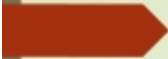# Machine Learning/Data Mining

Instructor: Chris Ding
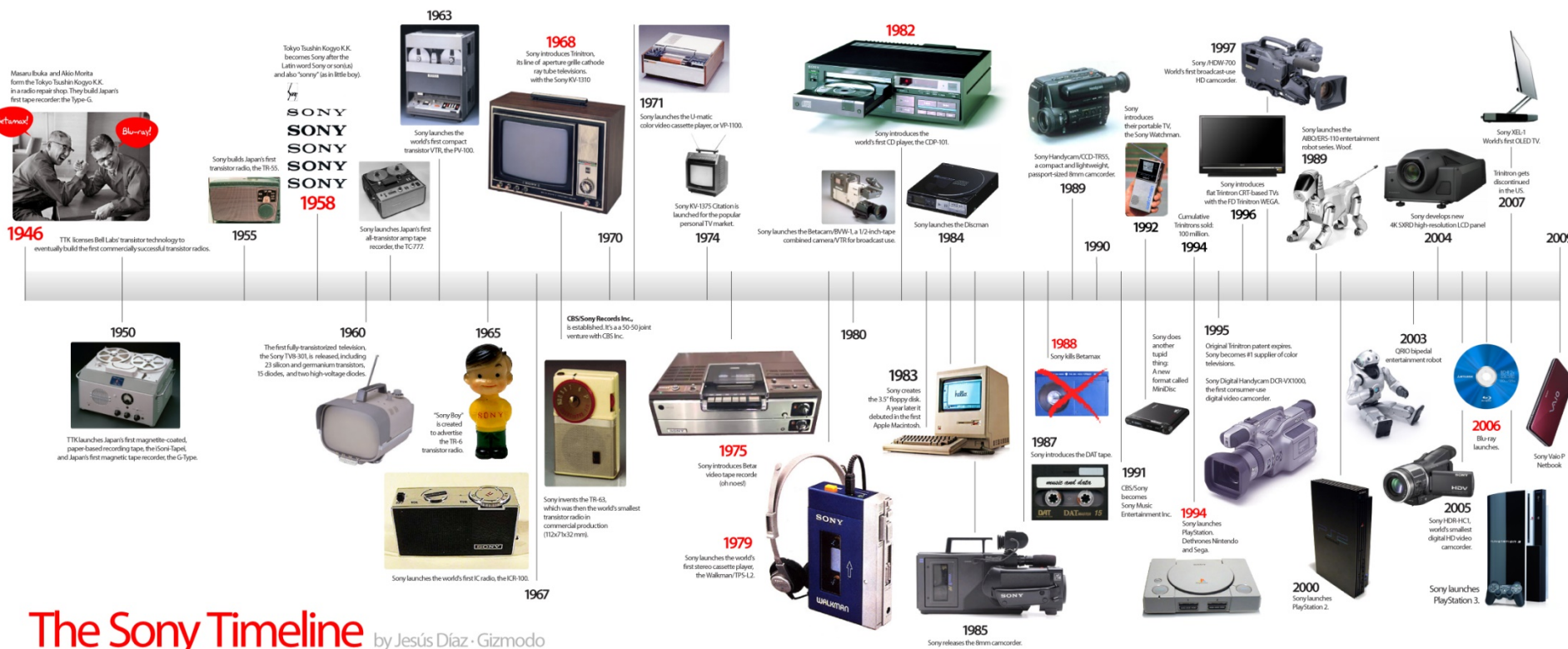
Information Technology Revolution

Revolution

Technology(and I.T) is advanced, such as with the advent of personal computer in the 1980s and its everyday use in the home and the workplaces. The world moved into the information age. I.T changed and it is changed our lifestyle.

The Sony Timeline
by Jesús Díaz · Gizmodo

1946
TTK licenses Bell Labs' transistor technology to eventually build the first commercially successful transistor radios.

Masaru Ibuka and Akio Morita form the Tokyo Tsushin Kogyo K.K. in a radio repair shop. They build Japan's first tape recorder: the Type-G.

Betamax!

Blu-ray!

1950
TTK launches Japan's first magnetite-coated, paper-based recording tape, the iSoni-Tapei, and Japan's first magnetic tape recorder, the G-Type.

1955
Sony builds Japan's first transistor radio, the TR-55.

1958
Tokyo Tsushin Kogyo K.K. becomes Sony after the Latin word Sony or son(u.s) and also "sonny" (as in little boy).

SONY
SONY
SONY
SONY
SONY
SONY

Sony launches Japan's first all-transistor amp tape recorder, the TC-777.

1960
The first fully-transistorized television, the Sony TV8-301, is released, including 23 silicon and germanium transistors, 15 diodes, and two high-voltage diodes.

1963
Sony launches the world's first compact transistor VTR, the PV-100.

1965
"Sony Boy" is created to advertise the TR-6 transistor radio.

Sony invents the TR-63, which was then the world's smallest transistor radio in commercial production (112x71x32 mm).

1967
Sony launches the world's first IC radio, the ICR-100.

1968
Sony introduces Trinitron, its line of aperture grille cathode ray tube televisions, with the Sony KV-1310.

1970

1971
Sony launches the U-matic color video cassette player, or VP-1100.

Sony KV-1375 Citation is launched for the popular personal TV market.

1974

CBS/Sony Records Inc., is established. It's a a 50-50 joint venture with CBS Inc.

1975
Sony introduces Betar video tape recorde (oh noes!)

1979
Sony launches the world's first stereo cassette player, the Walkman/TPS-L2.

SONY
WALKMAN

1980

1982
Sony introduces the world's first CD player, the CDP-101.

Sony launches the Betacam/BVW-1, a 1/2-inch-tape combined camera/VTR for broadcast use.

1983
Sony creates the 3.5" floppy disk. A year later it debuted in the first Apple Macintosh.

hello

1984
Sony launches the Discman.

1985
Sony releases the 8mm camcorder.

1987
Sony introduces the DAT tape.

music and data
DAT   DATmaster 15

1988
Sony kills Betamax.

1989
Sony Handycam/CCD-TR55, a compact and lightweight, passport-sized 8mm camcorder.

1990

1991
CBS/Sony becomes Sony Music Entertainment Inc.

1992
Sony introduces their portable TV, the Sony Watchman.

1994
Sony does another tupid thing: A new format called MiniDisc

1994
Sony launches PlayStation. Dethrones Nintendo and Sega.

1995
Original Trinitron patent expires. Sony becomes #1 supplier of color televisions.

Sony Digital Handycam DCR-VX1000, the first consumer-use digital video camcorder.

Cumulative Trinitrons sold: 100 million.

1996
Sony introduces flat Trinitron CRT-based TVs with the FD Trinitron WEGA.

1997
Sony /HDW-700 World's first broadcast-use HD camcorder.

Sony launches the AIBO/ERS-110 entertainment robot series. Woof.

1989

2000
Sony launches PlayStation 2.

2003
QRIO bipedal entertainment robot

2004
Sony develops new 4K SXRD high-resolution LCD panel

2005
Sony HDR-HC1, world's smallest digital HD video camcorder.

Sony launches PlayStation 3.

2006
Blu-ray launches.

Sony Vaio P Netbook

2007
Sony XEL-1 World's first OLED TV.

Trinitron gets discontinued in the US.
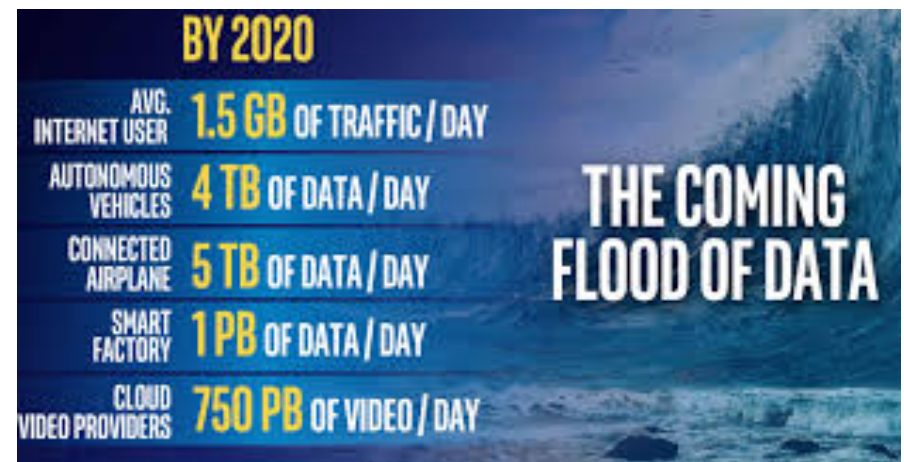
2009

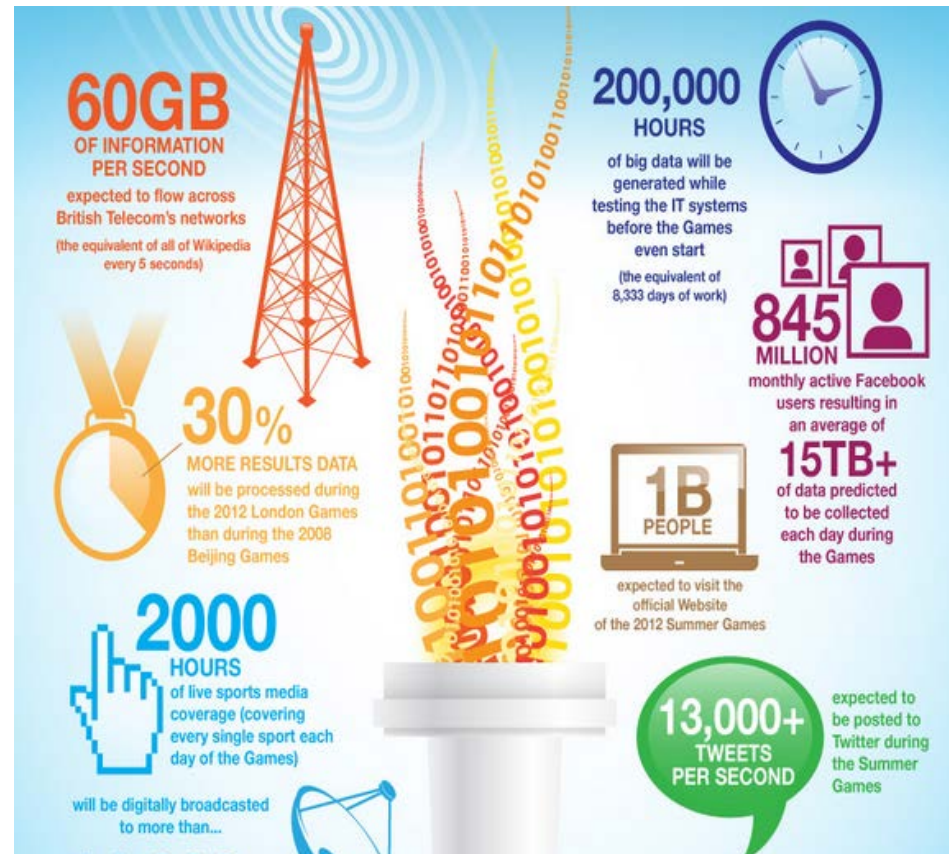Then comes the internet ---

The entire world is under your fingertips

A cellphone is most powerful knowledge and communication tool/database

It uses many many data mining technology

ARE YOU READY FOR THE
BIG DATA EXPLOSION?

More info is entering organisations than ever before, and it all requires management.

BY 2020

AVG. INTERNET USER	**1.5 GB**	OF TRAFFIC / DAY
AUTONOMOUS VEHICLES	**4 TB**	OF DATA / DAY
CONNECTED AIRPLANE	**5 TB**	OF DATA / DAY
SMART FACTORY	**1 PB**	OF DATA / DAY
CLOUD VIDEO PROVIDERS	**750 PB**	OF VIDEO / DAY

THE COMING FLOOD OF DATA

The Unstructured Data Explosion

- Growing 100X every 10 years
- Requires a new approach

New Data

Hi-Res: 50,000,000 KB / object

Video: 5,000,000 KB / movie

Audio: 5,000 KB / song

Images: 1,000 KB / picture

Text: 500 KB / record

Traditional Data

58674322
98323456
20419335
49567053

the quick brown fox jumped over the lazy brown dog.

Numbers: 5 KB / record

60GB OF INFORMATION PER SECOND expected to flow across British Telecom's networks (the equivalent of all of Wikipedia every 5 seconds)

200,000 HOURS of big data will be generated while testing the IT systems before the Games even start (the equivalent of 8,333 days of work)

30% MORE RESULTS DATA will be processed during the 2012 London Games than during the 2008 Beijing Games

845 MILLION monthly active Facebook users resulting in an average of **15TB+** of data predicted to be collected each day during the Games

1B PEOPLE expected to visit the official Website of the 2012 Summer Games

2000 HOURS of live sports media coverage (covering every single sport each day of the Games) will be digitally broadcasted to more than...

13,000+ TWEETS PER SECOND expected to be posted to Twitter during the Summer Games

What Happens in an **Internet Minute?**

639,800 GB of global IP data transferred

135
Botnet infections

6
New Wikipedia articles published

1,300
New mobile users

20
New victims of identity theft

204 million
Emails sent

47,000
App downloads

$83,000
In sales

100+
New Linkedin accounts

61,141
Hours of music

20 million
Photo views

3,000
Photo uploads

320+
New Twitter accounts

100,000
New tweets

277,000
Logins

6 million
Facebook views

2+ million
Search queries

And **Future Growth** is **Staggering**

30
Hours of video uploaded

1.3 million
Video views

Today, the number of **networked devices** = the global population

By **2015**, the number of **networked devices** = **2x** the global population

In **2015**, it would take you **5 years** to view all video crossing IP networks each **second**

(intel)

The New Technology:

data mining, machine learning, AI

Turning data into knowledge

Search Engine Google uses many data mining technology

Google

data mining 🔍

All   News   Books   Images   Videos   More   Settings   Tools

About 55,000,000 results (0.54 seconds)

Data Mining - ibm.com
[Ad] www.ibm.com/PredictiveAnalytics ▾
Data Mining Through IBM Allows For Analysis From Various Models.
   Platform & Technology   IBM Analytics for Spark
   IBM Cloud Data Services   Advanced Analytics

How Data Mining Works - Free White Paper Download - sas.com
[Ad] www.sas.com/**Data**-**Mining**/White-Papers ▾
Applying **data mining** and real-time analytics in oil and gas operations.
Machine Learning · Predictive Analytics · Business Analytics · Statistical Analytics
SAS Analytics Products · Analytics White Papers · Analytics Webinars · How to Buy SAS Analytics

da·ta min·ing

noun   COMPUTING

   the practice of examining large databases in order to generate new information.

      ﹀   Translations, word origin, and more definitions

                                                                          Feedback

Data mining - Wikipedia
https://en.wikipedia.org/wiki/**Data_mining** ▾

IBM Watson

SPEECH TO TEXT	Employs low latency speech recognition capabilities to convert English speech to text
TEXT TO SPEECH	Synthesizes natural-sounding speech from text in English and Spanish
VISUAL RECOGNITION	Analyzes the visual content of images and videos to understand their content
CONCEPT INSIGHTS	Explores the concepts behind your input, identifying associations beyond traditional text matching
TRADEOFF ANALYTICS	Helps users make better choices by weighing multiple and often conflicting goals

IBM Watson uses many many machine learning technology

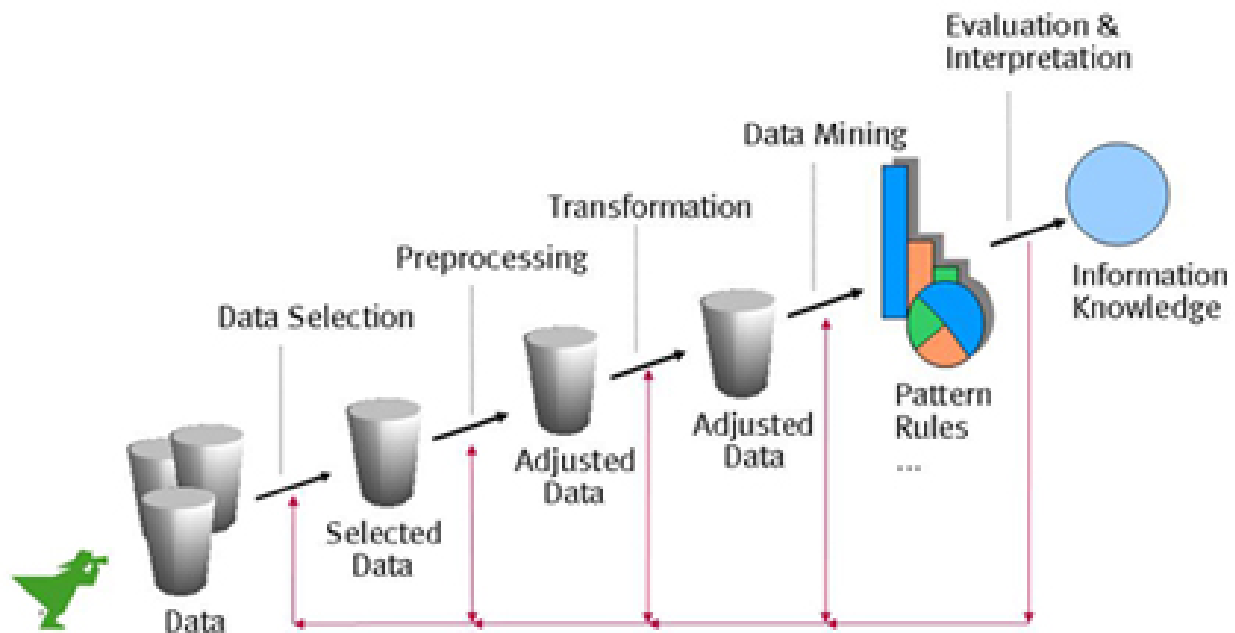Security. One that Watson should help mitigate.

Data Mining: extract knowledge from data
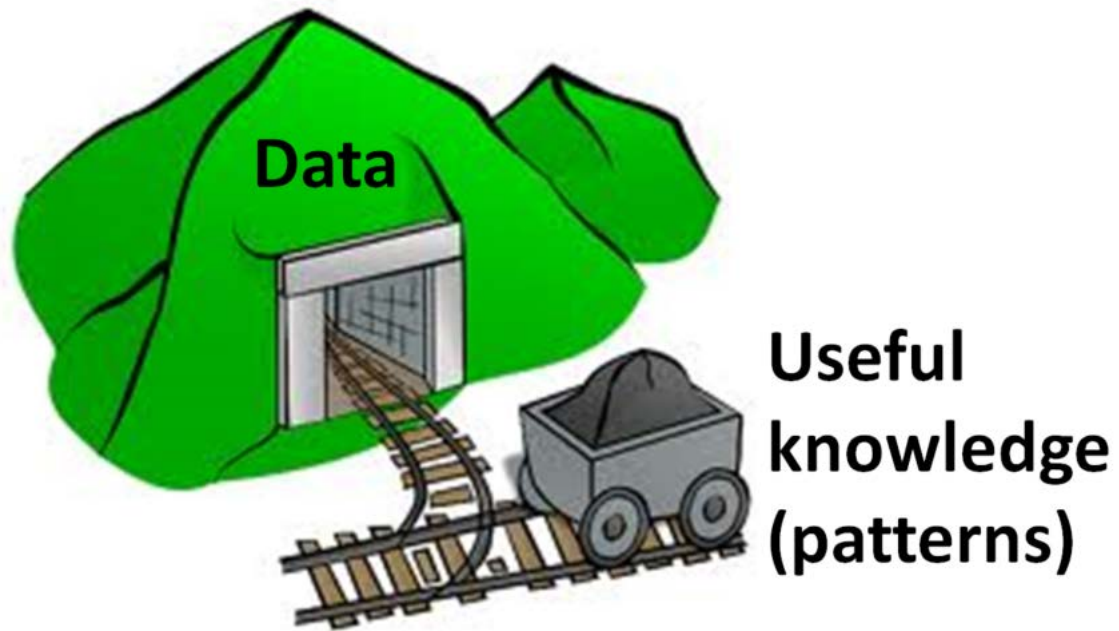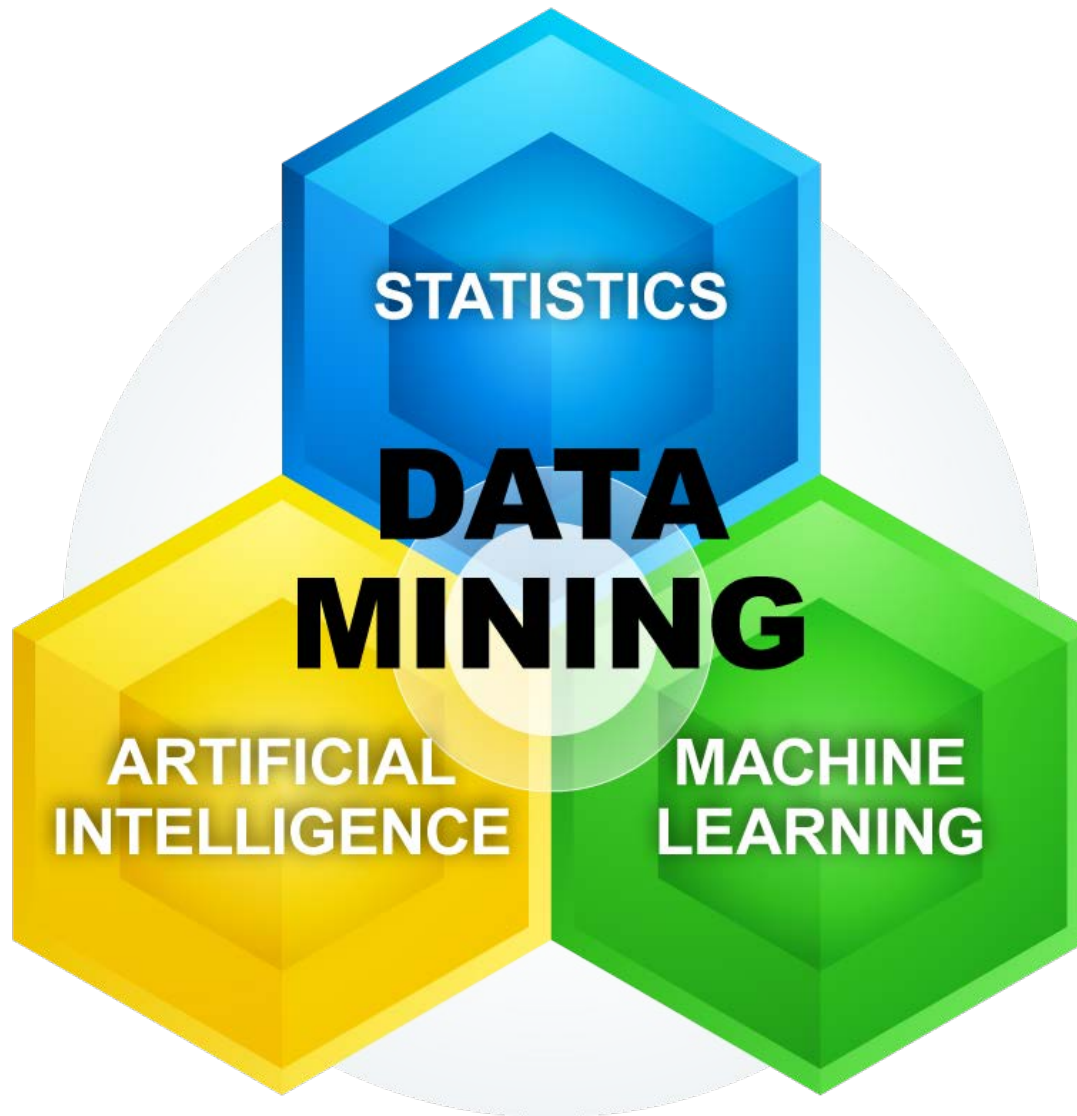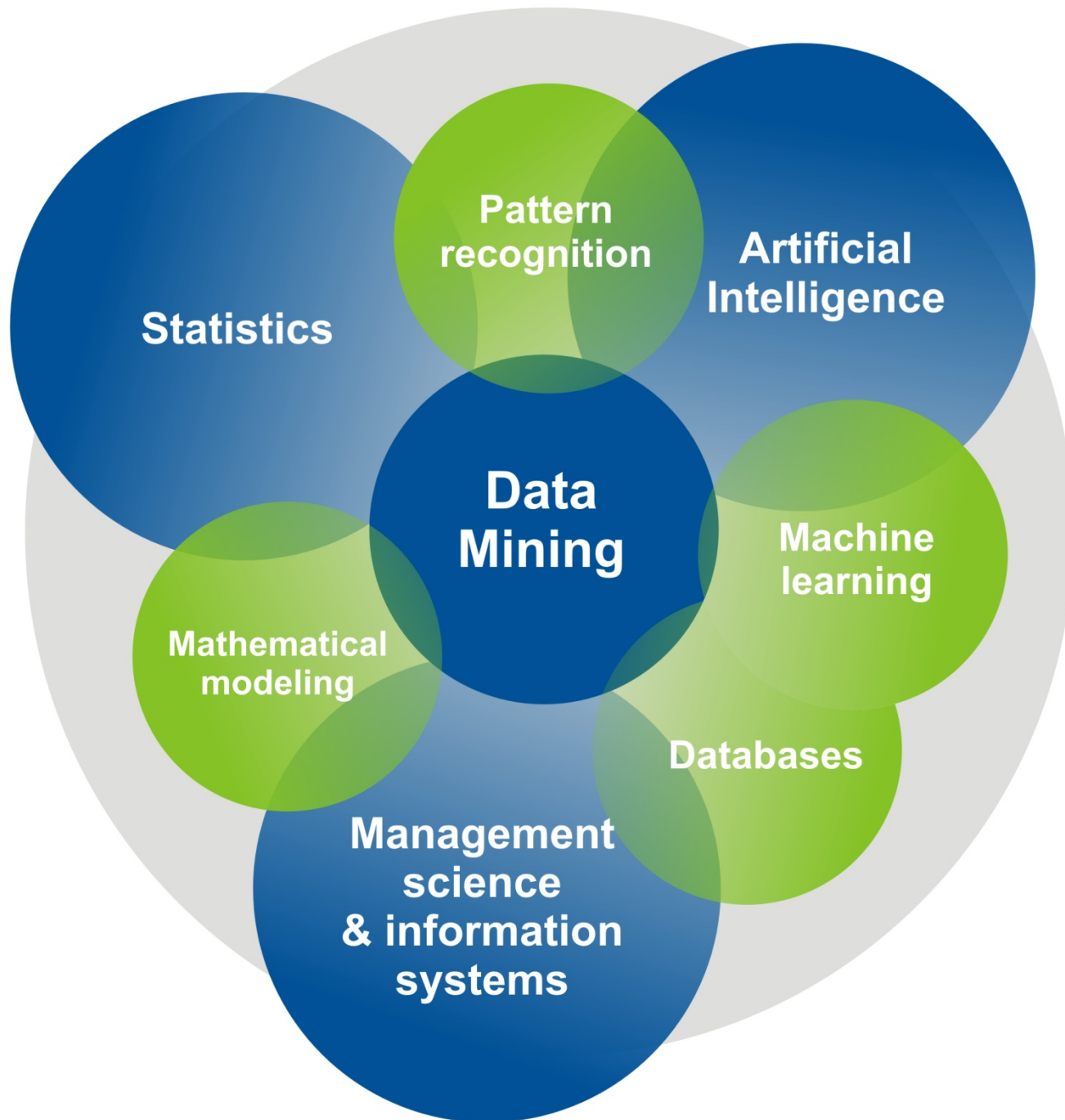
Data Mining Model

3 Examples of Data Mining

- Market basket data
- Handwritten digits recognition
- Cancel detection from gene expression data

Market Basket Data Analysis

Market Basket Example

? Where should detergents be placed in the Store to maximize their sales?

? Are window cleaning products purchased when detergents and orange juice are bought together?

? Is soda typically purchased with bananas? Does the brand of soda make a difference?

? How are the demographics of the neighborhood affecting what customers are buying?

Image source: deepclimate.org

Market Basket Data Analysis: find associations

Of transactions that included milk:
- 71% included bread
- 43% included eggs
- 29% included toilet paper
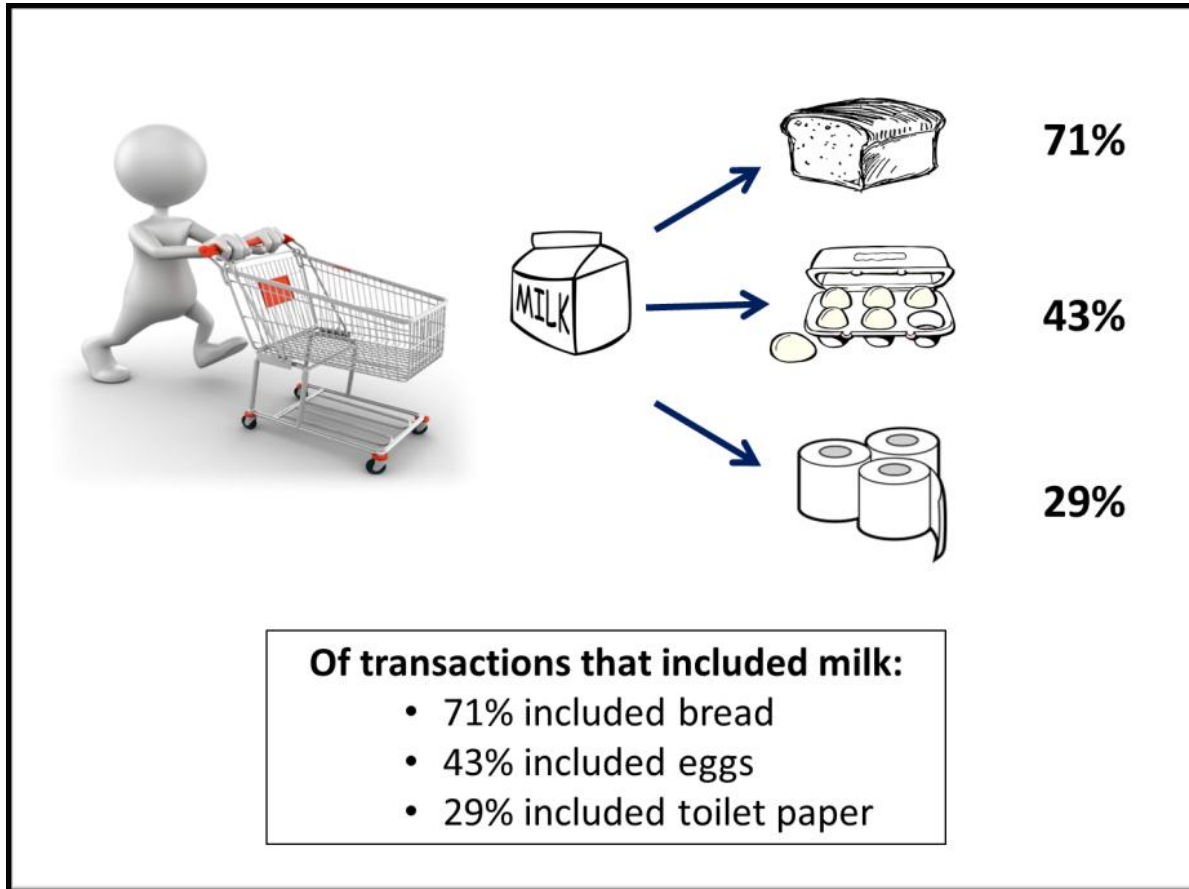
Table 1. Market basket transactions

Transaction ID	Items Bought
1	{Laptop, Printer, Tablet, Headset}
2	{Printer, Monitor, Tablet}
3	{Laptop, Printer, Tablet, Headset}
4	{Laptop, Monitor, Tablet, Headset}
5	{Printer, Monitor, Tablet, Headset}
6	{Printer, Tablet, Headset}
7	{Monitor, Tablet}
8	{Laptop, Printer, Monitor}
9	{Laptop, Tablet, Headset}
10	{Printer, Tablet}

Market Basket Analysis
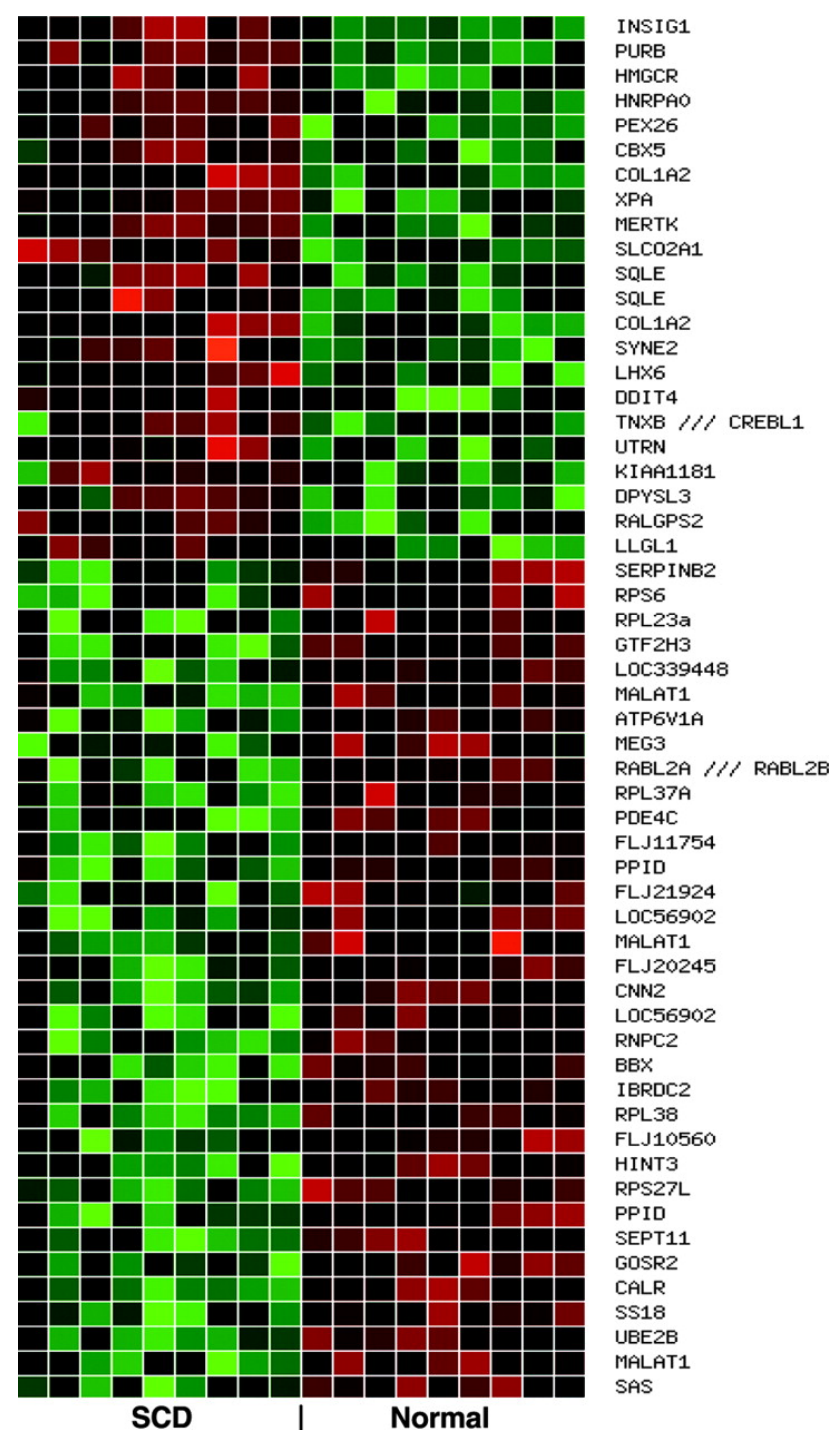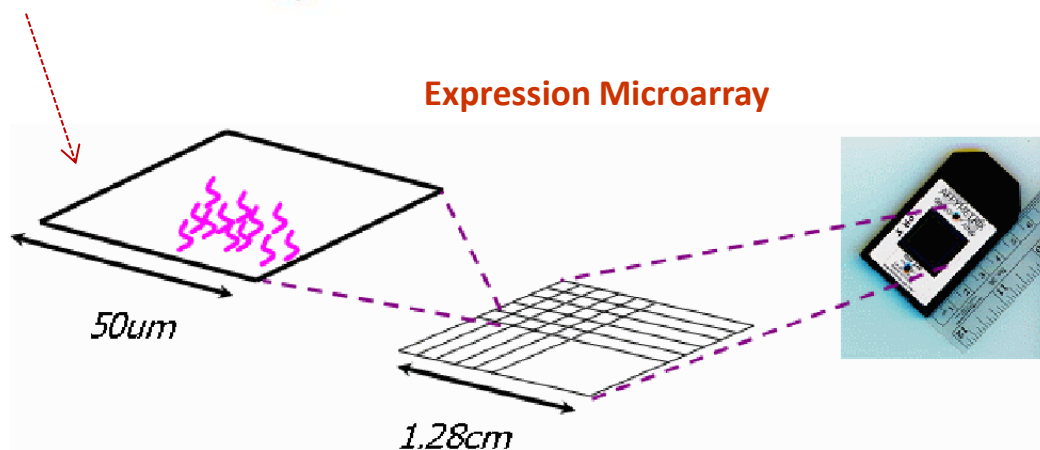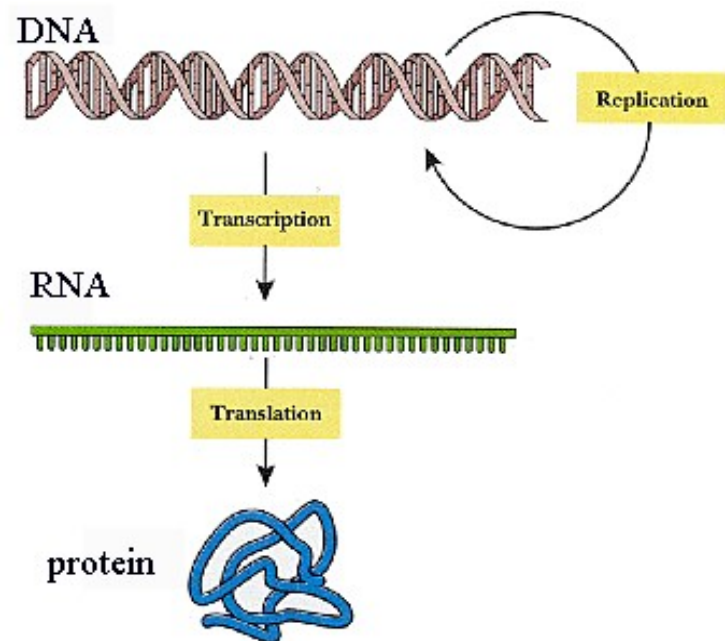
An example of market basket transactions.

TID -- Items
1. {Bread, Milk}
2. {Bread, Diapers, Beer, Eggs}
3. {Milk, Diapers, Beer, Cola}
4. {Bread, Milk, Diapers, Beer}
5. {Bread, Milk, Diapers, Cola}

DNA Gene Expressions to detect cancer

Expression Microarray

50um

1.28cm

INSIG1
PURB
HMGCR
HNRPA0
PEX26
CBX5
COL1A2
XPA
MERTK
SLCO2A1
SQLE
SQLE
COL1A2
SYNE2
LHX6
DDIT4
TNXB /// CREBL1
UTRN
KIAA1181
DPYSL3
RALGPS2
LLGL1
SERPINB2
RPS6
RPL23a
GTF2H3
LOC339448
MALAT1
ATP6V1A
MEG3
RABL2A /// RABL2B
RPL37A
PDE4C
FLJ11754
PPID
FLJ21924
LOC56902
MALAT1
FLJ20245
CNN2
LOC56902
RNPC2
BBX
IBRDC2
RPL38
FLJ10560
HINT3
RPS27L
PPID
SEPT11
GOSR2
CALR
SS18
UBE2B
MALAT1
SAS

SCD | Normal

Recognize handwritten digits and letters.

All post office mailed letters are machine scanned to automatically decide the destination zip-code