

EE226 Big Data Mining

Liyao Xiang (向立瑶)

<http://xiangliyao.cn/>

Shanghai Jiao Tong University

Spring 2019

About Me

- Position
 - Assistant Professor at John Hopcroft Center for CS since 2018
 - IIOT (Intelligent Internet of Things) Lab
 - Research: security, privacy, data mining/machine learning, mobile computing
- Education
 - Ph.D., ECE Dept., University of Toronto, 2014-2018
 - M.A.Sc., ECE Dept., University of Toronto, 2012-2014
 - B.Eng., EE, Shanghai Jiao Tong University, 2008-2012

Course Administration

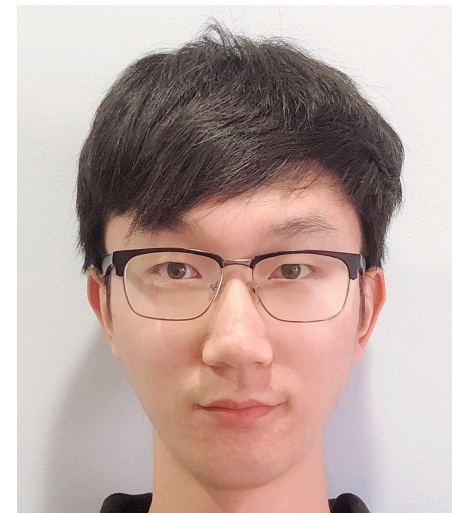
- No official textbook for this course, but the recommended books are
 - Jiawei Han, Micheline Kamber, Jian Pei, “Data Mining: Concepts and Techniques, 3rd Edition,” Morgan Kaufmann Series, 2012.
 - 周志华, “机器学习”, 清华大学出版社, 2016.
 - Avrim Blum, John Hopcroft, Eavindran Kannan, “Foundations of Data Science,” 上海交通大学出版社, 2017.
 - Christopher M. Bishop, “Pattern Recognition and Machine Learning,” Springer, 2011.

Course Administration

- Theory and hands-on experience are both valued.
 - No midterm, no final
 - One course work (30%)
 - Kaggle-in-Class competitions on image classification
 - One assignment (15%)
 - One in-class test (15%)
 - Three in-class quizzes (10%)
 - Poster project (30%)

TA Administration

- Teaching assistant: Hui Xu (徐辉), first-year Ph.D. student. Email: xhui_1@sjtu.edu.cn
- Join the mail list by sending your
 - Name
 - Student number
 - Email addressto Hui Xu xhui_1@sjtu.edu.cn with title “Check in EE226”
- Office hour: every Friday 8-9pm



Goal

- Know about the big picture of data science
- Understand the theoretical concepts in data mining
- Get familiar with fundamental data mining methodologies
- Get hands-on data mining experience
- Know about research frontiers on security and privacy in data mining

Course Landscape

1. Introduction
2. Fundamentals of DM
3. Basic DM Alg.
4. Supervised Learning 1
5. Supervised Learning 2
6. Supervised Learning 3
7. Unsupervised Learning
8. Graphical Prob. Models 1
9. Graphical Prob. Models 2
10. Knowledge Graphs (In-class Test)
11. Learning to Rank
12. Reinforcement Learning
13. Adversarial Attacks
14. Privacy-Preserving DM
15. Course Review
16. Poster Session

Introduction to Data Mining

Liyao Xiang

<http://xiangliyao.cn/>

Shanghai Jiao Tong University

Outline

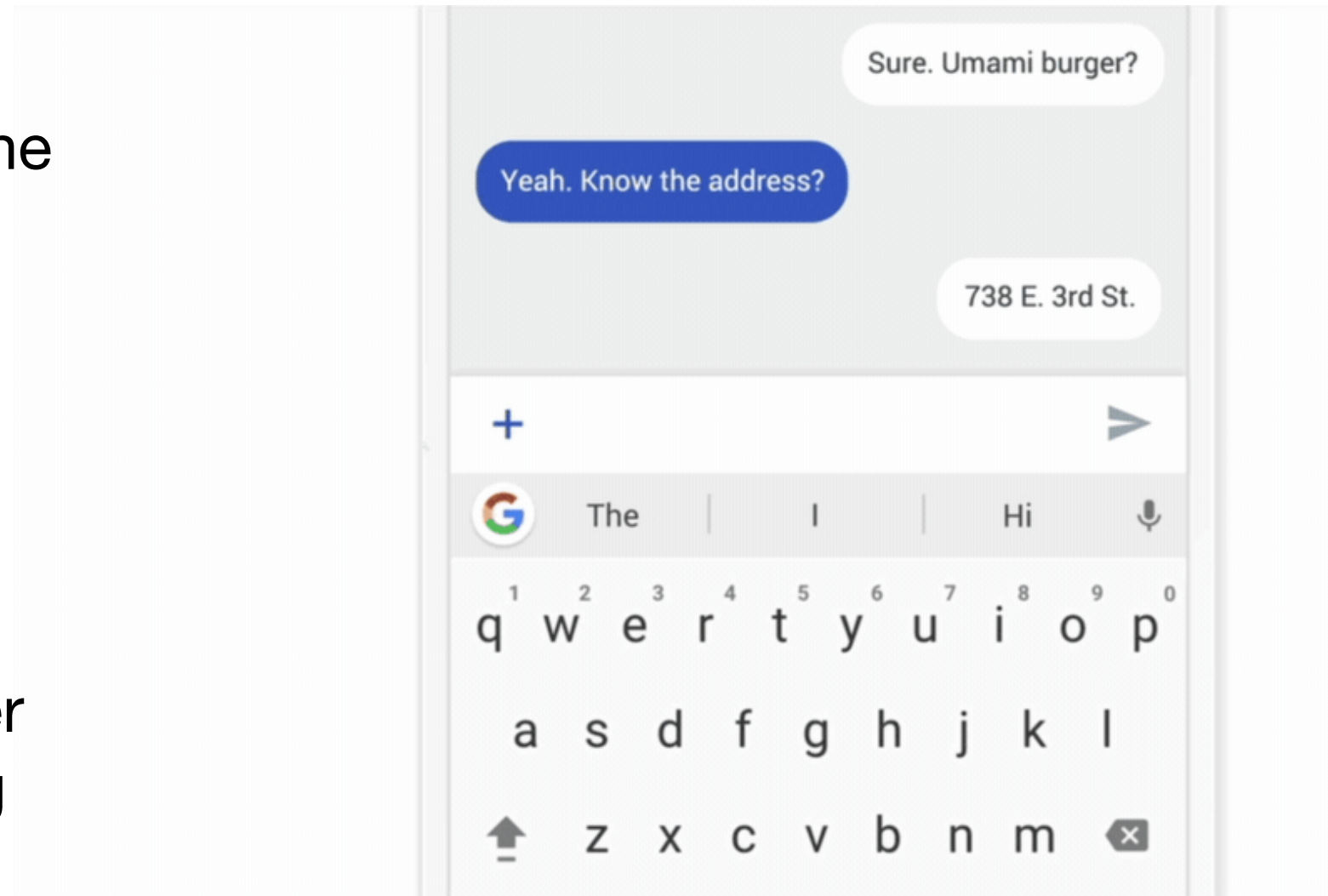
- Why Data Mining?
- What is Data Mining?
- What Kinds of Data Can be Mined?
- What Kinds of Knowledge Can be Mined?
- What are the Technologies?
- What are the Targeted Applications?

Outline

- Why Data Mining?
- What is Data Mining?
- What Kinds of Data Can be Mined?
- What Kinds of Knowledge Can be Mined?
- What are the Technologies?
- What are the Targeted Applications?


Gboard Example

- How does Gboard make the typing suggestion?
 1. Gboard shows a suggested query
 2. I clicked
 3. Next time, the answer shows up as a typing suggestion



Ads Display Example

- Weinan frequently visits emarketer.com

[Research Topics](#)[Products](#)[Why eMarketer](#)[Customer Stories](#)[Articles](#)

Advertisers Continue Rapid Adoption of Programmatic Buying

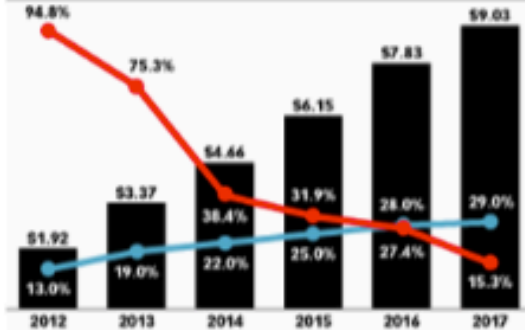
By 2017, advertisers will spend more than \$9 billion on RTB

Nov 26, 2013

[Share](#)[Print](#)[Email](#)

Advertisers are spending more than expected on real-time bidding, which is expected to account for a significant share of all display ad spending in the US

billions, % change and % of total digital display ad spending



Year	RTB digital display ad spending (billions)	% change	% of total digital display ad spending
2012	\$1.92	94.8%	13.0%
2013	\$3.37	75.3%	19.0%
2014	\$4.66	38.4%	22.0%
2015	\$6.15	31.9%	25.0%
2016	\$7.83	28.0%	27.4%
2017	\$9.03	15.3%	29.0%

■ RTB digital display ad spending
■ % change ■ % of total digital display ad spending

Note: includes all display formats served to all devices
Source: eMarketer, Dec 2013
164097 www.emarketer.com

advertising—which includes RTB—continues its rapid transition from infancy to a well-established display purchase method in just a few years.

eMarketer projects RTB digital display ad spending in the US will account for 29.0% of total US digital display ad spending by 2017, or \$9.03 billion. In 2013, it will account for 19.0%, or \$3.37 billion. These estimates are revised slightly upward from our previous forecast in August

Latest from eMarketer


[Latest Articles](#)[Latest Webinars](#)

[Hispanic Gen Xers Lead in Daily Tablet Usage](#)

[Chrysler's Multichannel Approach to Online Video Gets Greater Recall](#)

[Android Rules UK Smartphone Sales](#)

[More Articles »](#)[eMarketer Daily Newsletter »](#)



WATCH THE VIDEO.
DO WHAT CAN NOW BE DONE. ©

[@ Contact Sign-Up](#)[@ Contact Sign-Up](#)[@ Contact Sign-Up](#)[@ Contact Sales](#)

Ads Display Example

- Weinan booked a hotel on booking.com

Booking.com

Browse by destination theme [Shopping](#) [Fine Dining](#) [Culture](#) [Sightseeing](#) [Monuments](#) [Relaxation](#)

home → [uk](#) 16,378 properties → [greater london](#) 1,824 properties → [london](#) 1,574 properties → search results London, 2 adults, 11 nights (Jul 14 - Jul 25) [Change dates](#)

Search

Destination/Hotel Name:

Distance:

Check-in Date:

Check-out Date:

☐ I don't have specific dates yet

Guests

Search
Search properties

Filter by:

48% reserved

London is a top choice with fellow travelers on your selected dates (48% reserved).
Tip: Prices might be higher than normal, so try searching with different dates if possible.

[Try previous week](#) Jul 7 - Jul 18 [Try next week](#) Jul 21 - Aug 1

930 out of 1857 properties are available in and around London
Showing 1 – 15

Sort by: **Recommended** Stars Location Price Review Score [List](#) [Map](#)

Park Plaza Victoria London ★★★★★ [1736](#) **Very good 8.5**
Score from 1137 reviews

Central London, Westminster, London • [Nearby stop](#)

There are 13 people looking at this hotel.
Latest booking: 1 hour ago

Superior Double Room **£2,353.65**
7 more room types > **Book now**

Central Park Hotel ★★★ [1993](#) **6.6**

Ads Display Example

- Today, he found an ad on his facebook page.

The image is a screenshot of a Facebook profile page for a user named 'Weinan'. The page layout includes a left sidebar with navigation links like 'Family', 'UCL', 'SJTU', and 'Groups'. The main content area features a sponsored advertisement for 'Secret Escapes', which promotes exclusive discounts on luxury hotels and holidays. The ad includes a large image of a hotel interior and a 'Sign Up' button. To the right of the main content, there are friend suggestions for 'Bingkai Lin' and 'Zhaomeng Peng', and two more sponsored ads: '247 London Hostel' and 'Stale Marketing Stinks'. The bottom of the page shows a footer with language and privacy settings.

Search for people, places and things

Weinan Home

Family

- UCL
- SJTU 16
- UCL 20+
- Shanghai Jiao Ton... 16
- London, United Ki... 20+
- University College... 20+
- Close Friends
- Intern,Beijing,Microso...

GROUPS

- Microsoft Research C...
- Create group

INTERESTS

- Pages and Public Fig...

PAGES

- Like Pages 1
- Pages feed 9
- Create a Page...

DEVELOPER

Secret Escapes
Sponsored · ✱

Find the best rates on handpicked hotels

Secret Escapes | Exclusive Discounts
Get up to 70% off luxury hotels and holidays.
WWW.SECRETESCAPES.COM

Like - Comment - Share · 2,327 85 444

Bingkai Lin
43 mutual friends
Add Friend

Zhaomeng Peng
10 mutual friends
Add Friend

SPONSORED

247 London Hostel
booking.com
Book & Save! 247 London Hostel, London.

Stale Marketing Stinks
emarketer.com
Freshen up with eMarketer's reports, trends & data on digital marketing. Download Today!

English (UK) · Privacy · Terms · Cookies · More

Slide credit: Weinan Zhang

Ads Display Example

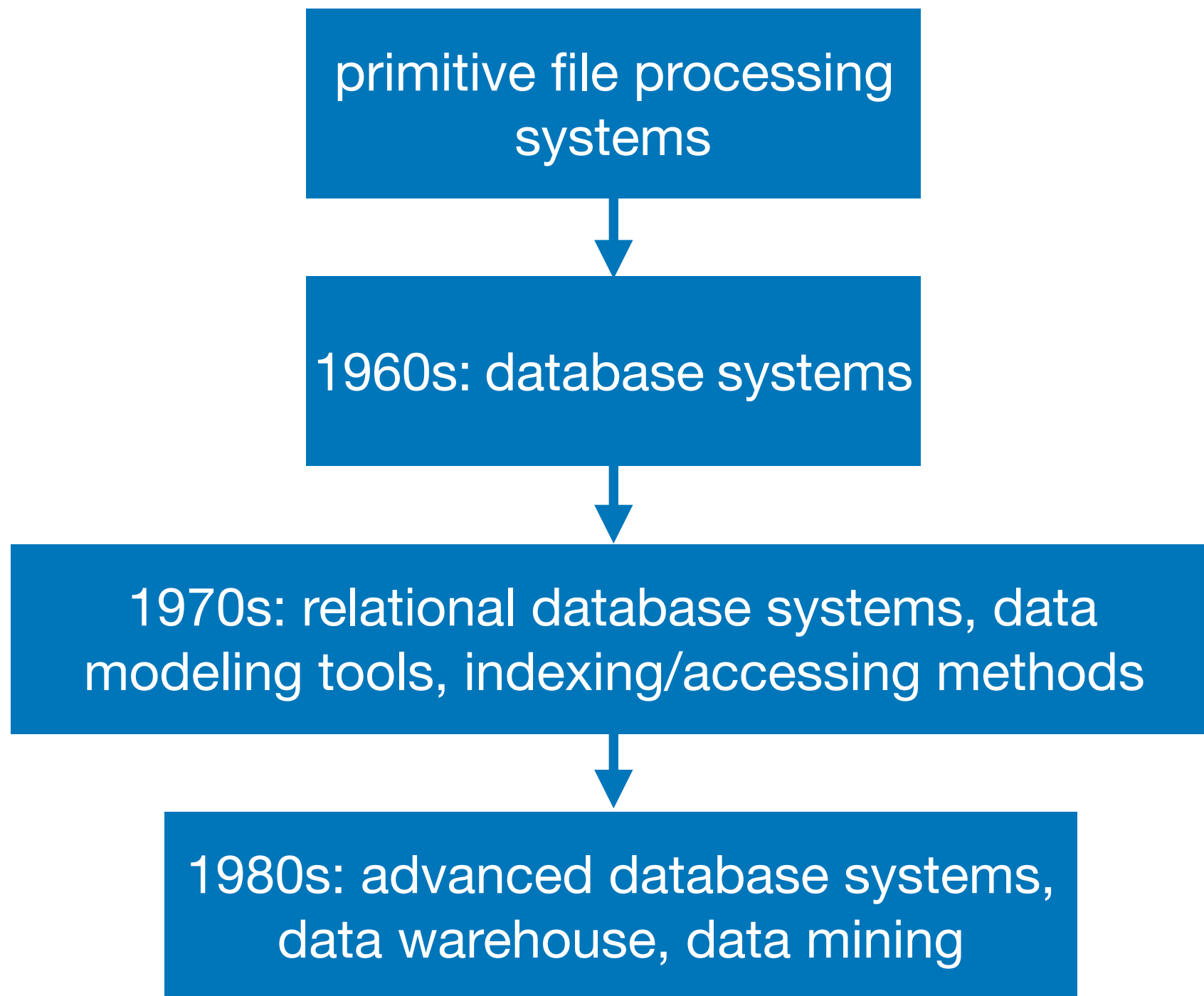
- Today, he found an ad on his facebook page.
- Why do the ads show to Weinan?
- How likely will he click on the ad?

Outline

- Why Data Mining?
- **What is Data Mining?**
- What Kinds of Data Can be Mined?
- What Kinds of Knowledge Can be Mined?
- What are the Technologies?
- What are the Targeted Applications?

Data Mining

- Definition: Knowledge Discovery from Data



Data Mining

- Definition: Knowledge Discovery from Data
- Iterative process includes:

1. Data cleaning

2. Data integration

3. Data selection

4. Data transformation

5. Data mining

6. Pattern evaluation

7. Knowledge presentation



data preprocessing



**may interact with user
or a knowledge base**

Data Mining

- Definition: Knowledge Discovery from Data
- Iterative process concludes:

1. Data cleaning
2. Data integration
3. Data selection
4. Data transformation
5. Data mining
6. Pattern evaluation
7. Knowledge presentation

Interesting if:

- 1. easily understood**
- 2. valid on new dataset**
- 3. potentially useful**
- 4. novel**

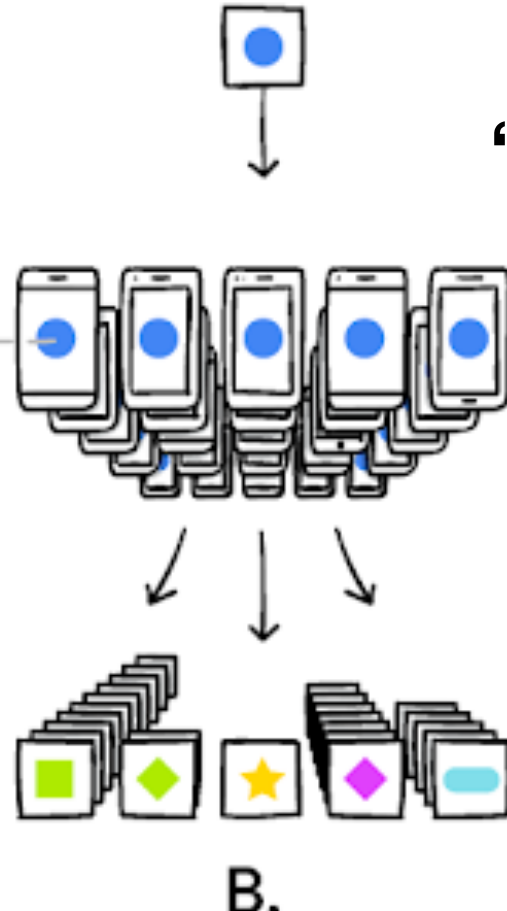
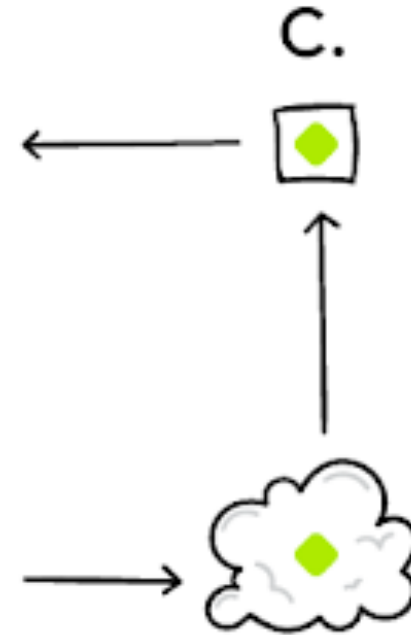
visualization etc.

Gboard Example

local data preprocessing
“click, context”



pattern evaluation & knowledge presentation
“show which suggestion”



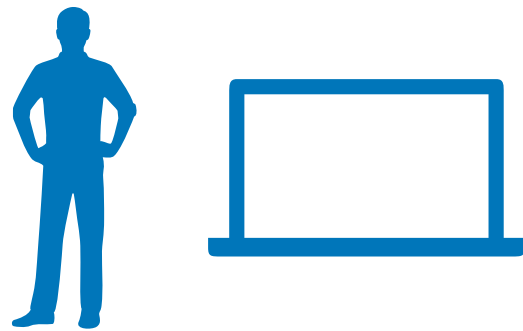
local model update

**data mining: interaction
between users and cloud**

Ads Display Example

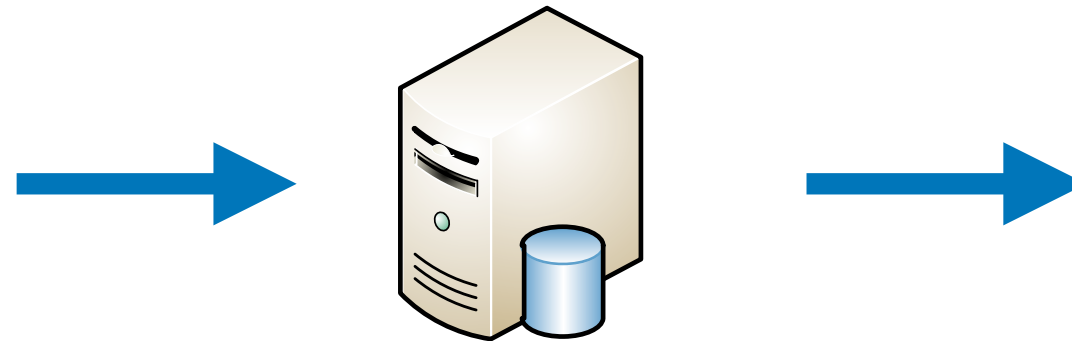
- Ads Display

**“click or not”
— raw data**



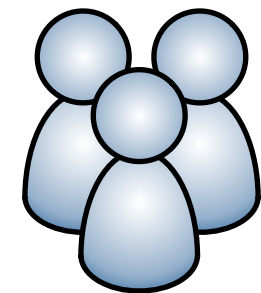
Users

**“20-40, male, travel”
— attributes**



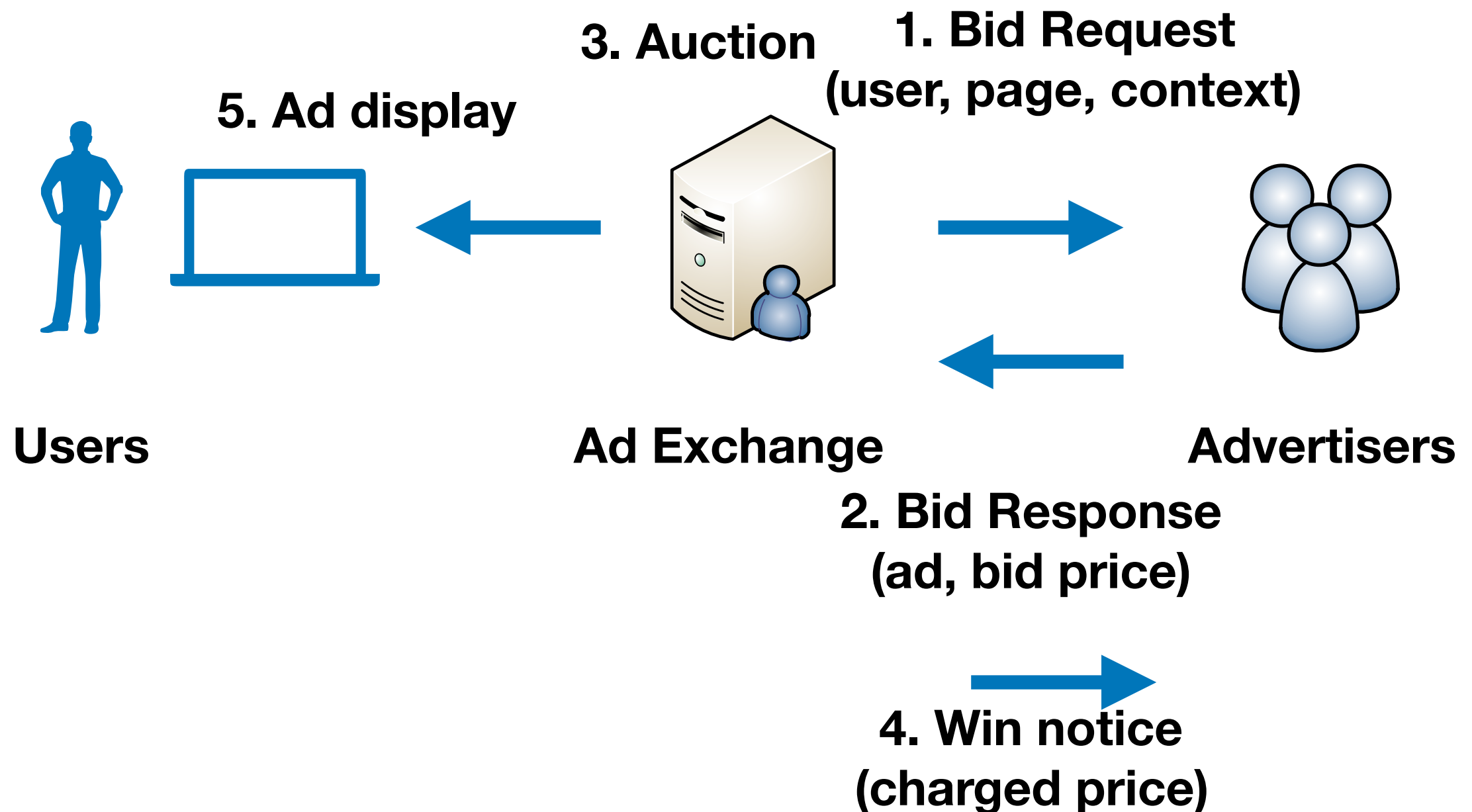
**Data Management Platform:
data preprocessing,
data mining**

**user information
matching**



**Advertiser:
targets a segment of
users**

Ads Display Example



Outline

- Why Data Mining?
- What is Data Mining?
- **What Kinds of Data Can be Mined?**
- What Kinds of Knowledge Can be Mined?
- What are the Technologies?
- What are the Targeted Applications?

Data Source

- Database
 - E.g. A relational database
 - a collection of **tables**, each consisting of a set of **attributes** (columns) and a large set of **tuples** (rows, **key** + **attribute values**)

cust_ID	name	address	age	occupation	income
1	Alice	21 Baker St.	30	Doctor	50k
2	Bob	40 St. George St.	22	Student	10k
3					

Data Source

- Database
 - E.g. A relational database
 - **relational queries**: “Show me the number of customers between the age of 20 to 30”
 - aggregate functions e.g.: sum, avg, count, max and min

cust_ID	name	address	age	occupation	income
1	Alice	21 Baker St.	30	Doctor	50k
2	Bob	40 St. George St.	22	Student	10k
3					

Data Source

- Database
 - E.g. A relational database
 - **relational queries**: “Show me the number of customers between the age of 20 to 30”
 - **aggregate** functions e.g.: sum, avg, count, max and min
 - Mining: predict credit risk of new customers

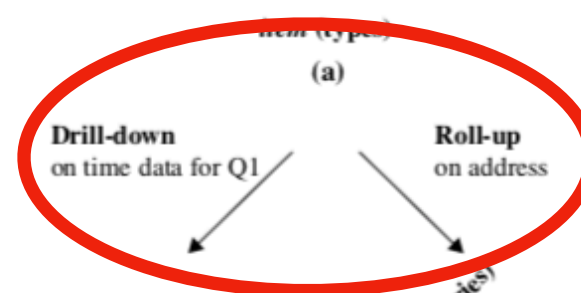
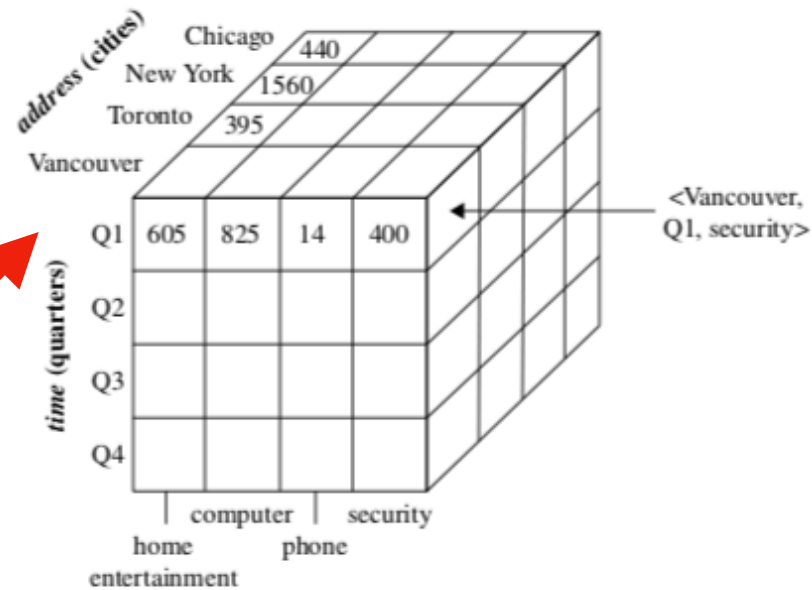
Data Source

- Data Warehouses
 - A repo of information collected from **multiple sources**, stored under a **unified schema**, residing at **a single site**
 - **data cube**: a multidimensional data structure
 - each **dimension** is an attribute or a set of attributes
 - each **cell** stores aggregate measure
 - operations include **drill-down**, **roll-up**
 - **Multidimensional mining**: explore multiple combinations of dimensions at varying levels of granularity

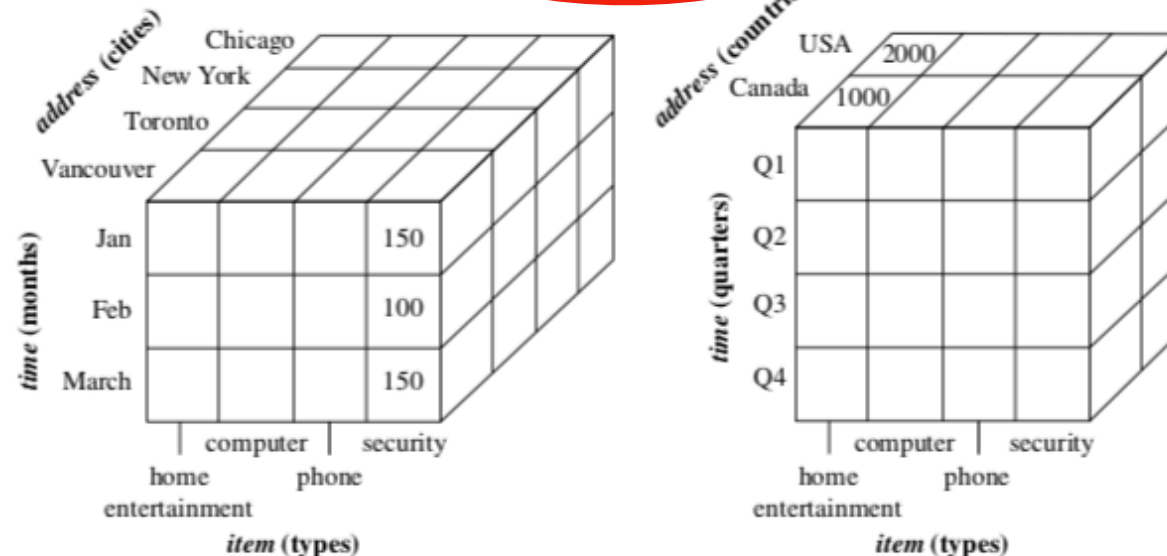
Data Source

3 dimensions:
address, time, item

aggregate value:
sales_amount



**differing degrees
of summarization**



Data Source

- Transactional data
 - transaction: **trans_ID** + a list of **items**
 - Mining **frequent itemsets**
- Sequence data, data streams, spatial data, hypertext and multimedia data, graph and networked data...

Question

- What is the difference between a *data warehouse* and a *database*?

Answer

- A data warehouse: information collected from multiple sources, over **a period of time**, stored under **a unified schema**, used for data analysis and decision support; whereas a database is a collection of interrelated data that represents the **current status** of the stored data. Could be multiple heterogeneous databases with **different schemas**.

Summary

- Data to be mined:
 - relational database
 - data warehouse
 - transactional data
 - sequential data, spatial data, data stream, multimedia data, graph data, networked data, ...

Outline

- Why Data Mining?
- What is Data Mining?
- What Kinds of Data Can be Mined?
- **What Kinds of Knowledge Can be Mined?**
- What are the Technologies?
- What are the Targeted Applications?

Knowledge

- **Characterization:** summarization of general characteristics of a class of data
 - e.g., summarize the characteristics of customers who spend over \$2000 a year on Apple products.
 - methods: statistical measures and plots, data cube roll-up, ...
 - outputs: pie charts, bar charts, curves, data cube, ...
- **Discrimination:**
 - e.g., compare the general features of books of which sales amount exceeds 1 million with those whose sales do not pass 5k
 - methods and outputs: same with characterization

Knowledge

- Association and Correlation

- e.g., **frequent itemset**, a set of items that frequently appear together in a transactional dataset
- lead to associations

$\text{buys (X, "computer")} \Rightarrow \text{buys (X, "software")}$

[support = 1%, confidence = 50%]

confidence: if one buys a computer, 50% chance it will buy software

support: computer and software are together in 1% of transactions

Knowledge

- Prediction
 - **classification** method: predict the class of (categorical, discrete) objects whose class label is unknown

IF-THEN rules

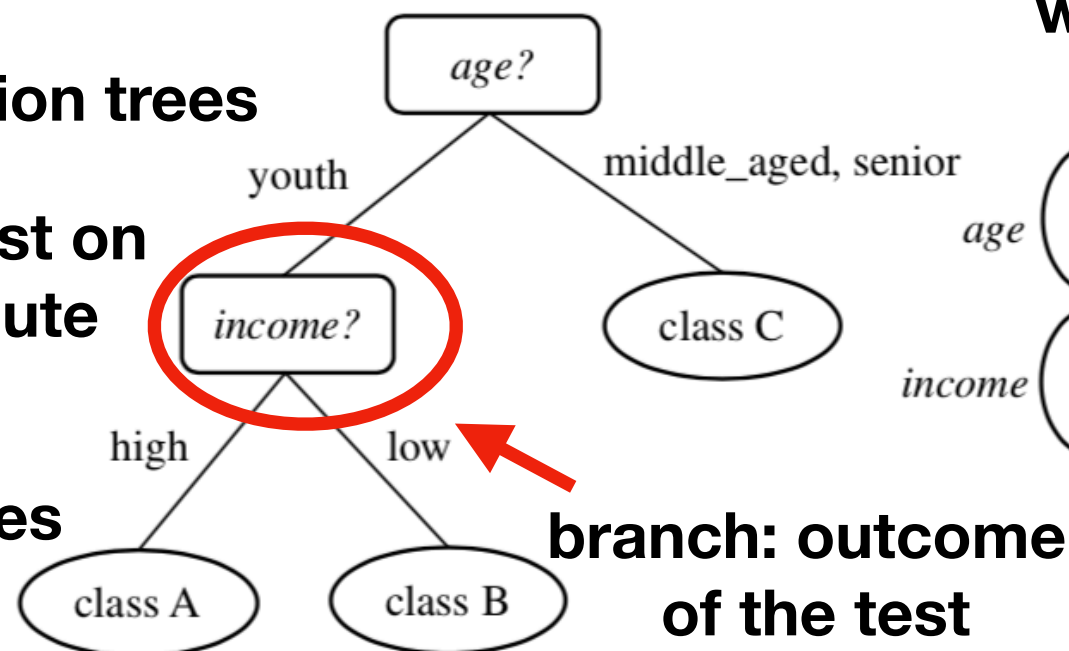
$age(X, \text{"youth"}) \text{ AND } income(X, \text{"high"}) \longrightarrow class(X, \text{"A"})$
 $age(X, \text{"youth"}) \text{ AND } income(X, \text{"low"}) \longrightarrow class(X, \text{"B"})$
 $age(X, \text{"middle_aged"}) \longrightarrow class(X, \text{"C"})$
 $age(X, \text{"senior"}) \longrightarrow class(X, \text{"C"})$

(a)

decision trees

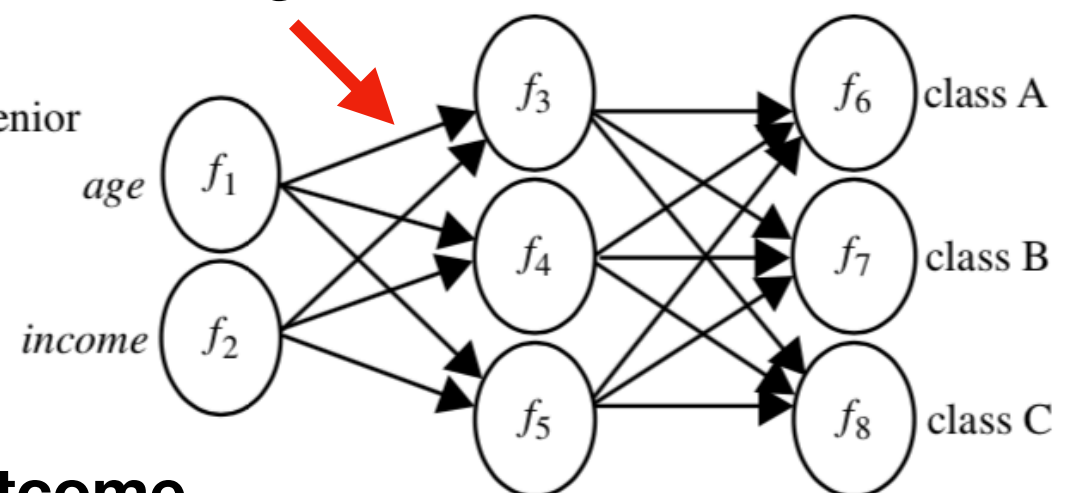
node: a test on
an attribute

leaves: classes



(b)

weights neural networks



(c)

Knowledge

- Prediction
 - **classification** method: predict the class of (categorical, discrete) objects whose class label is unknown
 - **regression** method: predict missing or unavailable numerical data values
 - e.g., predict the amount of revenue that each item generates

Knowledge

- Clustering

- group data without class label
- e.g., identify homogeneous subpopulations of customers

- Outlier

- e.g., uncover unusual usage of credit cards



Question

- What is the difference between discrimination and classification?
Between characterization and clustering?

Answer

- **Discrimination** is a **comparison** of features of target class data objects with features of objects from contrasting classes. **Classification** is the process of **finding models** that describe or distinguish data classes for the purpose of **predicting** objects with unknown class.
- **Characterization** is a **summarization** of features of a target class of data. **Clustering** is the **analysis** of data objects without knowing labels.

Summary

- Knowledge to be mined:
 - Characterization
 - Discrimination
 - Association and Correlation
 - Prediction
 - Clustering
 - Outlier
 - ...

Outline

- Why Data Mining?
- What is Data Mining?
- What Kinds of Data Can be Mined?
- What Kinds of Knowledge Can be Mined?
- **What are the Technologies?**
- What are the Targeted Applications?

Technologies

- **Statistical model**
 - model data, data class, noise, missing data values, ...
 - summarize or describe a collection of data
 - e.g., mean, median, mode, proximity measures
 - verify data mining results by **hypothesis test**
 - e.g. χ^2 -correlation test for nominal data. Suppose attribute A has c values: a_1, a_2, \dots, a_c . B has r values: b_1, b_2, \dots, b_r . Let o_{ij} be the actual count of joint event $(A = a_i, B = b_j)$ and e_{ij} be the expected frequency: $e_{ij} = \text{count}(A = a_i) \times \text{count}(B = b_j) / \text{total_num_data}$.
 χ^2 value is computed as:
$$\chi^2 = \sum_{i=1}^c \sum_{j=1}^r \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

Hypothesis: A and B are indep.

Technologies

- **Machine learning:** Computer programs automatically learn to recognize complex patterns and make intelligent decisions based on data
 - Methods: Supervised learning, unsupervised learning, semi-supervised learning ...
 - Difference between data mining and machine learning:
 - Data mining is the process to discover various types of **pattern** that are inherited in the data and which are accurate, new and useful.
 - Machine learning is the study of **algorithms** that improve automatically through experience based on data.

Technologies

- Machine Learning

- Difference between data mining and machine learning:

Data Mining

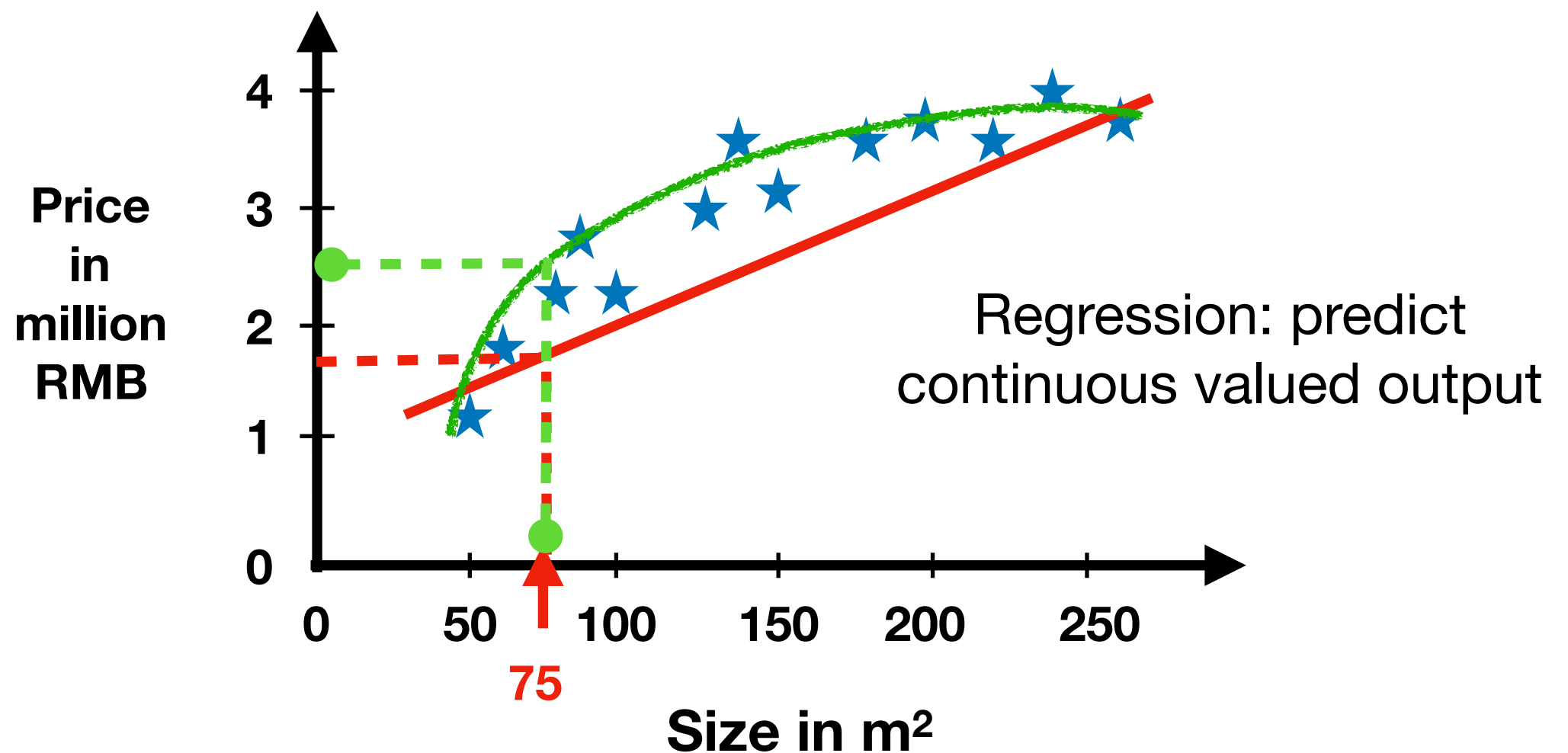
- Extracting **knowledge** from a large amount of data
- To get **rules** from the existing data
- Involve **more** manual effort
- Can use methods include machine learning

Machine Learning

- Introducing new **algorithms** from data and experience
- To teach computers to **learn and understand** the given rules
- Once design self-implemented, **no** human effort
- Can be used in areas outside data mining

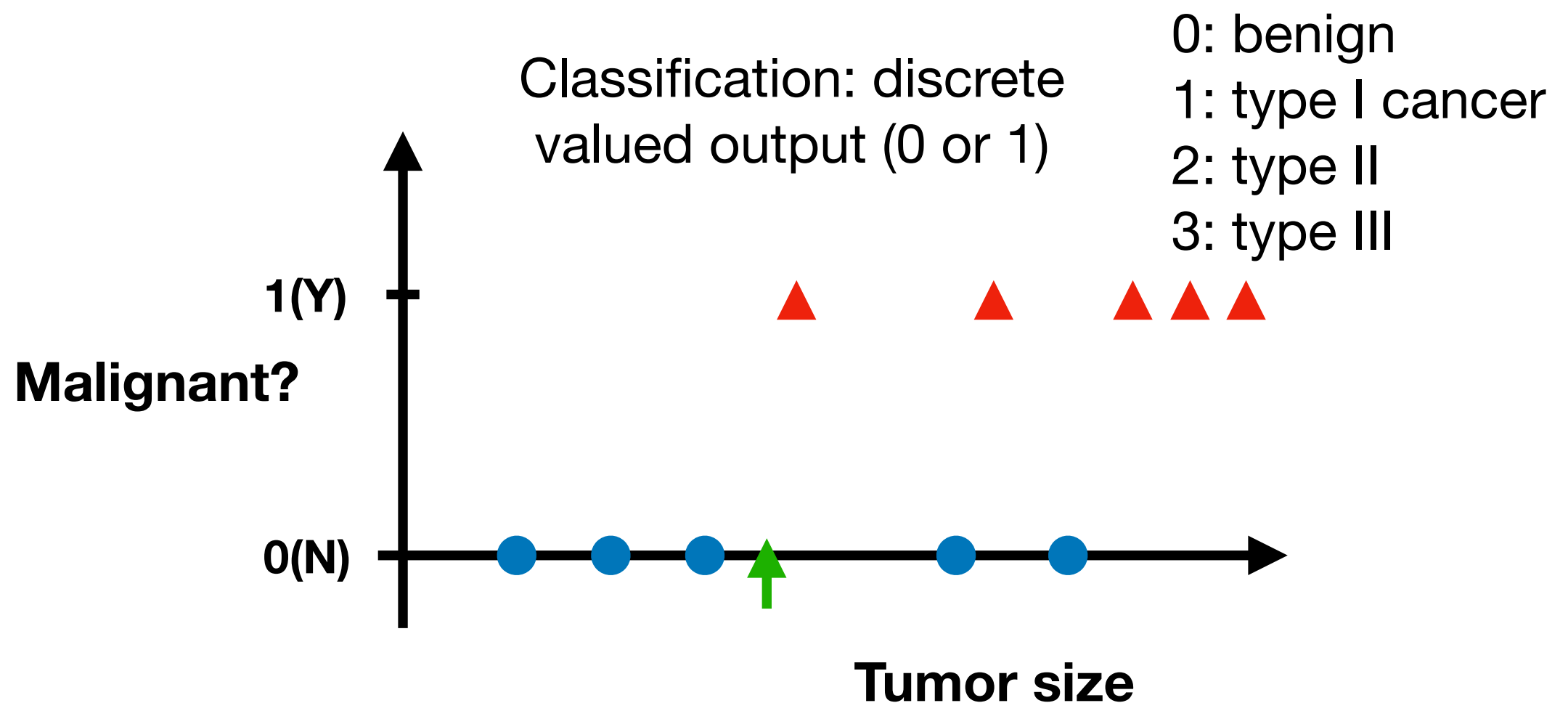
Technologies

- Machine Learning
 - Supervised Learning: We are given the algorithm and a dataset, in which the “right answer” were given.



Technologies

- Machine Learning
 - Supervised Learning:

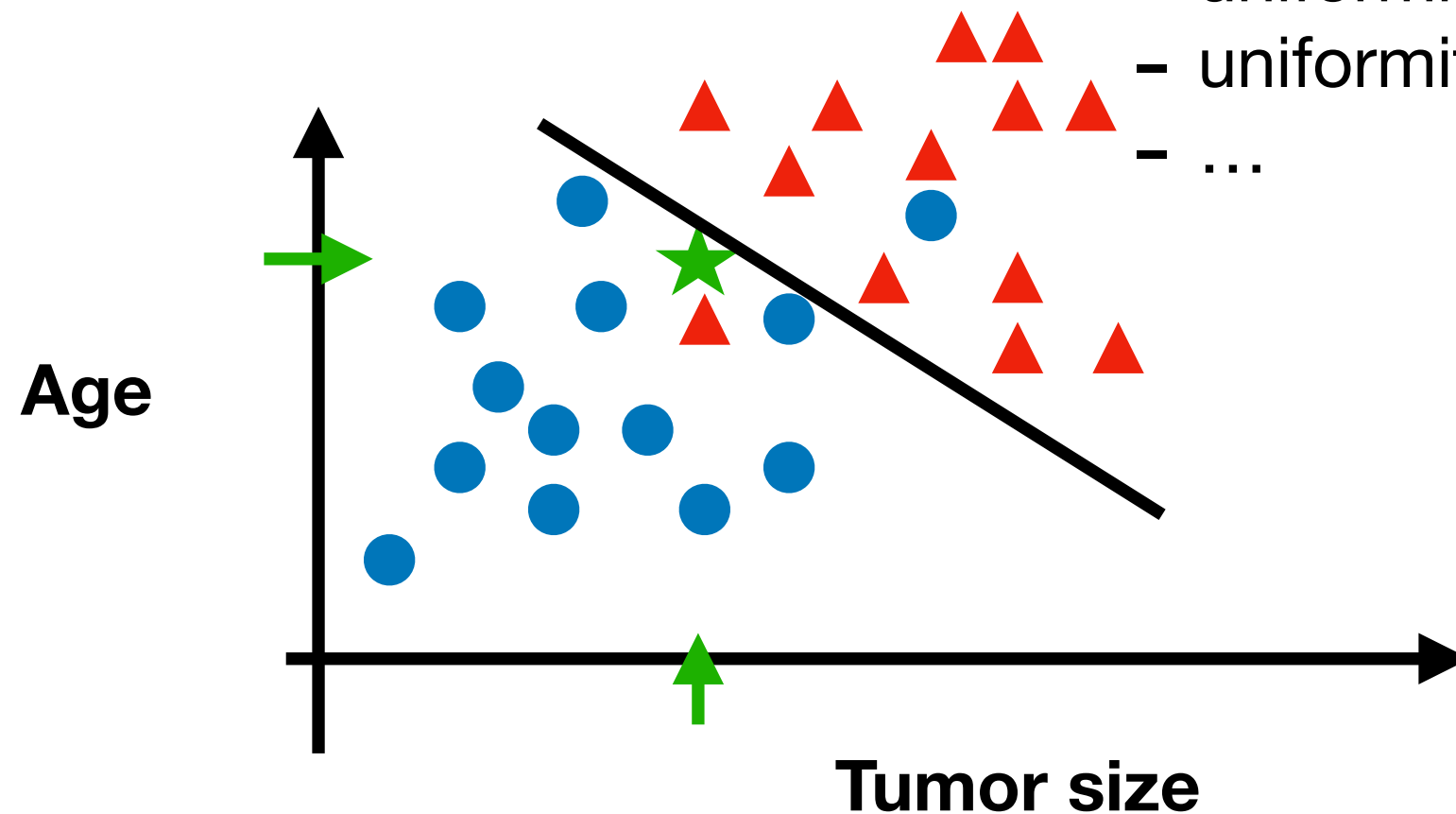


Technologies

- Machine Learning
 - Supervised Learning:

features:

- clump thickness
- uniformity of cell size
- uniformity of cell shape
- ...



Question

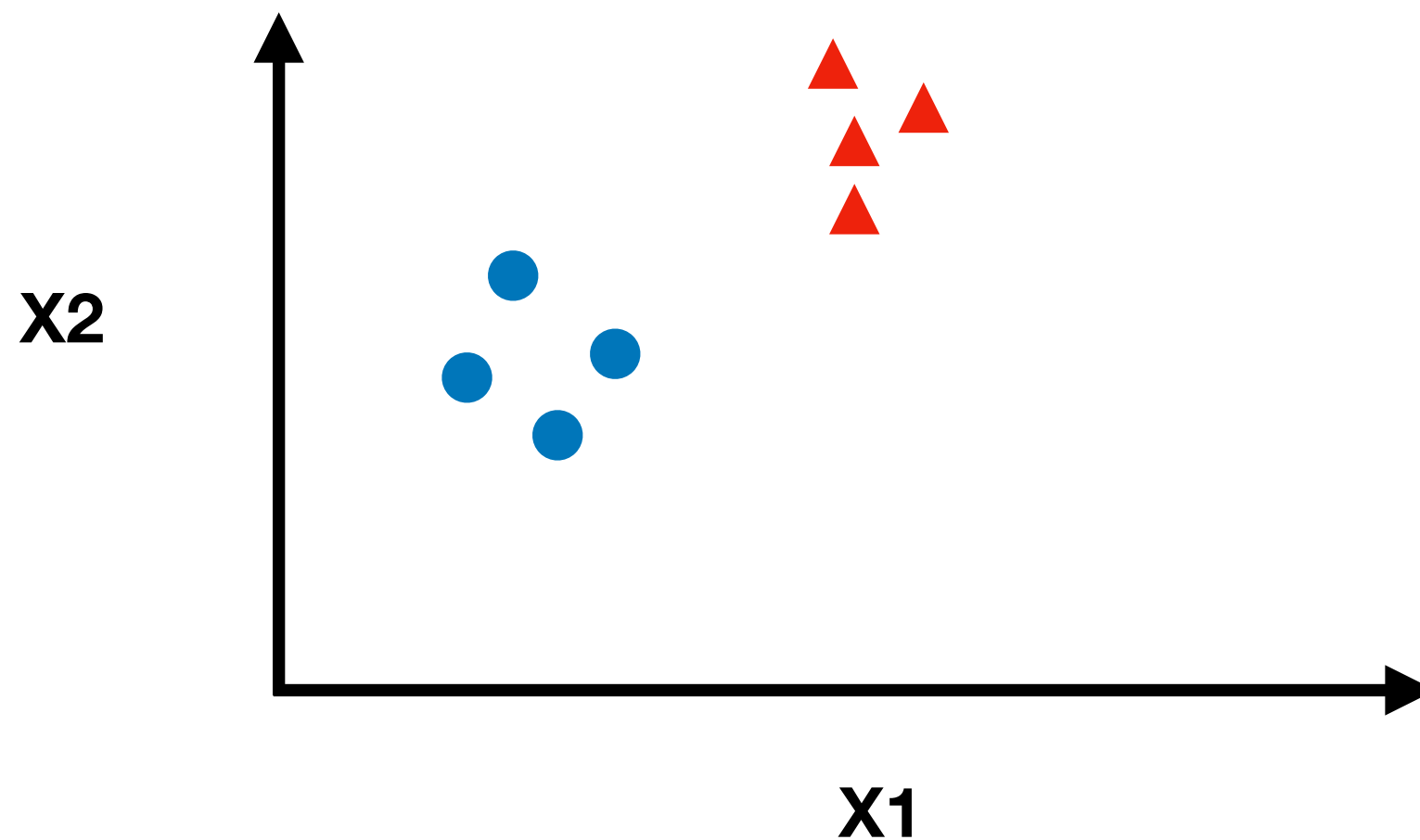
- What learning alg. would you use?
 1. You want to predict how many students will have lunch today in the No. 2 canteen?
 2. You want to examine individual lunch preferences. For each student decide which canteen he/she goes at noon today.

Answer

1. A regression problem
2. A classification problem

Technologies

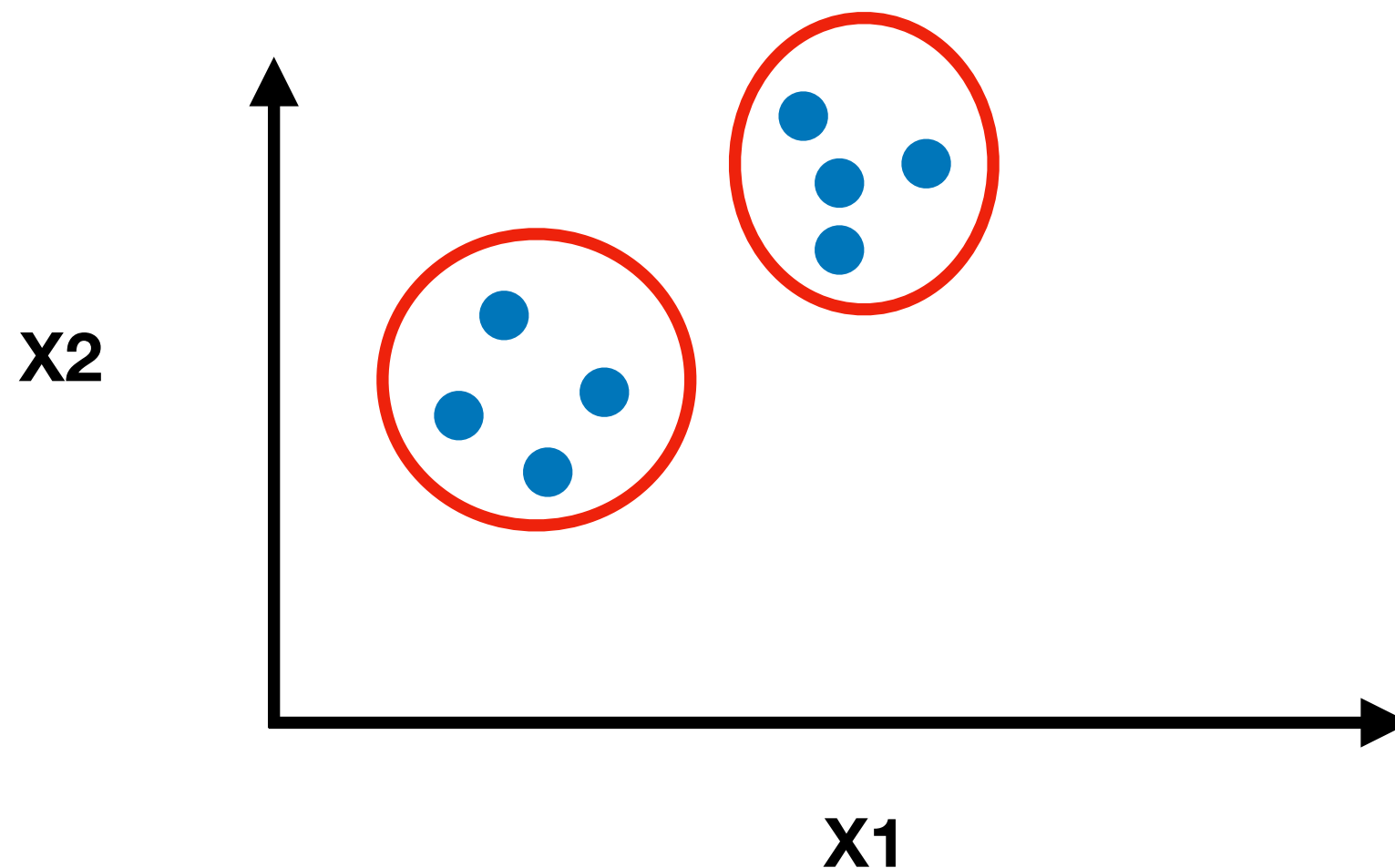
- Machine Learning
 - Unsupervised Learning:



Technologies

- Machine Learning
 - Unsupervised Learning:

Clustering



Technologies

- Machine Learning
- Unsupervised Learning:

The screenshot shows the Google News interface. On the left is a sidebar with navigation options: 'Top stories', 'For you', 'Favorites', 'Saved searches', and a list of categories (U.S., World, Business, Technology, Entertainment, Sports, Science, Health). At the top, there's a search bar and a 'Google News' logo. The main content area features the headline 'US and China extend trade talks' with a 'Full coverage' link and a 'Sort' dropdown. Below this, the 'Top coverage' section is highlighted with a red rounded rectangle. It contains four news items:

- CNN**: Trump predicts 'very good chance' of China trade deal (4 hours ago). Image of Donald Trump.
- USA TODAY**: Trump says U.S., China reach deal on currency manipulation as trade talks continue (today). Image of Donald Trump.
- CNBC**: Trump, Xi summit being discussed for late March as China commits to buying \$1.2 trillion in US goods (today). Image of Donald Trump.
- Bloomberg**: Chinese Official Extends U.S. Trade Trip as Tariff-Hike Threat Endures (today). Image of a group of people.

Below the 'Top coverage' section is a 'Videos' section, which currently shows a video player with a book cover titled 'TRUMPONOMICS' by Stephen Moore and Arthur R. Laffer, Ph.D. The book cover includes the subtitle 'Inside the America First Plan to Revive Our Economy'.

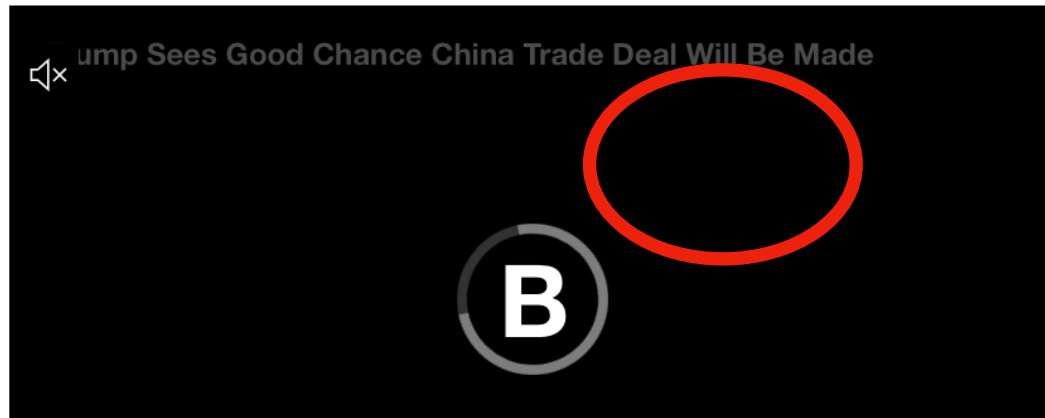
Politics

Chinese Official Extends U.S. Trade Trip as Tariff-Hike Threat Endures

By [Jennifer Jacobs](#), [Jenny Leonard](#), [Saleha Mohsin](#), and [Belinda Cao](#)

February 23, 2019, 4:02 AM GMT+8 Updated on February 23, 2019, 8:42 AM GMT+8

- ▶ Trump and Liu signal deal is likely; Lighthizer cites hurdles
- ▶ The two sides agreed on a currency provision, Mnuchin says

You have 9 free articles remaining. [Subscribe for unlimited access.](#)[View Subscription Offers](#)[Sign in](#)Bloomberg Anywhere
clients get **free access**

CNN travel

CNN politics

45

CONGRESS

SUPREME COURT

2018 ELECTION RESULTS

Trump predicts 'very good chance' of trade deal

By [Betsy Klein](#) and [Donna Borak](#), CNN

Updated 0705 GMT (1505 HKT) February 23, 2019



NEW



CNBC

MENU

MARKETS

BUSINESS

INVESTING

TECH

POLITICS

CNBC TV

PRE-MKT

US

EU

CHINA

ASIA

WORLD

FX

CRYPTOCURRENCY

COMMODITIES

BONDS

ETF MARKETS

SIGN IN

PRO

WATCHLIST

MAKE IT

SEARCH QUOT

USA

INTL

Trump, Xi summit being discussed for late March as China commits to buying \$1.2 trillion in US goods

- The United States and China are discussing a late March meeting between President Donald Trump and Chinese President Xi Jinping, according to sources.
- CNBC sources also confirmed that China has committed to buying up to \$1.2 trillion in U.S. goods.
- The two sides remain far apart on certain issues including the forced transfer of technology and intellectual property.

Kayla Tausche | Thomas Franck

Published 17 Hours Ago | Updated 17 Hours Ago

CNBC



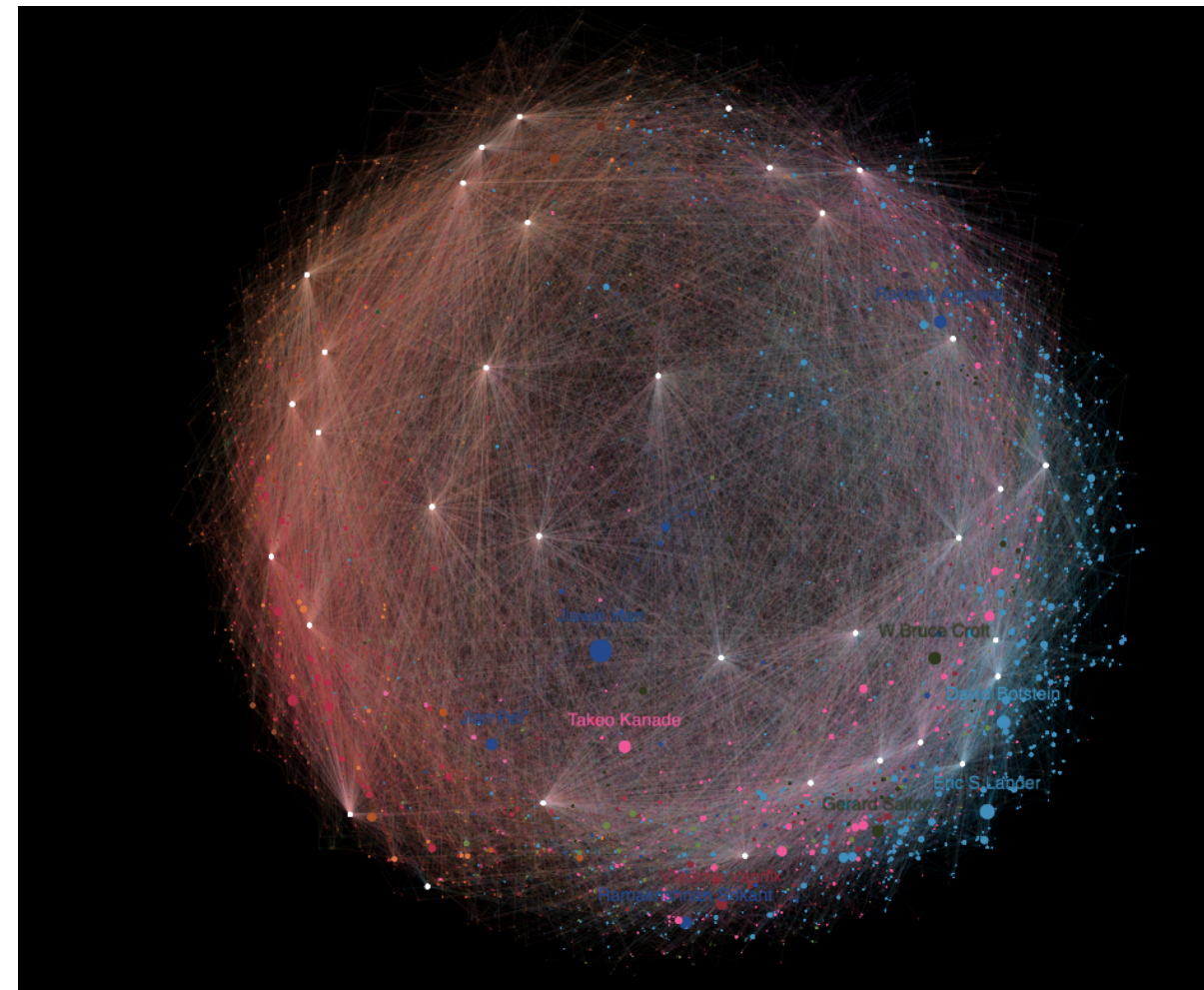
FROM THE WEB

Sponsored Links by Taboola

**This New \$89 Trap Finally Solves The Japan Mosquito Problem**
Electric Mosquito Killer**Autonomous Vehicle Reality Check: The State of Consumer Readiness ...**
LexisNexis



Organize computing clusters



Social network analysis

Figures from acemap.info



Market segmentation



Genome analysis

Question

- Of the following examples, which would you address using an unsupervised learning alg.?
 - Given email labeled as spam/not spam, learn a spam filter.
 - Given a set of news articles found on the web, group them into set of articles about the same story.
 - Given a database of customer data, automatically discover market segments and group customers into different market segments.
 - Given a dataset of patients diagnosed as either having diabetes or not, learn to classify new patients as having diabetes or not.

Answer

- Of the following examples, which would you address using an unsupervised learning alg.?
 - ☐ Given email labeled as spam/not spam, learn a spam filter.
 - ☒ Given a set of news articles found on the web, group them into set of articles about the same story.
 - ☒ Given a database of customer data, automatically discover market segments and group customers into different market segments.
 - ☐ Given a dataset of patients diagnosed as either having diabetes or not, learn to classify new patients as having diabetes or not.

Technologies

- **Database Systems & Data Warehouses**: focuses on the creation, maintenance, and use of databases for organizations and users.
 - data mining use scalable database technologies to achieve high **efficiency** and **scalability**
- **Information Retrieval**: searching for documents or information in documents
 - differ from database systems in that:
 1. data under search are **unstructured**
 2. queries are formed by **keywords**
 - method: **probabilistic models**
 - e.g., language model, topic model...

Summary

- Technologies used to mine data:
 - Statistics
 - Machine learning
 - Database systems and data warehouses
 - Information retrieval

Outline

- Why Data Mining?
- What is Data Mining?
- What Kinds of Data Can be Mined?
- What Kinds of Knowledge Can be Mined?
- What are the Technologies?
- What are the Targeted Applications?

Applications

- Case 1: Frequent Item Set Mining



frequent item set: {milk, bread}

association rules: milk => bread [support = 2%, confidence = 60%]

Applications

- Case 2: Web Search Engines

Google

MacBook pro

macbook pro price
macbook pro 2016
macbook pro best buy
macbook pro 2017
macbook pro touch bar
macbook pro 2018
apple macbook pro 2017
macbook pro refurbished

query recommendations

Report inappropriate predictions

Adobe has fixed a Premiere Pro CC issue that blew some MacBook Pro speakers
The Verge
9 hours ago

Amazon is blowing out MacBook and MacBook Pro refurbs, today only
BGR.com
1 day ago

MacBook Pro 2019: New 16-Inch Could Tap Intel Gen 9
Forbes
6 hours ago

→ More for MacBook pro

MacBook Pro - Apple
<https://www.apple.com/macbook-pro/>
The ultimate pro notebook, **MacBook Pro** features faster processors, upgraded memory, the Apple T2 chip, and a Retina display with True Tone technology.

Buy MacBook Pro

13-inch MacBook Pro

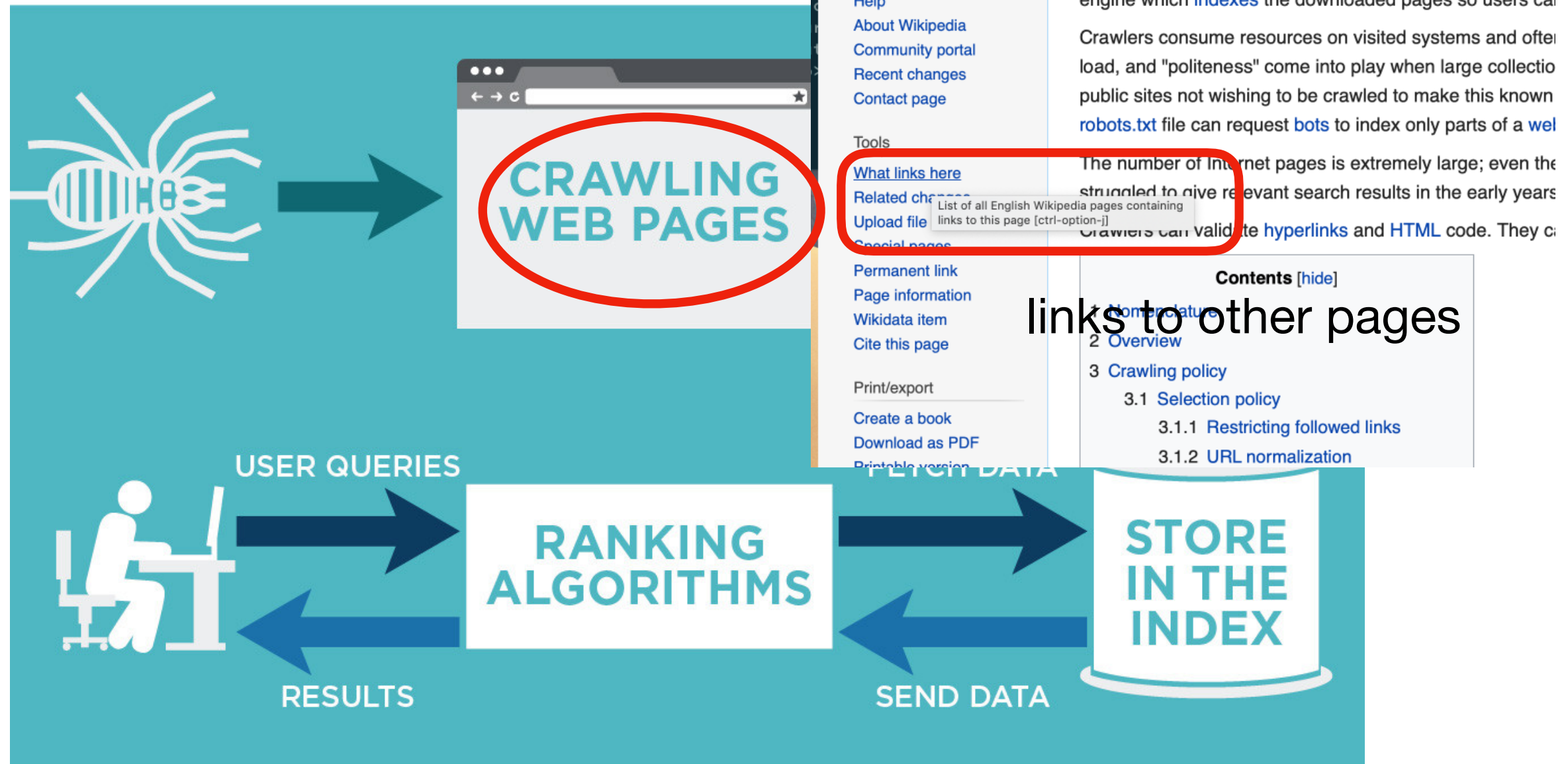
15-inch MacBook Pro - Space ...
Radeon Pro 560X with 4GB of GDDR5 memory. Radeon Pro ...

MacBook Pro - Technical ...
Complete technical specifications for MacBook Pro, including ...

page ranking

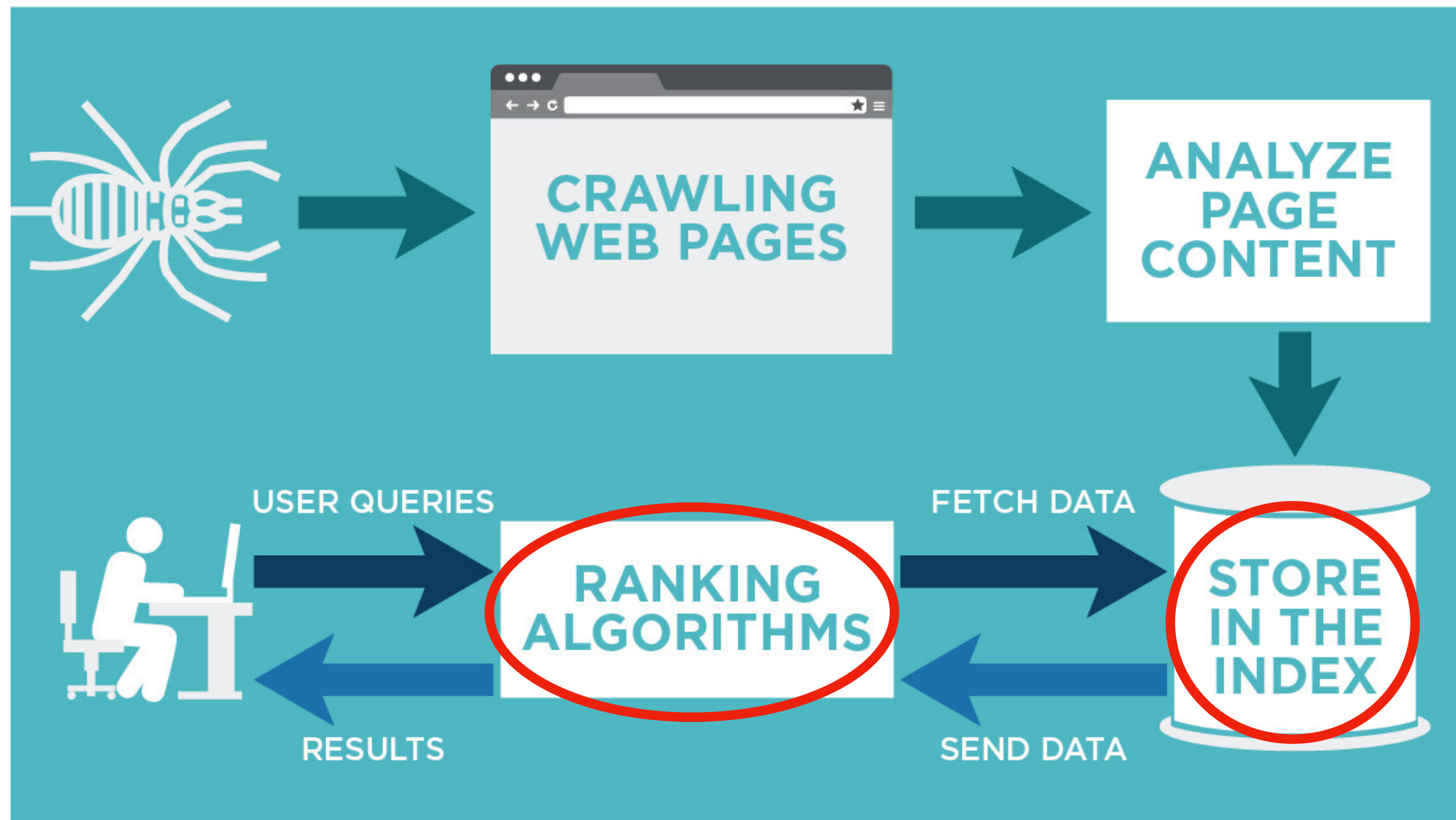
Applications

- Case 2: Web Search Engines



Applications

- Case 2: **Web Search Engines**



Applications

- Case 3: **Ads Display**



All Shopping News Images Videos More Settings Tools

About 552,000,000 results (0.41 seconds)

See iPhone Xs

Sponsored ⓘ



Apple iPhone XS

\$999.00

Apple
Free shipping



Apple iPhone XS Max

\$1,249.00

Apple
Free shipping



Apple iPhone X

\$18.75/mo
For 18 months
Sprint

[iPhone XS Special Offer | AT&T® Official Site | att.com](#)

[Ad] [www.att.com/iPhone](#) ▼

★★★★★ Rating for att.com: 4.5

Learn How to Get \$750 Off An **iPhone XS** When You Buy One w/ AT&T®. Check Order Status.

[iPhone XS Max Offer](#) · [Apple Watch Nike+ Offer](#) · [AT&T® Unlimited &MoreSM](#)

[iPhone XS | Limited time | From \\$699 with select trade-in](#)

[Ad] [www.apple.com/](#) ▼

Welcome to the big screens. Trade in your current **iPhone** and upgrade to a new one. Dual SIM.

[iPhone XS - Apple](#)

<https://www.apple.com/iphone-xs/> ▼

- Whether user likes the ads.
- How advertisers set bid price.

Applications

- Case 4: Information Extraction

Sample Job Posting:

Job Title: Senior DBMS Consultant

Location: Dallas,TX

Responsibilities:

DBMS Applications consultant works with project teams to define DBMS based solutions that support the enterprise deployment of Electronic Commerce, Sales Force Automation, and Customer Service applications.

Desired Requirements:

3-5 years exp. developing Oracle or SQL Server apps using Visual Basic, C/C++, Powerbuilder, Progress, or similar.

Recent experience related to installing and configuring Oracle or SQL Server in both dev. and deployment environments.

Desired Skills:

Understanding of UNIX or NT, scripting language. Know principles of structured software engineering and project management

Filled Job Template:

title: Senior DBMS Consultant

state: TX

city: Dallas

country: US

language: Powerbuilder, Progress, C, C++, Visual Basic

platform: UNIX, NT

application: SQL Server, Oracle

area: Electronic Commerce, Customer Service

required years of experience: 3

desired years of experience: 5



Job Template

title:

state:

city:

country:

language:

platform:

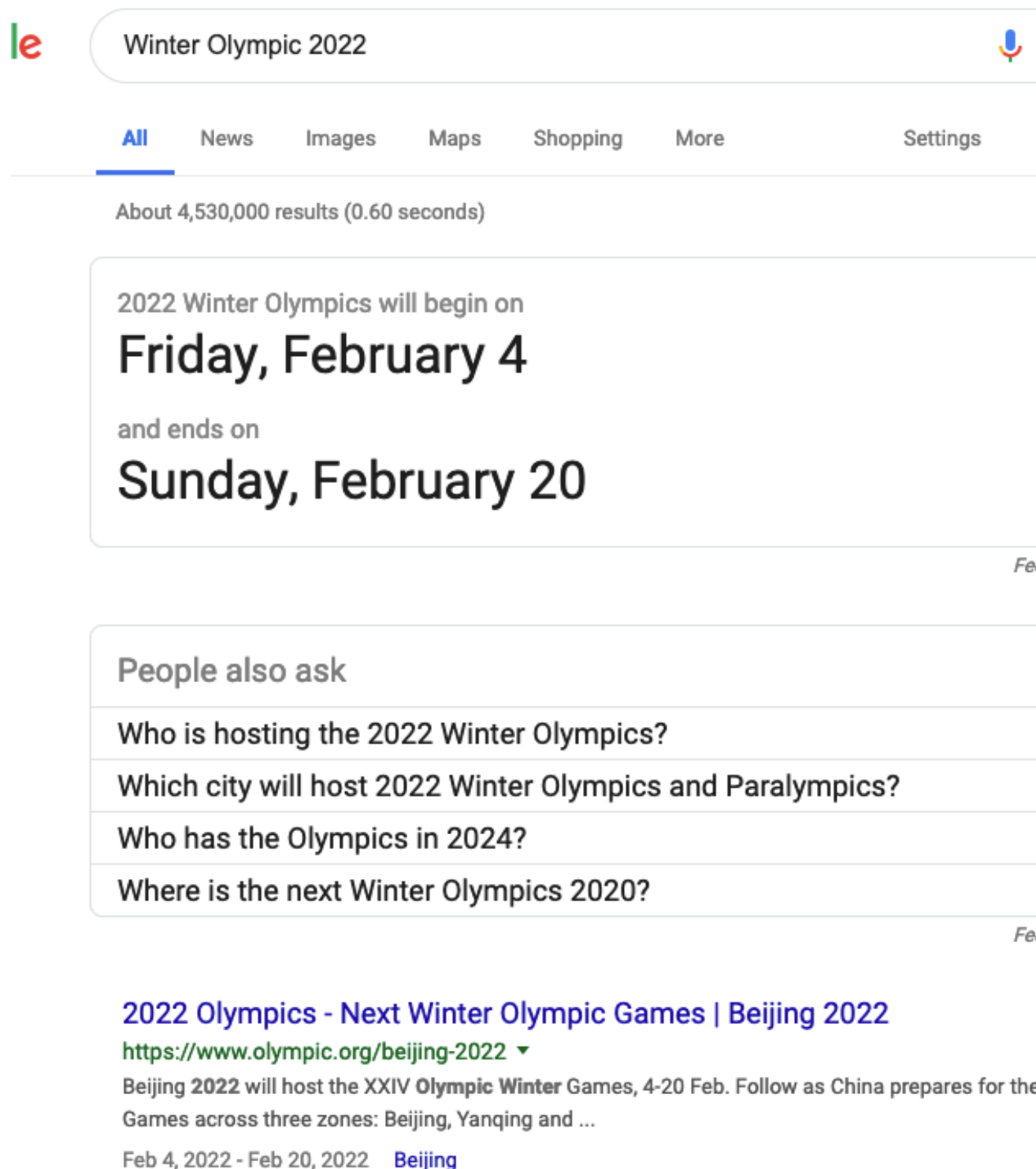
application:

area:

...

Applications

- Case 4: Information Extraction



Winter Olympic 2022

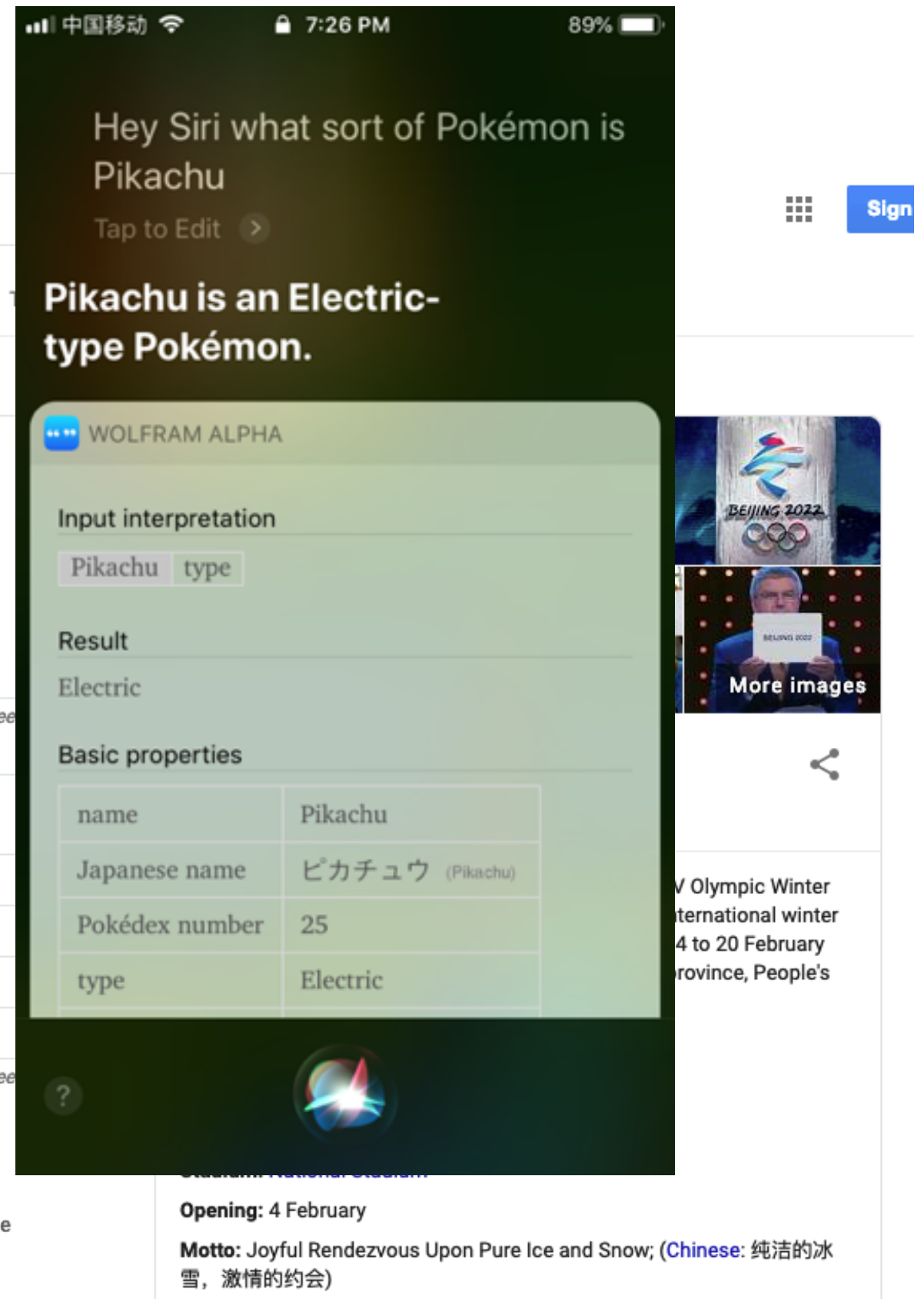
About 4,530,000 results (0.60 seconds)

2022 Winter Olympics will begin on
Friday, February 4
and ends on
Sunday, February 20

People also ask

- Who is hosting the 2022 Winter Olympics?
- Which city will host 2022 Winter Olympics and Paralympics?
- Who has the Olympics in 2024?
- Where is the next Winter Olympics 2020?

2022 Olympics - Next Winter Olympic Games | Beijing 2022
<https://www.olympic.org/beijing-2022>
Beijing 2022 will host the XXIV Olympic Winter Games, 4-20 Feb. Follow as China prepares for the Games across three zones: Beijing, Yanqing and ...
Feb 4, 2022 - Feb 20, 2022 [Beijing](#)



Hey Siri what sort of Pokémon is Pikachu

Pikachu is an Electric-type Pokémon.

WOLFRAM ALPHA

Input interpretation
Pikachu type

Result
Electric

Basic properties

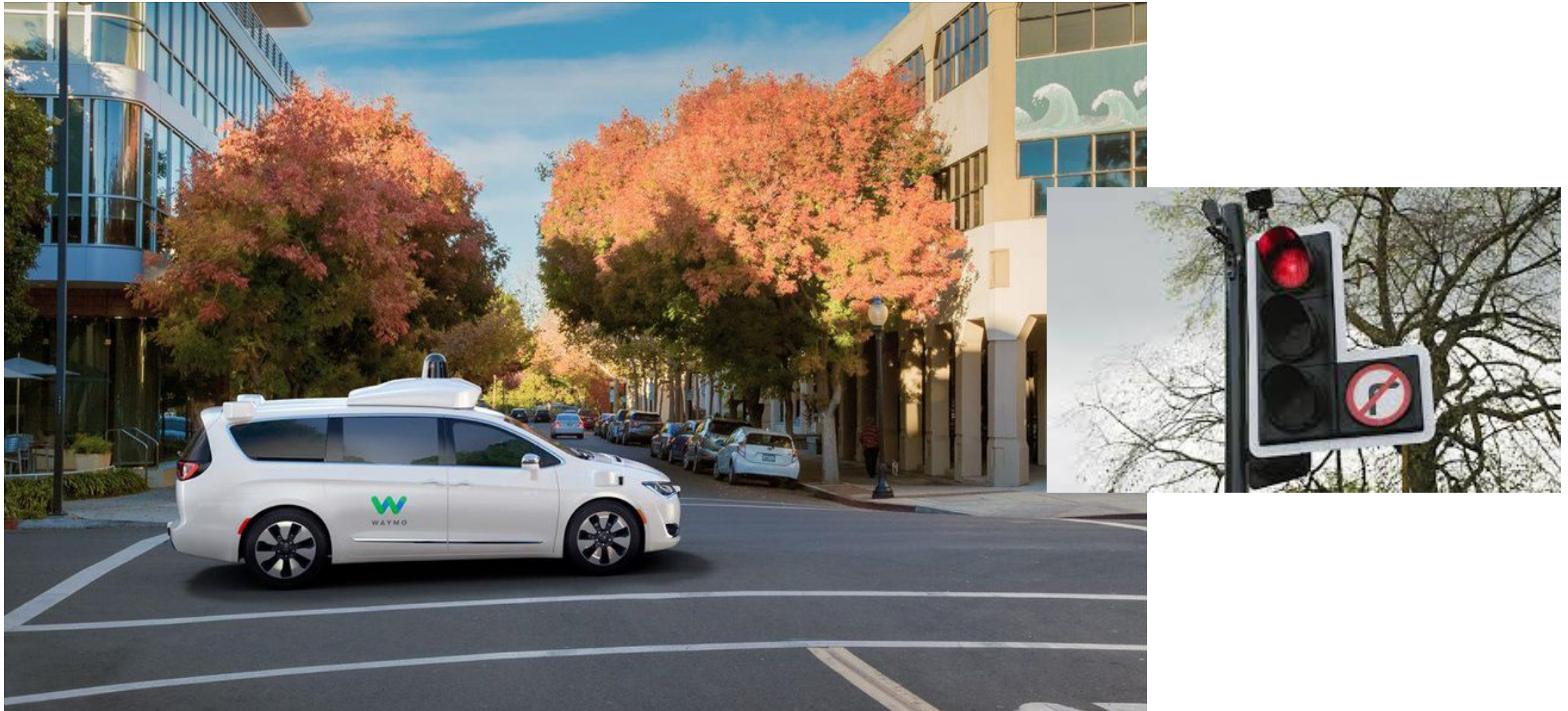
name	Pikachu
Japanese name	ピカチュウ (Pikachu)
Pokédex number	25
type	Electric

Opening: 4 February

Motto: Joyful Rendezvous Upon Pure Ice and Snow; (Chinese: 纯洁的冰雪, 激情的约会)

Applications

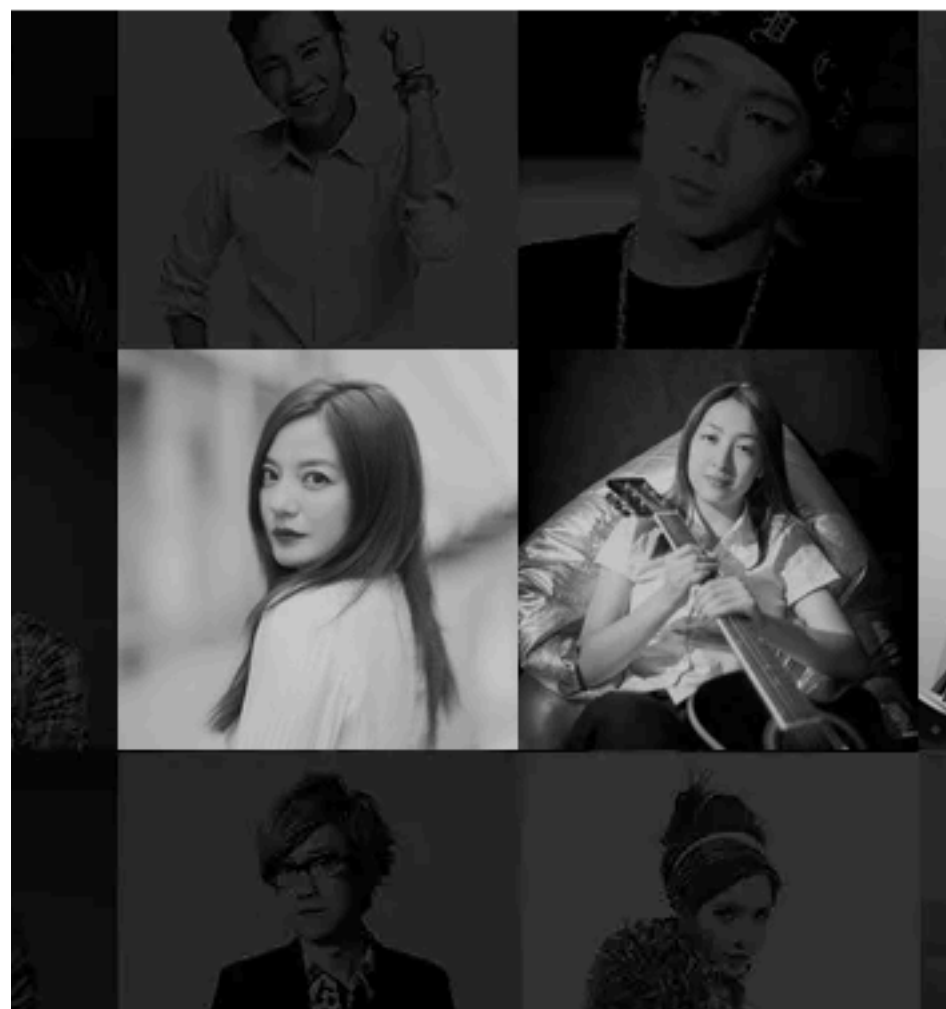
- Case 5: **Computer Vision**



Minivan Waymo under tests.

Applications

- Case 6: Interactive Recommendation



日本动漫中
播放量: 39

C D E F G H I J K L M N O P Q R S T U V

港台 欧美 日本 韩国 其他

组合

Summary

- Data mining is to discover implicit knowledge through massive data
- Data sources
- Knowledge types
- Technologies
- Applications