

Social Studies and Communication

Introduction to Strategic Data Science

1st lecture

Ryuichiro Ishikawa

Lecture: 5th on Mon. @ Room 508 of Building 11

Office hour: 4th period on Mon. &
2nd period on Wed.

HOW TO DELIVER THE CLASS

- **Model Lecture:** Master's degree programs in University of Liverpool in the U.K.
 - Online text-based method.
- This lecture is based on the following textbook:
 - **Joel Grus (2019) Data Science from Scratch 2nd eds., O'Reilly.**
【邦訳】 菊池彰 訳「ゼロからはじめるデータサイエンス 第2版」オライリー.
 - ✓ Please buy either original or translated version of the textbook. The Univ. coop. does not put this. So try to get a bookstore or an online bookstore.
- The purpose of this lecture is to study
 1. *fundamental knowledge of statistics & data analysis;*
 2. *Basic grammars of "Python" to visualize/analyze data.*

COURSE OUTLINE

15th(1/25): Final Report instead of No lecture

	#	Date	Contents
Introduction to Statistics and Python	1 st	9/28	Introduction: Data Analytic Thinking
	2 nd	10/5	Quick study of Python & Visualizing data
	3 rd	10/12	Basic statistics: Data and their description
	4 th	10/19	Basic statistical analysis I: Inference
	5 th	10/26	Operation of data on Python
	6 th	11/2	Basic statistical analysis II: Hypothetical tests
	7 th	11/9	Introduction to machine learning: Similarity, Neighbors, and Clusters
Introduction to machine learning	8 th	11/16	Naïve Bayesian analysis
	9 th	11/23	Single regression analysis/Multiple regression analysis
	10 th	11/30	Logistic analysis
	11 th	12/7	Decision tree
	12 th	12/14	Natural language processing
	13 th	12/21	Recommend system
	14 th	1/18	Summary of lecture

3

HOW TO PROCEED THE CLASS

- **Lecture Note:** All the materials are uploaded on the Moodle. So please check the Moodle class page every week. (Due to a technical reason, we upload the materials of the first and second lectures on my web page)
- **Credit !** : How you get the credit of this course.
 - ✓ Class exercises 70% & Final report: 30%, respectively.

Under the COVID-19, this class is also restricted. So I need the evidences that you study in the class. So please submit all the assignments if you'd like to get the credit.

I approve of discussing the questions of assignments together as you work through them. I set the discussion threads on the Moodle. Please discuss on the threads which are considered as class contributions.

4

PURPOSE OF THIS LECTURE:

- **Taking new trends in Data Science:**

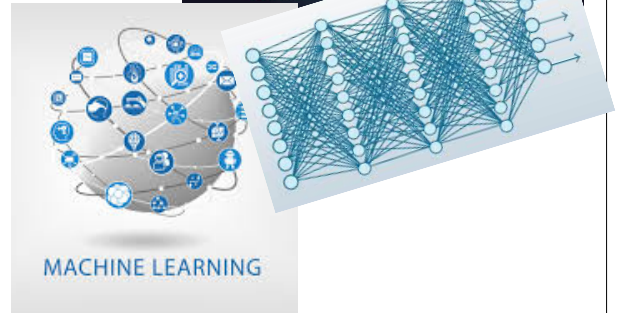
- ✓ Availability of 'Big data';
- ✓ Development of new algorithm;
- ✓ Provision of API on new algorithm.

- **Expanding Business by Data**

- Recommendation systems;
- Prediction of Customers' Behavior.

In this lecture, you study
Basic statistics to analyze data by *Python*.

What happen in the world



5

GET THINKING ON DATA

- Walmart is the first company to make use of customers' sales data:
 - Hurricane Frances in 2004 was on its way, threatening a direct hit on Florida's Atlantic coast. A week ahead of the storm's landfall, Linda M. Dillman, chief information officer, pressed her staff to come up with forecasts based on what had happened when Hurricane Charley struck several weeks earlier.
 - They found that the stores would indeed need...



Strawberry Pop-Tarts

with



6

DESCRIPTIVE AND INFERENTIAL STATISTICS

- **Descriptive statistics:** Compiling and organizing the collected data into tables and graphs.
- **Inferential statistics:** to sear out the overall characteristics of the collected data, including those that were not collected, from some of the data.

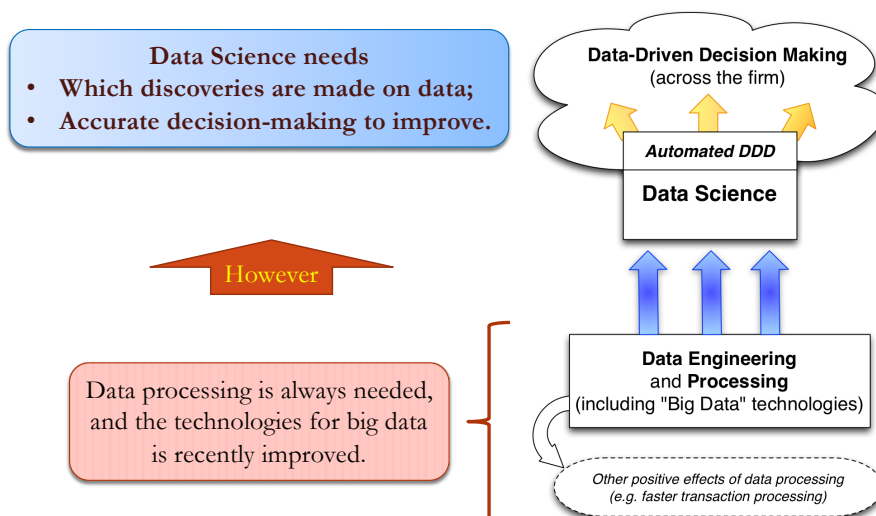
*There are three kinds of lies:
lies, damned lies, and statistics.*

By Benjamin Disraeli, the British Prime Minister
in the 19th century

7

SOME COMMENTS OF DATA PROCESSING

Data processing is basically NOT data science



8

INDIVIDUAL DATA

