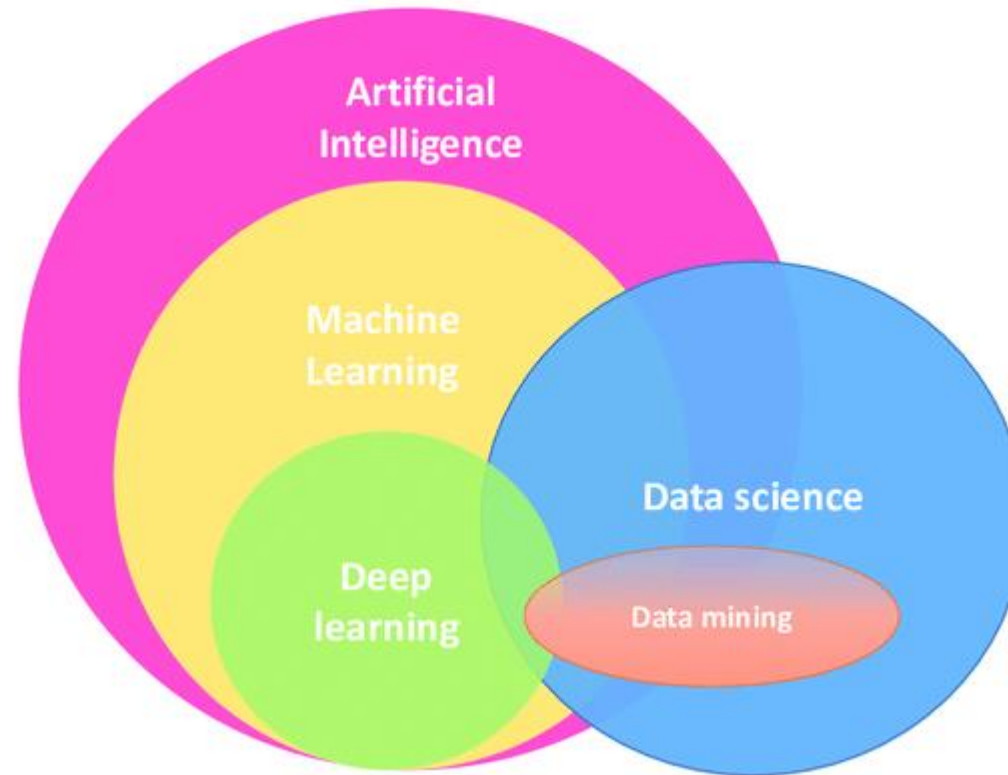
Data Mining & Machine Learning



VS



Data Mining & Machine Learning



Data Mining & Machine Learning

Data Mining



Extracting knowledge
from a large amount of
data

Data Science



Introduce new algorithm from data

Data Mining & Machine Learning

Data Mining



Get the rules or pattern
from the existing data

Data Science



Teach the machine or computer to
understand the rules or pattern

Data Mining & Machine Learning

Data Mining



Involves human
interference

Data Science



Automated



Gain Insight through Data



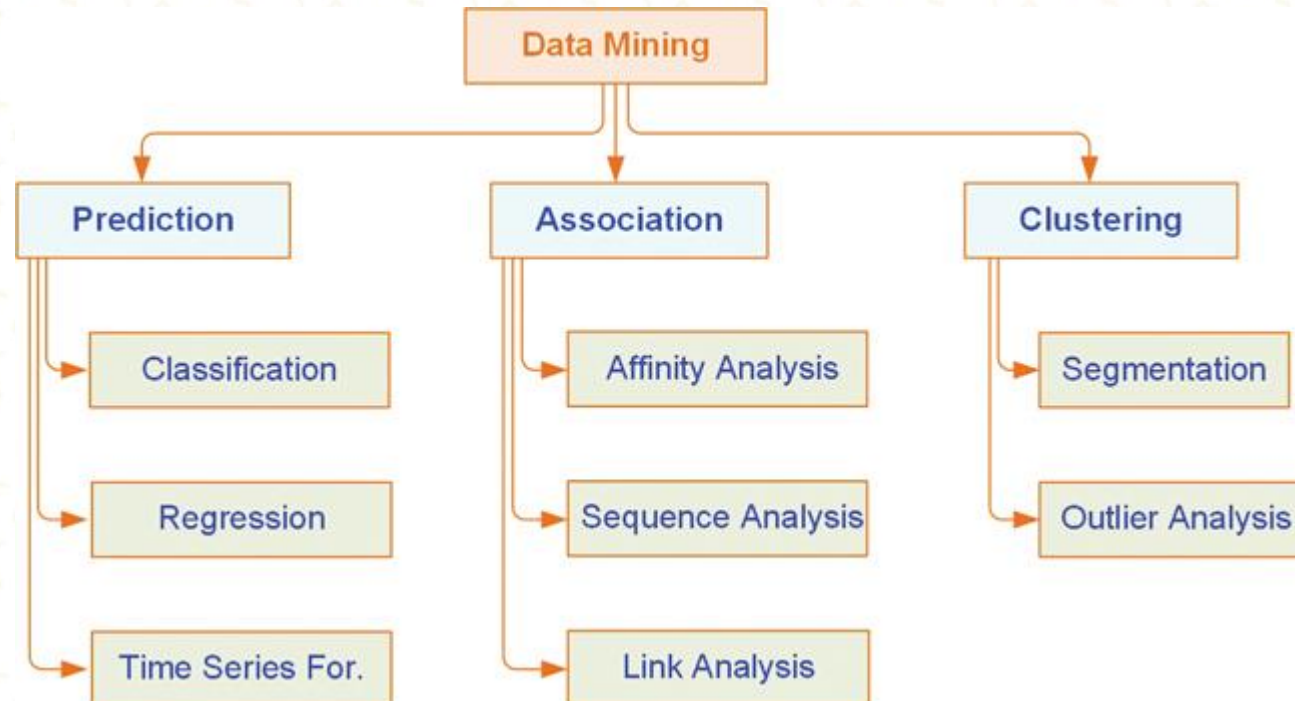


WEB CONNECTION
AGGREGATION INNOVATION BUSINESS
SOFTWARE INFORMATION CODE SOLUTION COLLECT BUG ERROR STORAGE
CONCEPT TRANSFORMATION EXTRACTION NETWORK ANALYTICS
ELABORATION BINARY INTERNET ANALYSIS
PATTERN PROCESS
DATA MINING
RECOGNITION
SYSTEM CALCULATION
MANAGEMENT DECISION SECURITY COMPUTER TECHNOLOGY
CONNECTIVITY
MACHINE LEARNING
CUSTOMER TROUBLESHOOT





Tasks in Data Mining





Tasks in Data Mining

Prediction

- Menggunakan beberapa variabel (atribut) untuk memprediksi nilai yang tidak diketahui atau nilai yang akan datang dari variabel (atribut) lain.
- Menemukan model (fungsi) yang menjelaskan dan membedakan kelas atau konsep untuk prediksi mendatang

Contoh : Memprediksi harga rumah berdasarkan atribut-atributnya seperti luas tanah, jumlah kamar, dsb.





Tasks in Data Mining

Association

- Mencari keterhubungan antara dua atau lebih kejadian.

Contoh :

Customer membeli kopi juga membeli gula

Jika stok payung habis, stok berikutnya yang mungkin habis adalah jas hujan (substitusi)





Tasks in Data Mining

Clustering







- Dengan label yang tidak diketahui, melakukan pengelompokan data (objek, kejadian, dsb) menjadi grup-grup yang antar grup masing-masing memiliki karakteristik yang similar.











Contoh Kasus Pola Prediksi: Decision Tree



STEP 1

Gender	Age	App
F	15	
F	25	
M	32	
F	40	
M	12	
M	14	

Gender	Age	App
F	15	
F	25	
M	32	
F	40	
M	12	
M	14	

Gender	Age	App
F	15	
F	25	
M	32	
F	40	
M	12	
M	14	

STEP 2

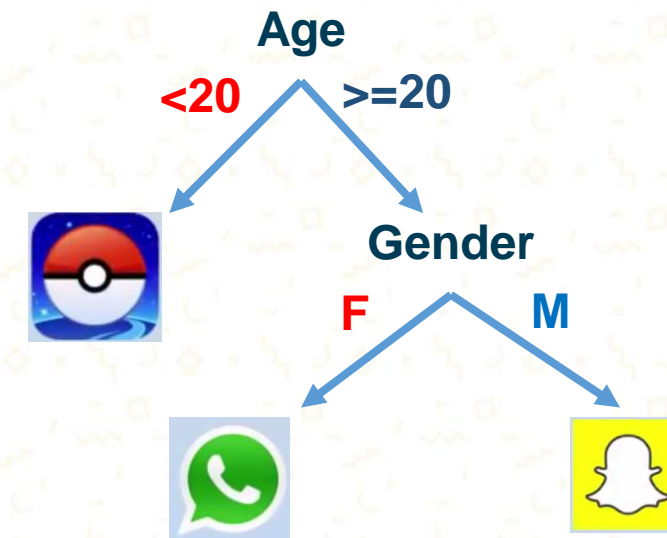
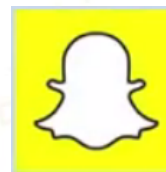
Gender	Age	App
F	25	
M	32	
F	40	

Test Subject



Antonio

Male
20 years old



Thank You

