Technology Science Information Networks Computing

Lecturer: Ting Wang (王挺)

利物浦大学计算机博士

清华大学计算机博士后

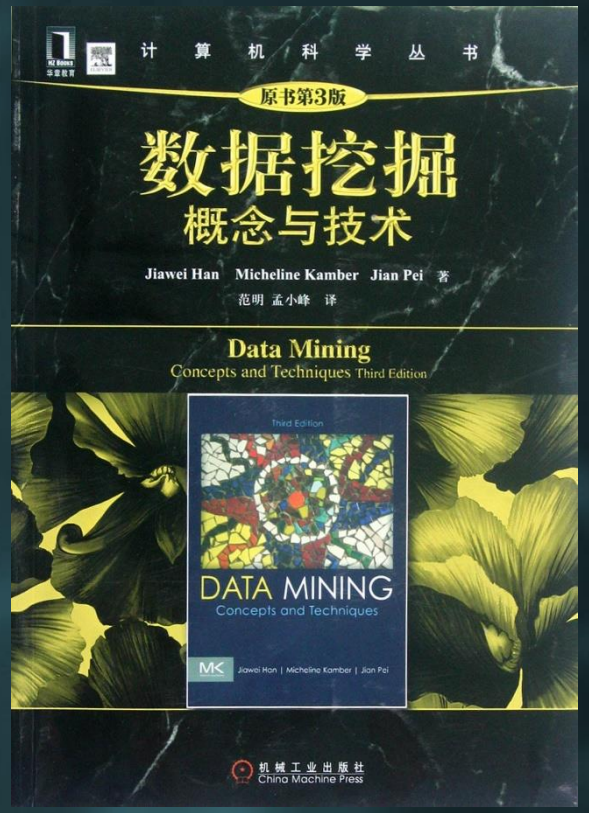
电子信息技术高级工程师

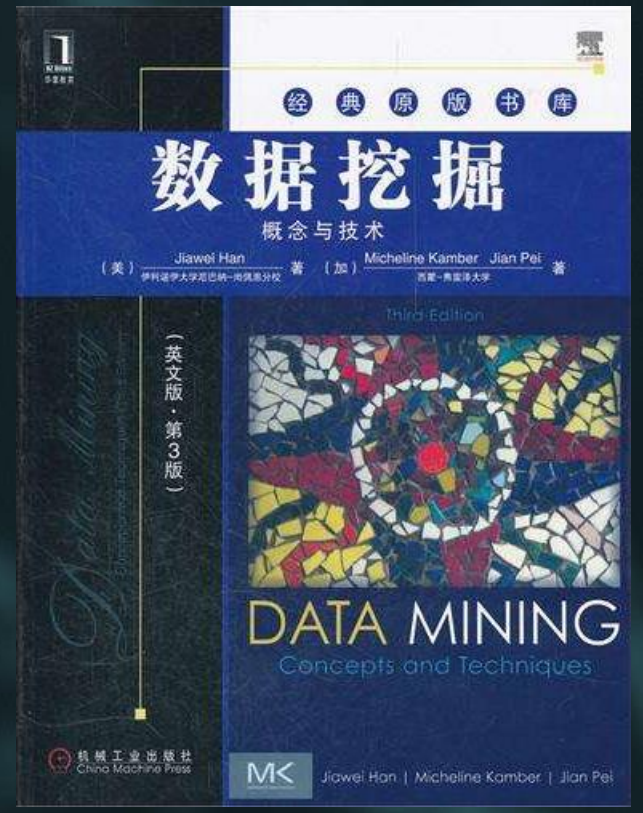
上海外国语大学网络与新媒体副教授

浙江清华长三角研究院海纳认知与智能研究中心主任

Chapter 1

Introduction to Data Mining

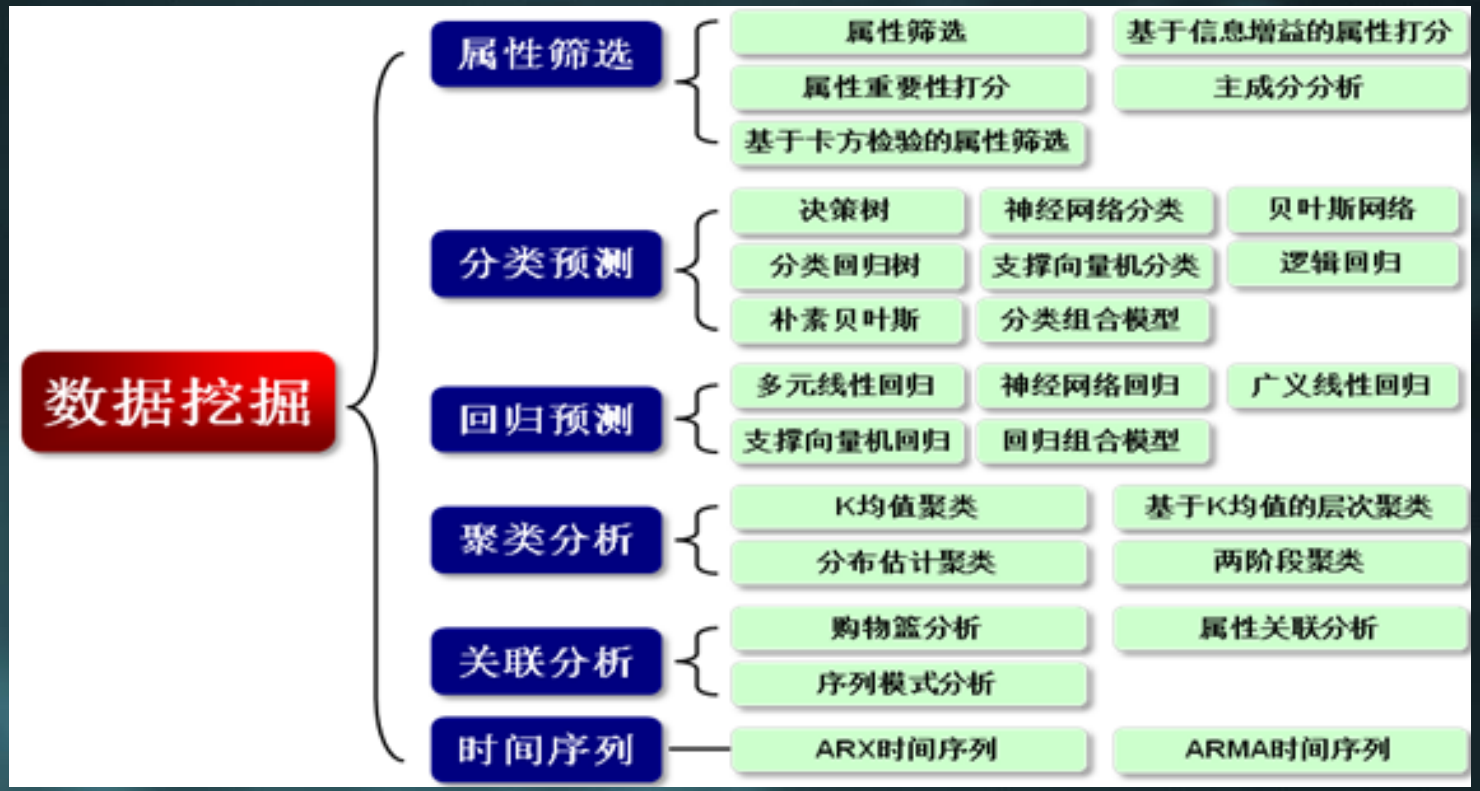




Chapter 1
Introduction to Data Mining

1. What is Data Mining? (The Concepts of Data Mining)

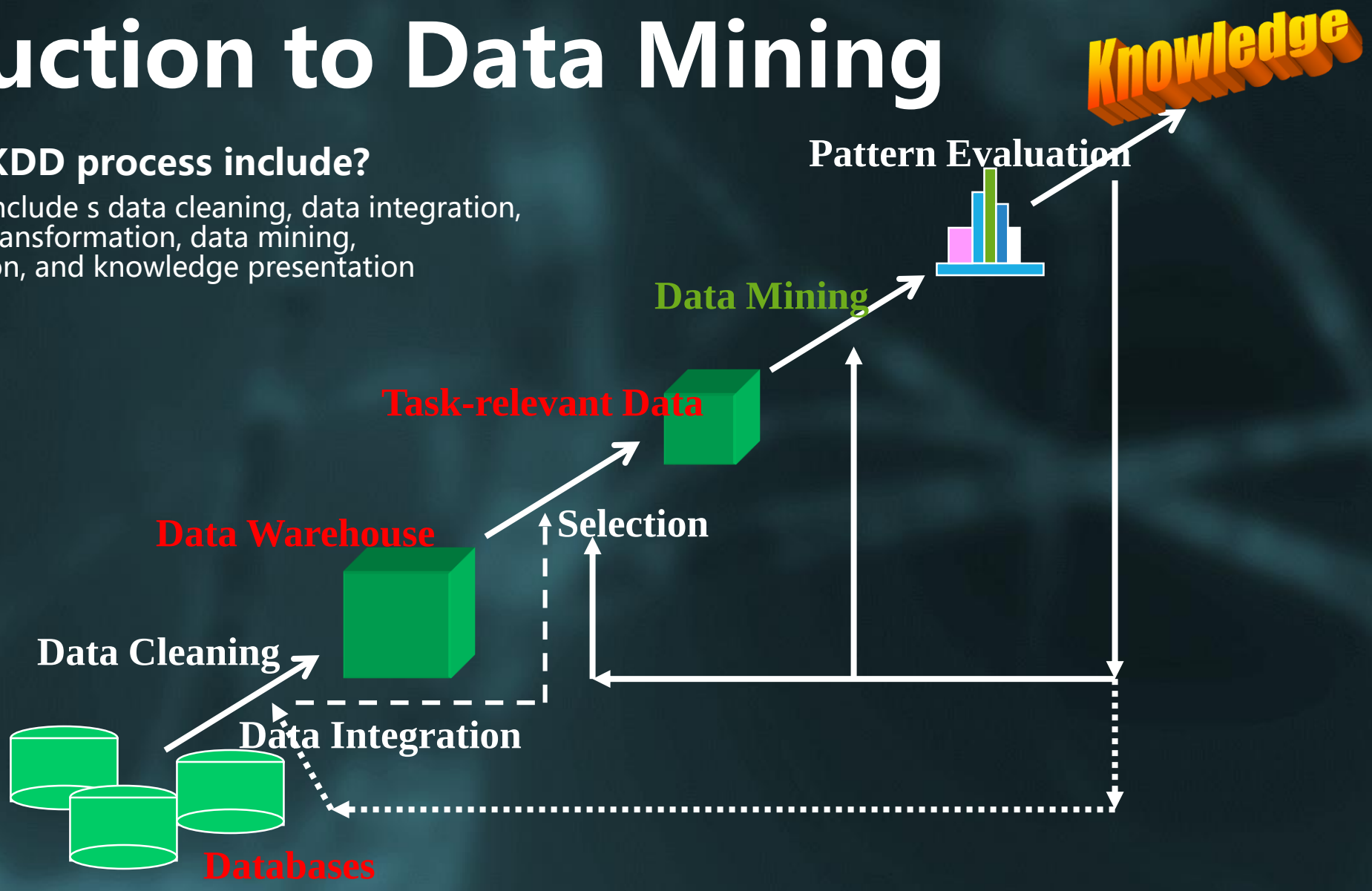
Discovering interesting patterns and knowledge from massive amount of data



Chapter 1
Introduction to Data Mining

2. What does a KDD process include?

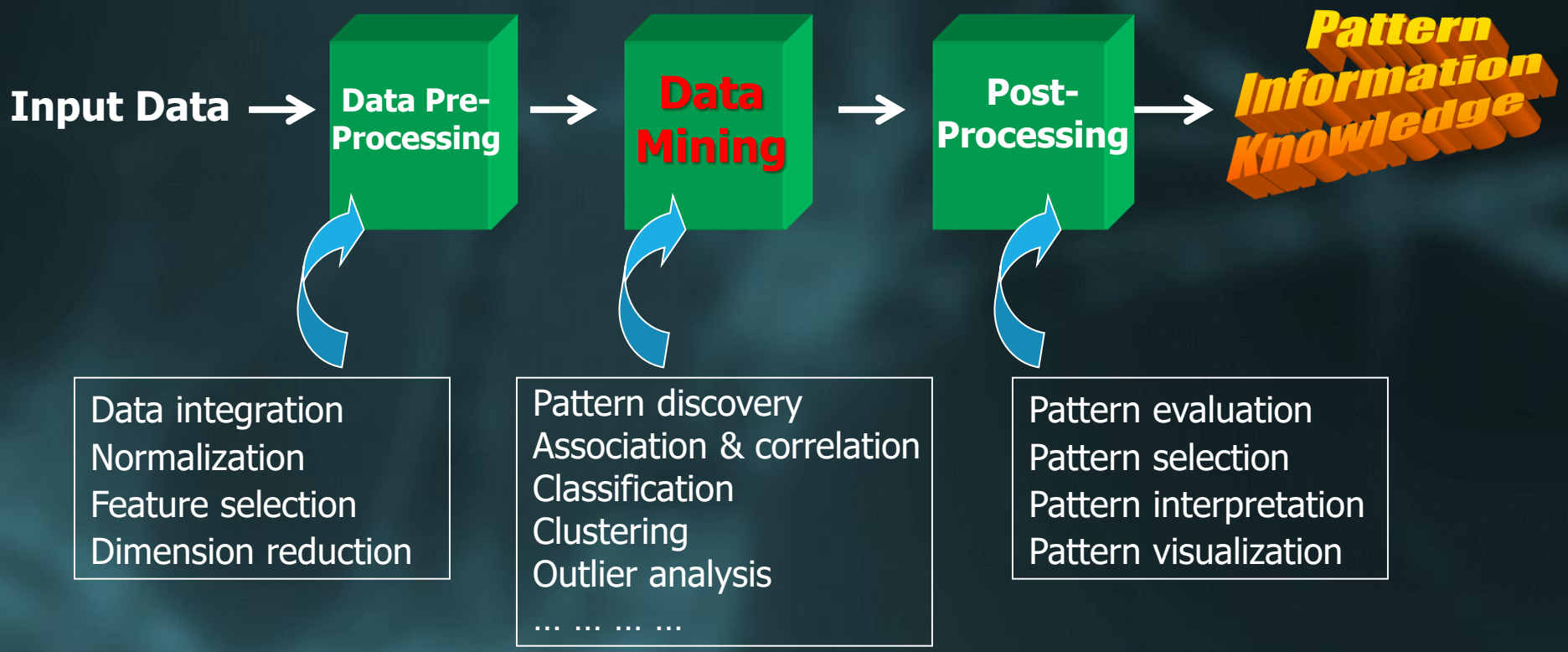
A KDD process include s data cleaning, data integration, data selection, transformation, data mining, pattern evaluation, and knowledge presentation



Chapter 1
Introduction to Data Mining

3. What are the functionalities of data mining?

Data mining functionalities: characterization, discrimination, association, classification, clustering, outlier and trend analysis, etc.



Chapter 1
Introduction to Data Mining

4. Do you know any application scenarios of Data Mining?

7 Best Real-Life Example of Data Mining

https://prowebscraper.com/blog/data-mining-examples/

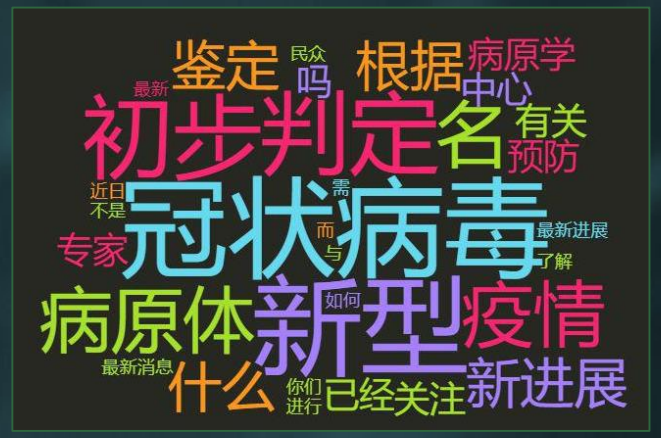
Chapter 1
Introduction to Data Mining

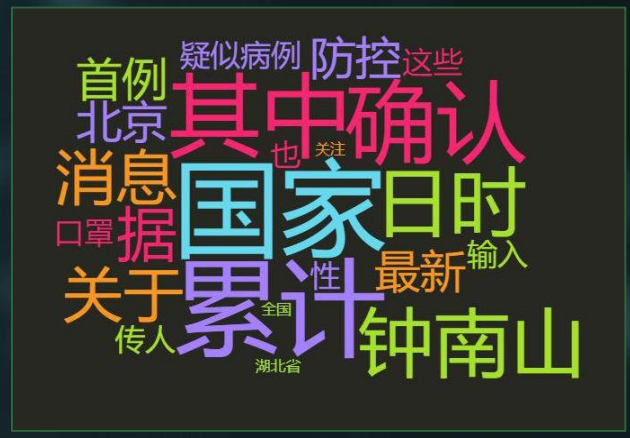
Chapter 1
Introduction to Data Mining

Analysis 1:

The history of Wuhan coronavirus in January 2020

http://www.myzaker.com/article/5e2e9518b15ec033014c366a/



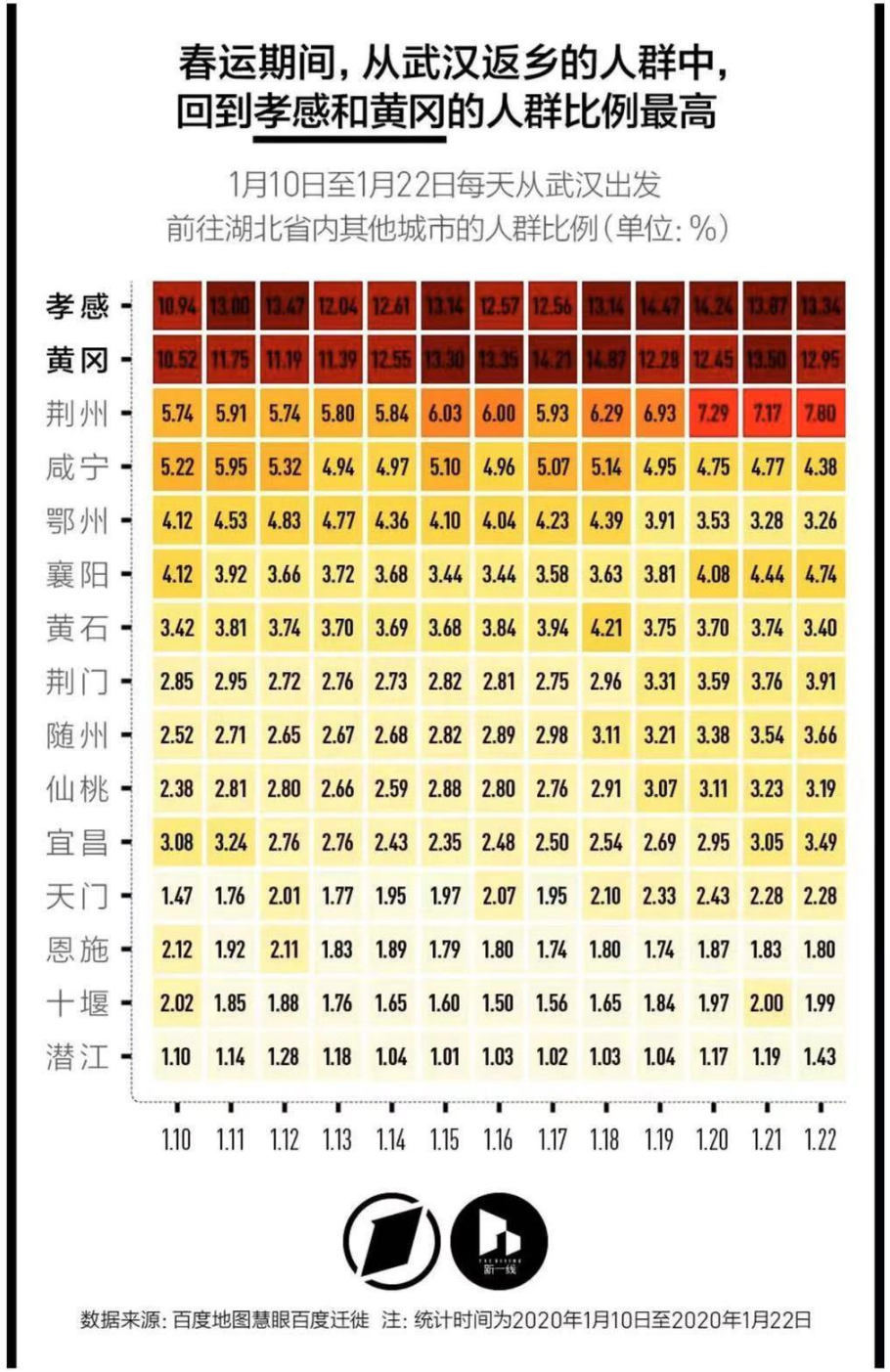


Time

Analysis 2:

Five million people left Wuhan before the lockdown, where did they go?

https://www.yicai.com/news/100481655.html



地区	确诊	死亡	治愈
▼ 湖北	33366	1068	2646
武汉	19558	820	1379
孝感	2751	45	179
黄冈	2398	54	338
随州	1129	12	56
荆州	1110	21	80
襄阳	1088	12	60
黄石	874	6	96
鄂州	861	28	72
宜昌	784	8	62
荆门	696	24	81
十堰	536	1	70
咸宁	525	6	65
仙桃	460	13	43
天门	293	10	12
恩施州	203	3	38
潜江	90	5	8
神农架林区	10		7

2020年2月12日湖北疫情
https://ncov.dxy.cn/ncovh5/view/pneumonia

More people went to Xiaogan and Huanggang from Wuhan, which indicates that more people in these two cities would get virus. Reports on February 12 verified this.

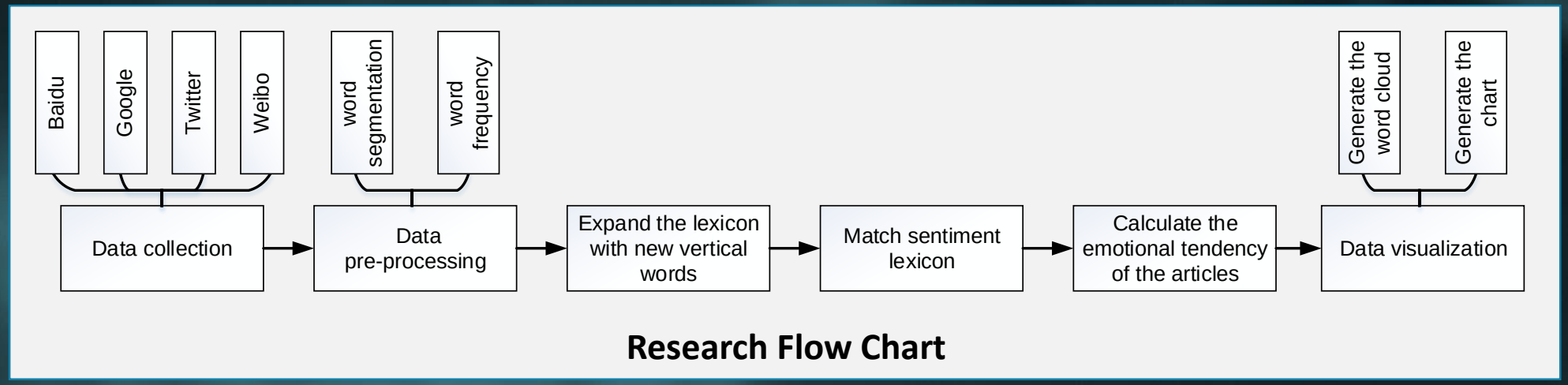
Chapter 1
Introduction to Data Mining

Chapter 1
Introduction to Data Mining

On November 26, 2018, the so-called “anti-AIDS” gene editing baby was born in China, which has aroused wide discussion and controversy internationally.

This study takes gene editing baby event as the main research topic, investigates the emotional tendency of Chinese and Western people towards this issue.

- This research was done by Miss Xinyu LIU and Dr. Ting WANG in 2019.
- Introduction to the Authors:
  I. Miss Xinyu LIU is now an undergraduate student in Shanghai International Studies University (SISU), and will be a postgraduate student in Journalism and Communication at SISU in September 2020.
  II. Dr. Ting WANG is the supervisor of Miss Xinyu LIU.



Research Flow Chart

Chapter 1
Introduction to Data Mining

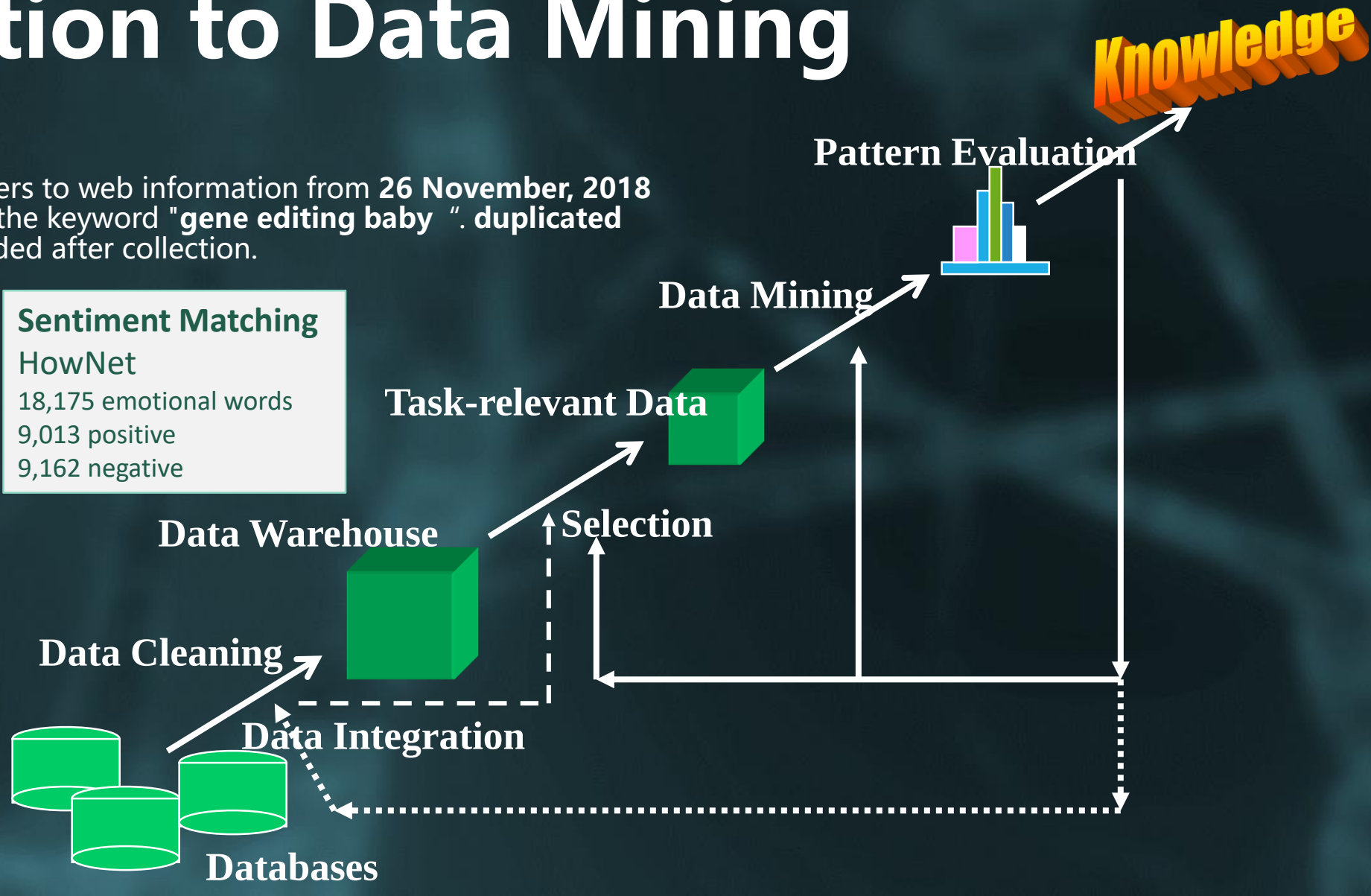
Data Description

the study uses the web crawlers to web information from **26 November, 2018 to 15 December, 2018**, with the keyword "**gene editing baby** ". **duplicated information** has been discarded after collection.

Data Collection
Baidu: 740
Google: 131
Weibo: 500
Twitter: 131

Data Preprocessing
Word quantity
Baidu: 195,710
Google: 44,080
Weibo: 24,755
Twitter: 3,409

Sentiment Matching
HowNet
18,175 emotional words
9,013 positive
9,162 negative



Chapter 1
Introduction to Data Mining

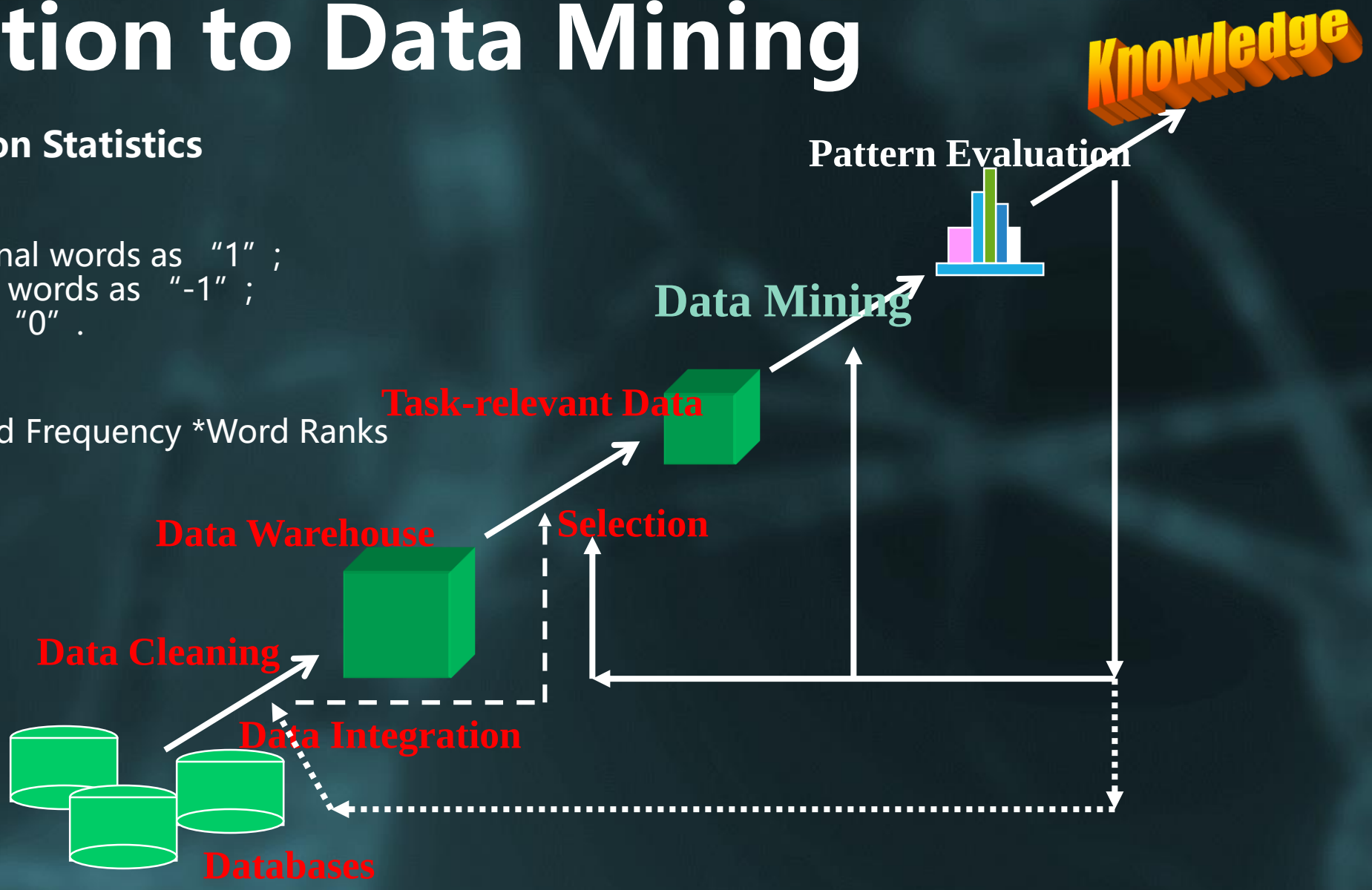
Data Mining: Based on Statistics

Step 1: Ranking

set each positive emotional words as “1”;
each negative emotional words as “-1”;
if there is no match, it is “0”.

Step 2: Calculation

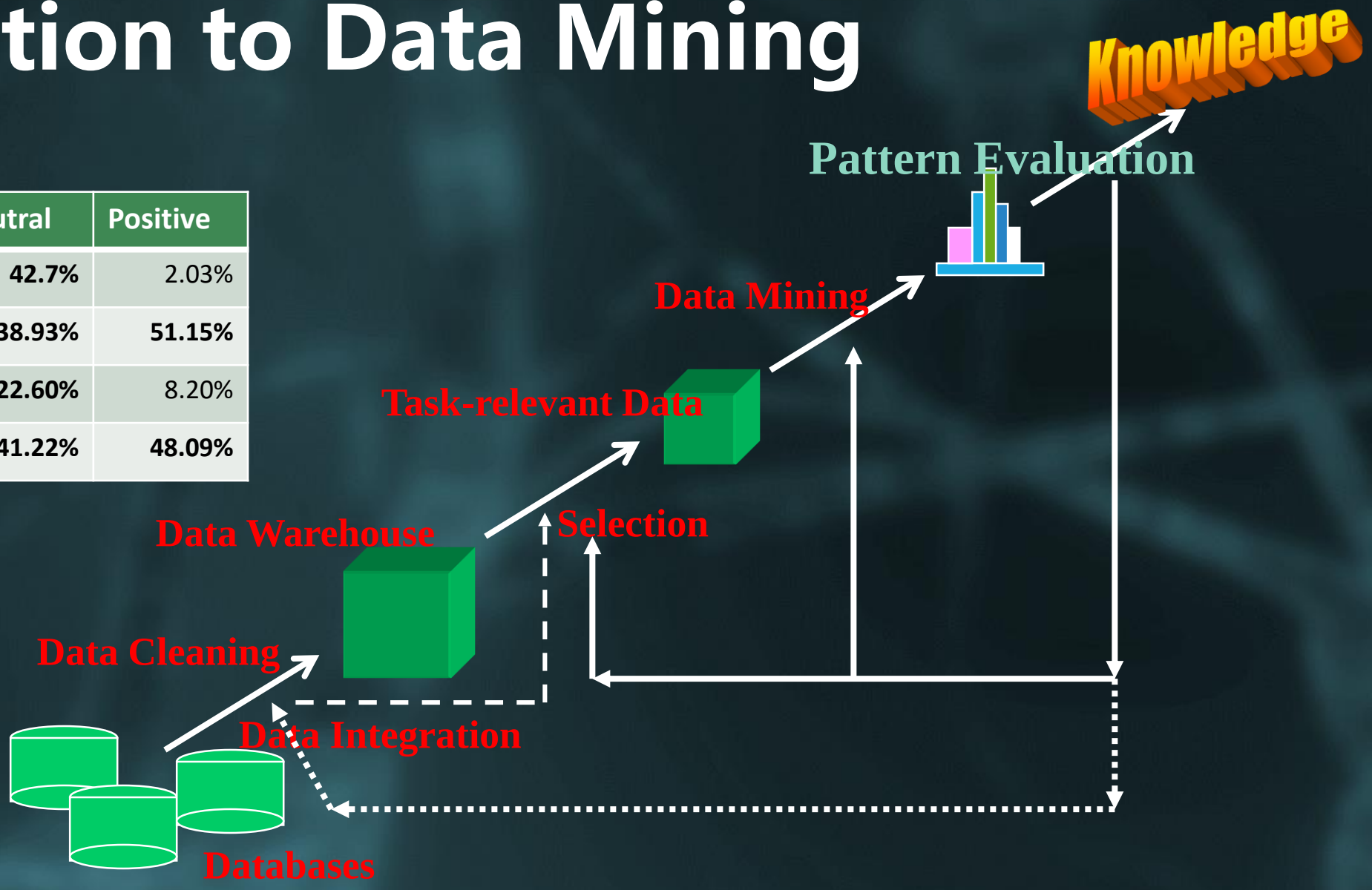
Article Sentiment=∑Word Frequency *Word Ranks



Chapter 1 Introduction to Data Mining

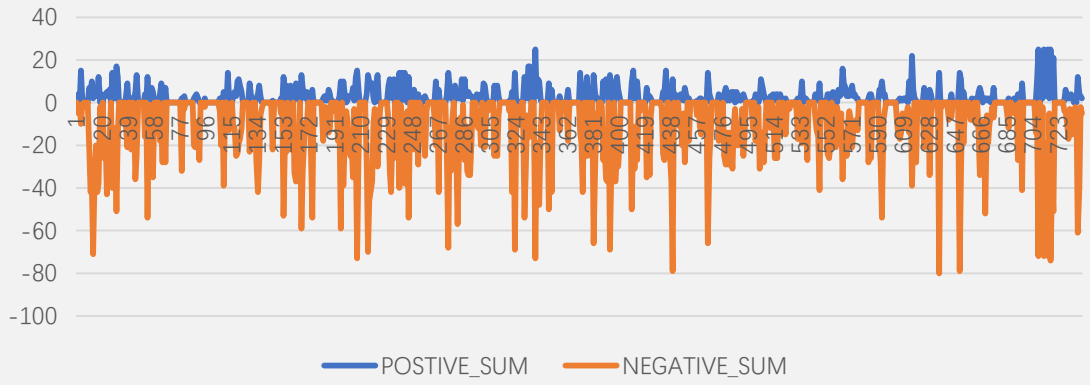
Pattern Evaluation

	Negative	Neutral	Positive
Baidu	**55.14%**	**42.7%**	2.03%
Google	9.92%	**38.93%**	**51.15%**
Weibo	**69.20%**	**22.60%**	8.20%
Twitter	10.69%	**41.22%**	**48.09%**

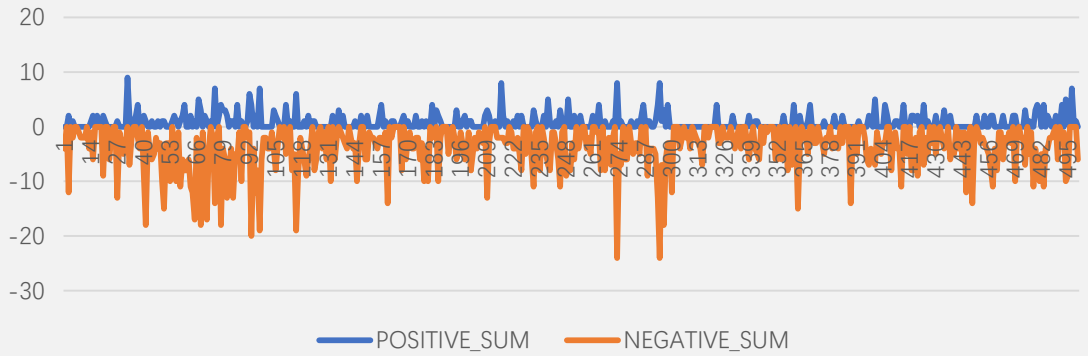


Chapter 1
Introduction to Data Mining

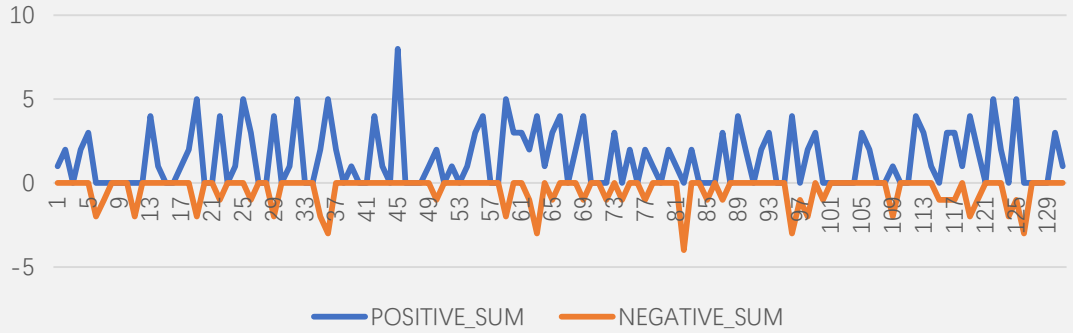
Pattern Evaluation



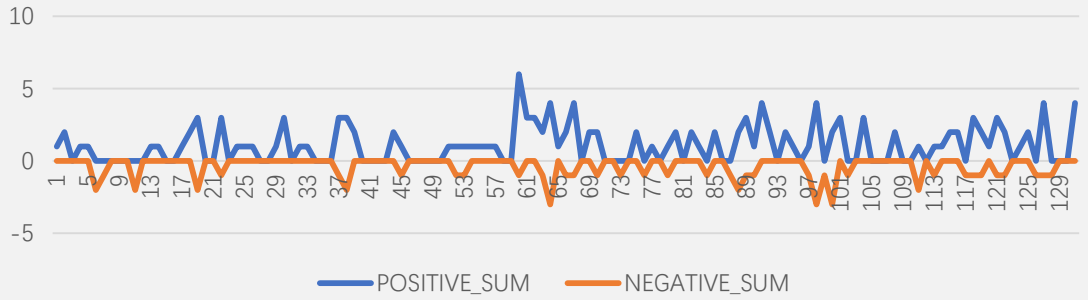
Baidu



Weibo



Google



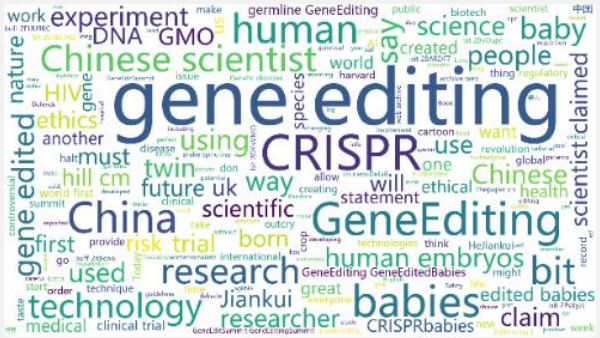
Twitter

Chapter 1
Introduction to Data Mining

Pattern Evaluation

Google

Baidu



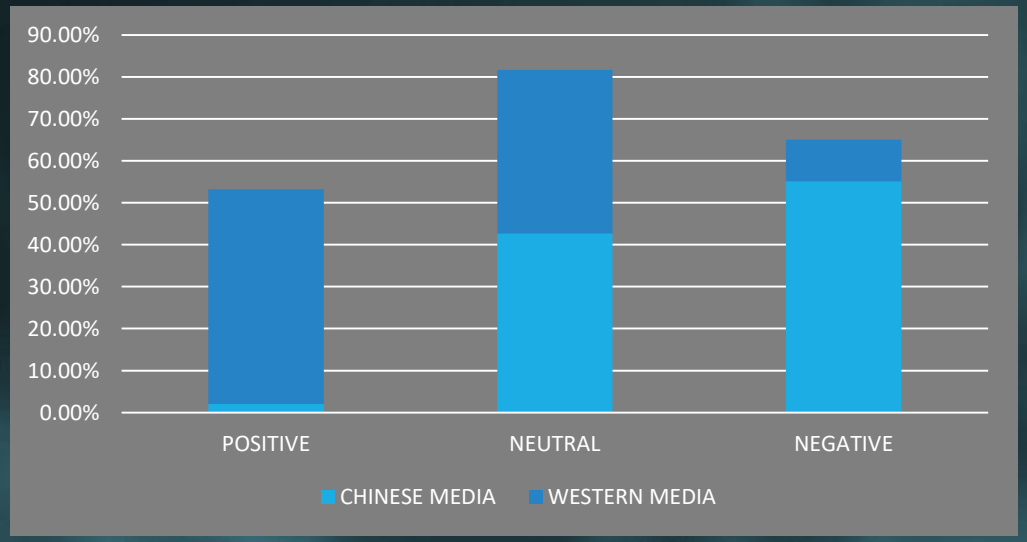
Twitter

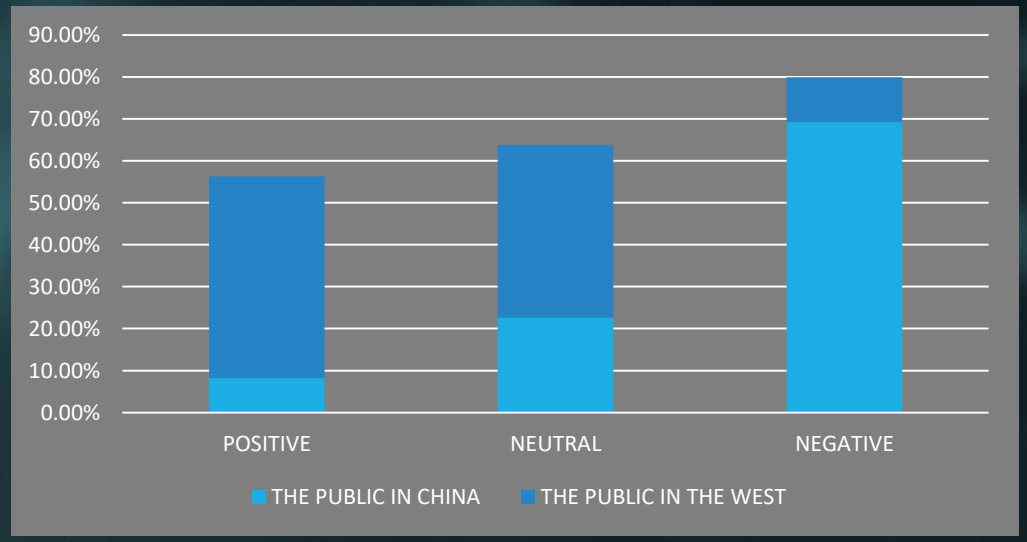
Weibo

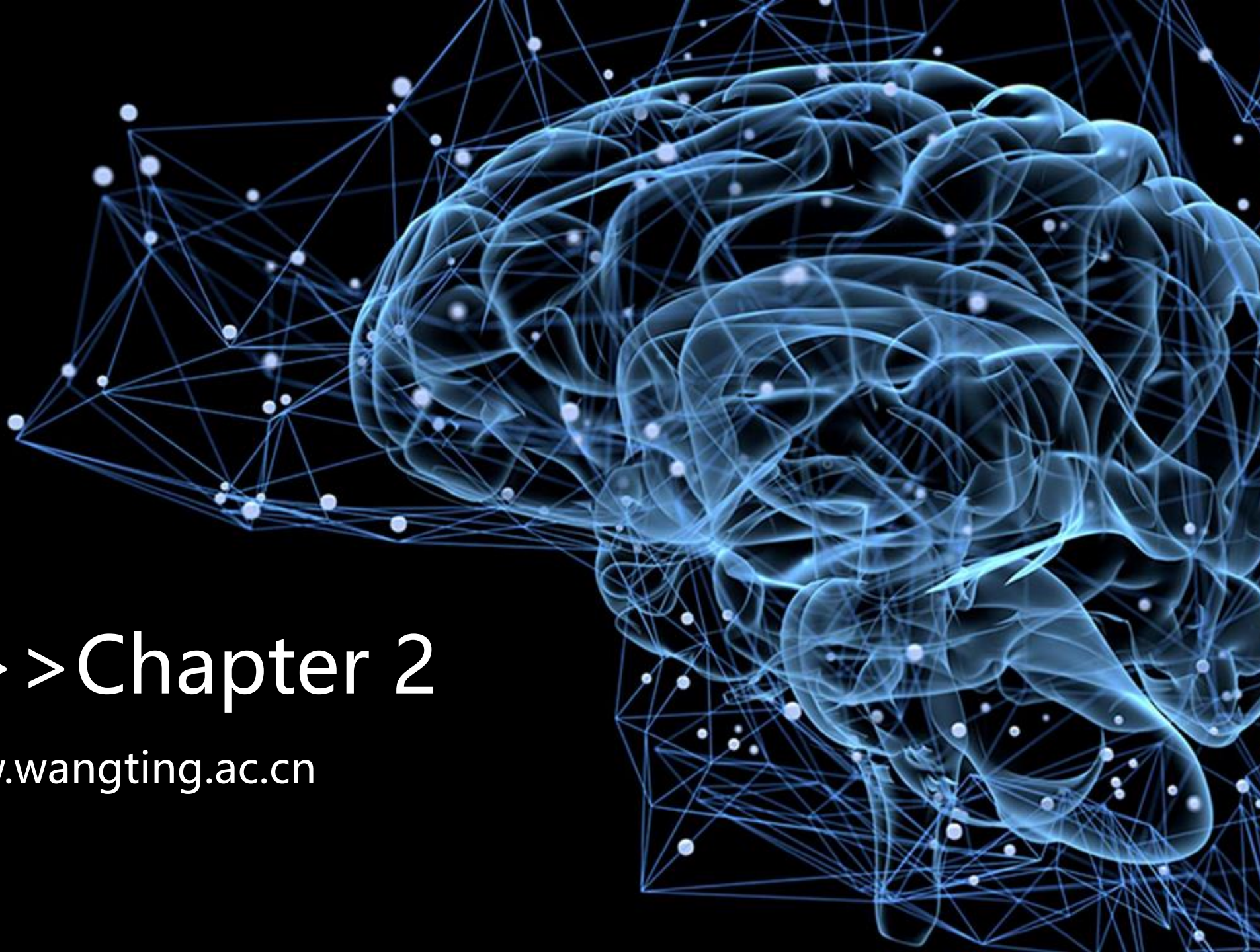
Chapter 1
Introduction to Data Mining

Conclusions:

- Chinese and western people have **great differences** in their emotional attitudes towards gene edited baby event.
- Western media and the public mainly talk about the technology and development related to gene editing.
- Chinese pays more attention to the social impacts of the incident and considers it from the perspective of ethics.







Next>>Chapter 2

www.wangting.ac.cn