

Lecture 0

Introduction to the Course

Kim, Yang Sok
Dept. of MIS, Keimyung University

- Who am I?
- What is Data?
- What is Business Data Mining?
- What & How We Learn in this Class?

Who am I?

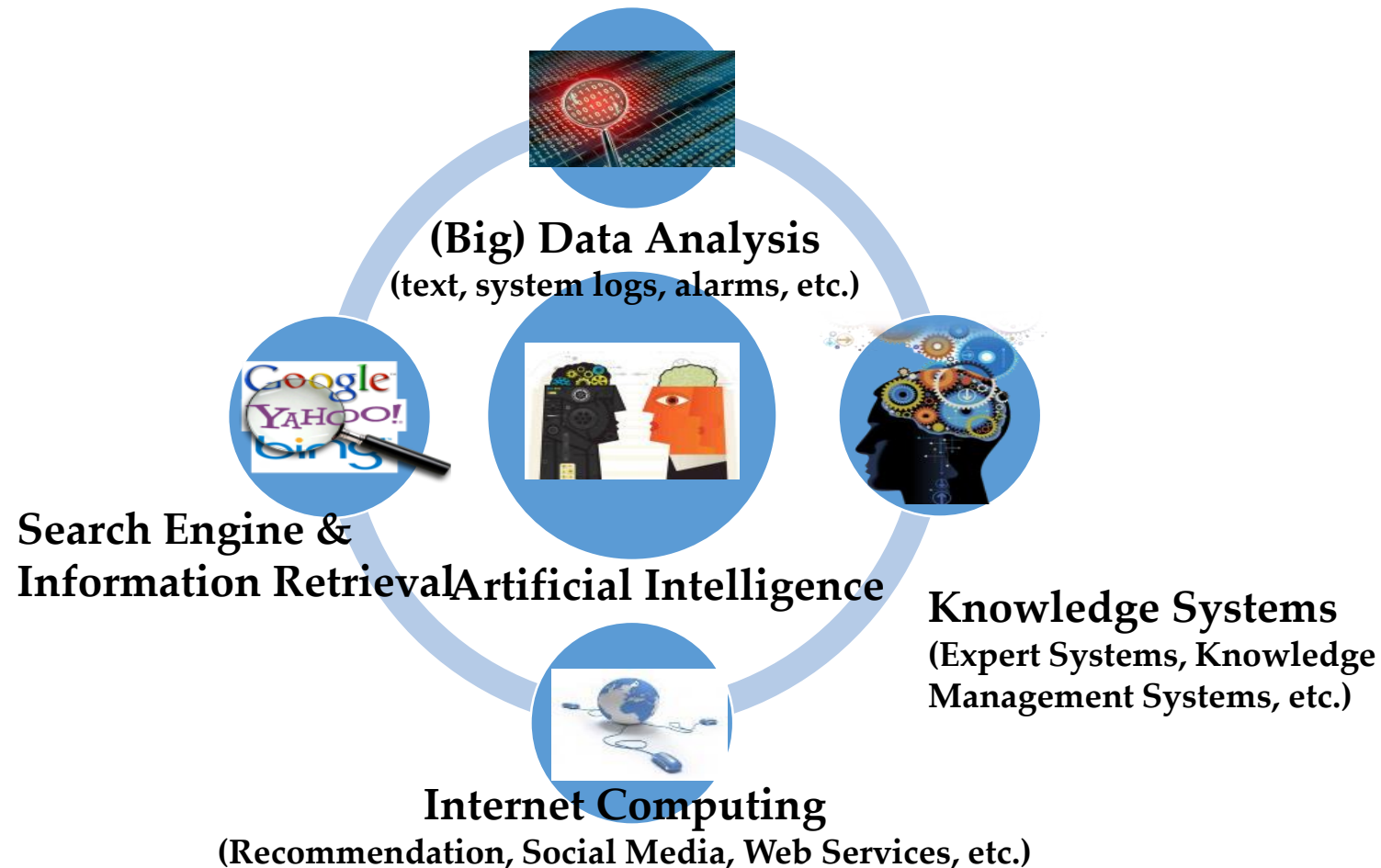
Who am I?

Short Biography

- Kim, Yang Sok
- Studied **economics** at the University of Seoul (1987 – 1995)
- Worked as a **business management specialist** at Hyundai Information Technology Co., Ltd. (1994 – 2001)
- Studied **computer science** at the University of Tasmania (2002 – 2009)
- Worked as a **researcher** at the University of New South Wales and at the University of Tasmania (2009 - 2004)
- Worked at Keimyung University since 2015

Who am I?

Interested in ...



You can find my research papers from DBLP (http://dblp1.uni-trier.de/pers/hd/k/Kim:Yang_Sok)

Who am I?

How to find me?

- Office: Room 333@ 의양관 별관
- Tel: 6147
- Email: YangSok.Kim@kmu.ac.kr
- Consultation Time: Wednesday 10:30 –12:00

If you wish to meet me, please make a booking through Edward System

What is Data?

What is Data?

Definition

- 1 : **Factual information** (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation
- 2 : **information in digital form** that can be transmitted or processed
- 3 : **information output by a sensing device or organ** that includes both useful and irrelevant or redundant information and must be processed to be meaningful



What is Data?

Basic Data Types

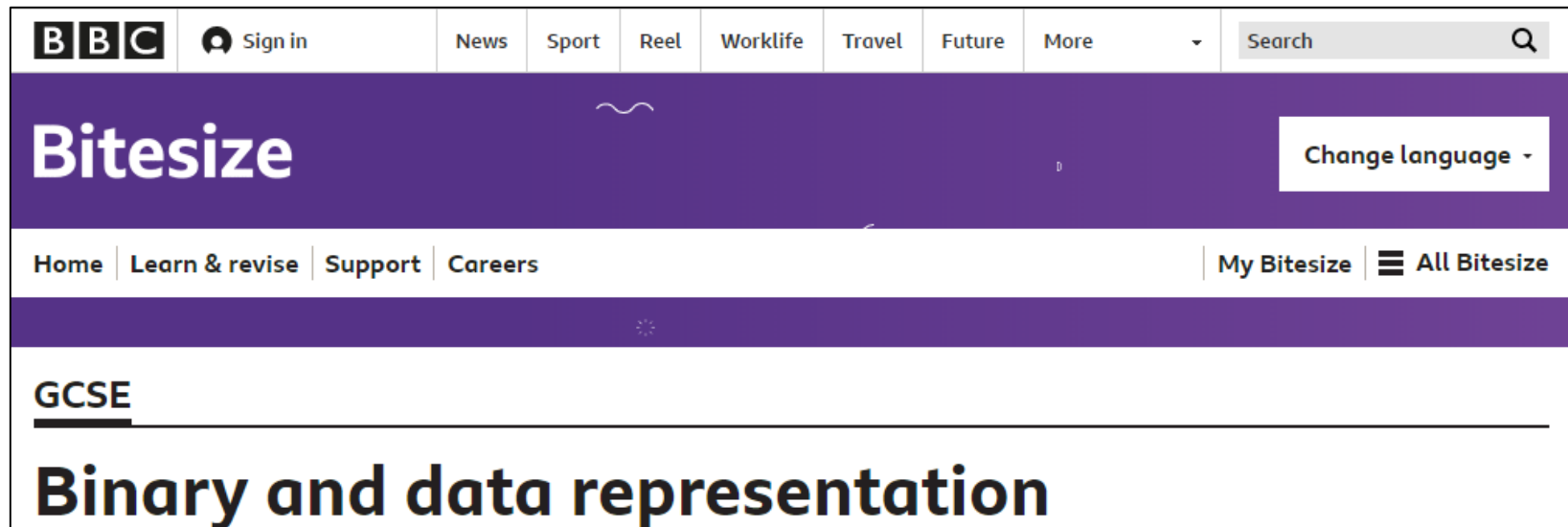
- **Alphanumeric data**
 - Combination of numbers and letters
- **Text data**
 - Words, sentences and paragraphs used in writing communication
- **Image data**
 - Graphics, shapes, figures etc.
- **Audio**
 - Human voice & other sounds



What is Data?

Representing Data in Digital Format

- All data types can be represented by digital format, namely by bits.
- This course does not teach data representation method, but you can read the following web pages to understand digital representation of data.
 - Binary and data representation - <https://www.bbc.co.uk/bitesize/topics/zd2xsbk>



What is Data?

Processing & Context

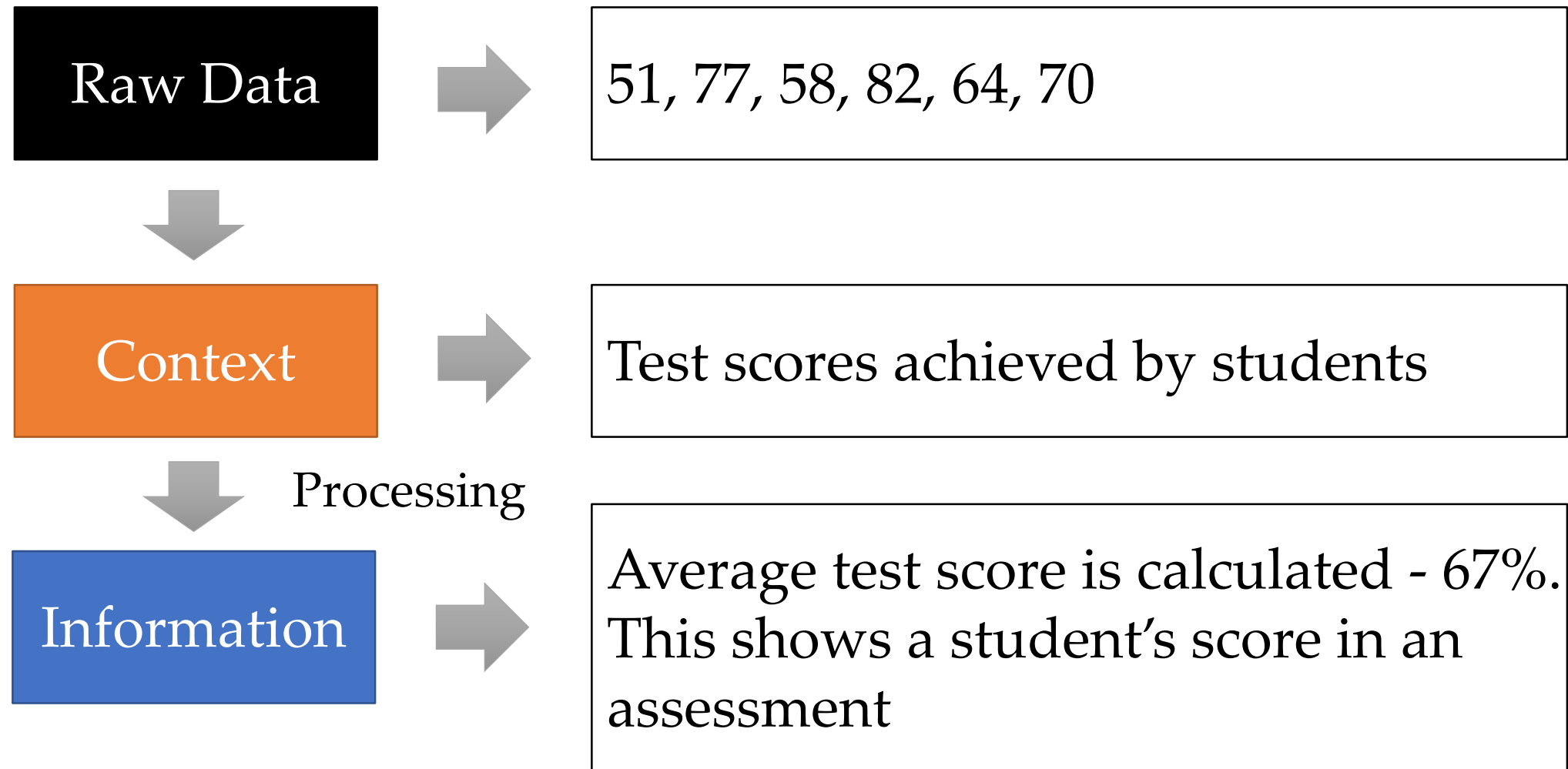
What do these numbers mean?

51, 77, 58, 82, 64, 70

Without processing and/or context this data is meaningless.

What is Data?

Processing & Context



What is Data Mining?

What is Data Mining?

Why?

- **Data grows explosively!**
- **Data collection and data availability**
 - Automated data collection tools, database systems, Web, computerized society
- **Major sources of abundant data**
 - Business: Web, e-commerce, transactions, stocks, ...
 - Science: Remote sensing, bioinformatics, scientific simulation, ...
 - Society and everyone: news, digital cameras,
- **We are drowning in data, but starving for knowledge!**
- **“Necessity is the mother of invention” — Data mining enable automated analysis of massive data sets**

What is Data Mining?

Definition

- **Data mining (knowledge discovery from data)**
 - Extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data
- **Alternative name**
 - Knowledge discovery in databases (KDD)
- **Watch out: Is everything “data mining”?**
 - Query processing
 - Expert systems or statistical programs

What is Data Mining?

Potential Applications

- **Market analysis and management**
 - Target marketing, customer relationship management (CRM), market basket analysis, market segmentation
- **Risk analysis and management**
 - Forecasting, customer retention, quality control, competitive analysis
- **Fraud detection and detection of unusual patterns (outliers)**
- **Text mining and Web mining**
 - news group, email, documents, etc.
- **Stream data mining**
- **Bioinformatics and bio-data analysis**

What is Data Mining?

Potential Applications

Walmart increase sales by analyzing customers' behavior!

Walmart's market researcher found that when male customers visit the baby department to pick up diapers for their little ones, it's very likely for them to buy themselves a couple of beers. Thus, Walmart relocated the beers next to diapers, which had led the sales of beers and diapers increased significantly.

market basket
analysis



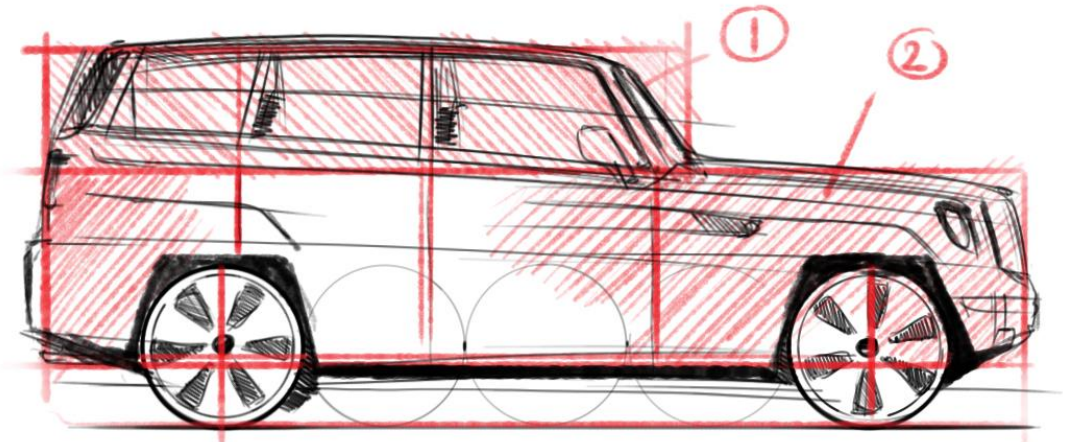
What is Data Mining?

Potential Applications

Ford improves vehicle models via the social media platform!

Ford's R&D team once made an analysis on the ways to open up the back truck of their SUV's (such as open manually or automatically). Though their routine surveys do not reflect this as a potential issue as no customer was reporting it, by analyzing the social platforms, Ford's team found that people actually talked over it a lot.

**Text mining with
social media data**



Gracenote analyze user behaviors to suggest songs!

Gracenote got the technics that use the built-in microphones on smartphones and tablets to recognize songs played on user's TV or stereo, detect reactions such as applause or boos, and even detect if the user has turned up the volume. In this way, Gracenote can study the songs that are liked by the users and the specific time and place when this song is played.



Recommendation

What is Data Mining?

Potential Applications

PredPol forecast crime scenes

PredPol, working with Los Angeles and Santa Cruz police and a team of researchers, predicts the odds of a crime occurring based on a variation of earthquake prediction algorithms and crime data that can be accurate to within 500 square feet. In Los Angeles, where the algorithm was applied, the distribution of theft and violent crime dropped by 33% and 21% correspondingly.



Crime prediction

What is Data Mining?

Potential Applications

German football team use data mining to improve players' performance

The German football federation partnered with SAP AG in order to analyse video data and both individual and team performance. This allowed them to give individuals feedback as to how they can improve their performances and how they could integrate themselves better with one another.

Using analytics, they cut down average possession time from 3.4 seconds to 1.1 seconds, a critical improvement that made all the difference when they defeated Brazil in semi-final and against Argentina, where Mario Gotze's goal in extra-time gave them the most prestigious prize in sport.



**Sport data
analytics**

What & How We Learn in this Class?

What & How We Learn in this Class?

Goal of the Class

- **Students learn to analyze business data mining using Rapidminer, a predictive analytics tool.**
- **Students learn about basic knowledge on business data mining.**
- **Students learn practical skills for analyzing data using Rapidminer.**
- **Students conduct a team-based data mining project.**

- **Textbook**

- 토닥토닥 래피드마이너

- **Supplementary**

- A. Chisholm, Exploring Data with Rapidminer
 - RapidMiner Manual
(<http://docs.rapidminer.com/downloads/RapidMiner-v6-user-manual.pdf>)
 - Data Mining for the Masses
(<http://docs.rapidminer.com/downloads/DataMiningForTheMasses.pdf>)

- **Lecture Notes**

- Available on the CTL unit website (교수학습시스템)

What & How We Learn in this Class?

Course Schedule

• 1st Term (Week 1 – Week 8)

Week	Contents	Assignment
1	Unit Overview	
2	CRISP-DM: A Data Mining Methodology	
3	Business Understanding	
4	Data Understanding with Rapidminer (1)	Assignment Specification Release
5	Data Understanding with Rapidminer (2)	
6	Data Preparation with Rapidminer(1)	
7	Data Preparation with Rapidminer(2)	
8	Middle Exam	

What & How We Learn in this Class?

Course Schedule

• 2nd Term (Week 9 – Week 16)

Week	Contents	Assignment
9	Modeling (1)	
10	Modeling (2)	
11	Modeling (3)	
12	Modeling (4)	
13	Evaluation & Deployment	Assignment Due
14	Project Presentation (1)	
15	Project Presentation (2)	
16	Final Exam	

What & How We Learn in this Class?

Assessment

- **Report/Project (30%)+Exams(60%)+Participation (10%)**
- **Report/Project (30%): Due Date: Week 13**
 - Data analysis with real world dataset.
 - Detailed assignment specification will be released in Week 4.
- **Exams (60%)**
 - Middle Exam(30%)
 - Final Exam (30%)
- **Participation (10%)**
 - Deduct 1 marks for an absent and 0.5 mark for late attendance.
 - If you lose 10 marks, you will receive F automatically.

What & How We Learn in this Class?

Report Assessment

- **Report assignment should be submitted via teaching & learning system(교수학습시스템)**
- **In addition to electronic submission, you also need to submit a printed copy. Please keep a copy of assignment by yourself.**
- **Submit assignment within the due date.**
 - Unless appropriate arrangements have been made, supported by a sensible and valid reason, late submissions will attract a penalty of 5% per day or 25% per week.
- **Students have a responsibility to uphold the university standards on ethical scholarship.**
 - Cheating, plagiarism, and falsification of data are dishonest practices that contravene academic values.

What & How We Learn in this Class?

Exam Assessment

- **Exam 1: In Class Exams**
 - A timed, proctored exam.
 - Typically last 1 hours.
- **Exam 2: In Class Exams**
 - A timed, proctored exam.
 - Typically last 1 hours.
- **Details of exam questions will be discussed during the class**

What & How We Learn in this Class?

Participation Assessment

- **Attending the class is important!!**
 - **Deduct 1 point for each absent**
 - **Deduct 0.5 point for each lateness**
- **Students should be aware that the subject will be examined on material covered in classes and it is the individual student's responsibility to ensure that they are sufficiently familiar with this material.**
- **Do not make a mistake by assuming that the materials provided online perfectly substitute for class attendance.**

What & How We Learn in this Class?

Announcement

- **Announcements are made throughout the unit, typically to clarify requirements for assignments.**
- **Any such announcements will be placed on the unit web site.**
- **Such announcements are deemed, within two working days, to be made to the whole group.**
- **Announcements made at an organized session are deemed to be made to the whole group.**



QUESTIONS?