University Of Danang
University Of Economics
THE FACUTY OF ELECTRONIC COMMERCE

AI in Business

AI in Business introduction

Tran Van Loc

AI and Business

- Artificial intelligence (AI) is about getting computers to do things that require human intelligence
  - Understanding language
  - Reasoning
  - Navigating the physical world
  - Learning
- Machine learning - focused on getting computers to learn without explicitly programming them
- AI is increasingly being seen as the next phase of digital transformation

AI application on your phone?

AI application?

- Guide to make your very own chatbot!

AI application?

- NOW!

Digital Transformation Timeline

- Multiple digital technologies have helped transform business
- Companies that were slow to react paid the price

- AI is likely to be equally transformative and can potentially be viewed as a general-purpose technology

What are General-Purpose Technologies (GPTs)?

- GPTs are “characterized by the potential for pervasive use in a wide range of sectors and by their technological dynamism” (Bresnahan & Trajtenberg)
  - GPTs stimulate innovation and economic growth
  - GPTs inform products strategy & organizational design/strategy
- 3 factors can indicate that a new technology shows promise as a GPT:
  1. Widespread use of the technology across many industries
  2. High volume of research jobs related to the technology
  3. Research jobs widespread across many industries

How Does AI Show Promise as a GPT?

Research by Goldfarb et al (2019)

Technology	Total Jobs	Total Research Jobs	Percent Research Jobs
Machine Learning	370,572	5715000%	14.6%
GIS Software	142,645	899300%	5.8%
CRISPR	6,716	525000%	78.4%
Quantum Computing	6,355	134800%	17.4%
Fracking	2,957	14100%	2.8%
Robotics	284,474	2244800%	7.3%
Nanotechnology	4,883	251900%	46%
Internet-of-things	124,506	575000%	4.3%
Cloud Computing	1,415,289	2939200%	2.2%

High volume of jobs and research jobs

- Research jobs are a particularly important indicator of GPTs because they demonstrate that a technology is “capable of ongoing improvement” and could have significant future potential, beyond what is currently recognized

ML's Promise as a GPT (cont.)

→ Jobs widespread across industries

Industry NAICS2	AI	GIS	Quantum	Fracking	Robotics	Nanotech	IoT	CRISPR	Cloud	Total
Accommodation and Food Services	5,389	883	4	5	1,516	5	828	2	15,090	5,594,113
Admin & Support & Waste Mgmt & Remediation Srvcs.	4,142	2,099	0	75	2,363	6	1,836	9	26,763	2,237,225
Agriculture, Forestry, Fishing and Hunting	820	1,388	0	3	384	0	146	0	2,310	97,644
Arts, Entertainment and Recreation	437	917	0	2	925	7	109	0	2,591	578,889
Construction	335	2,406	1	45	2,330	22	361	0	3,709	850,081
Educational Services	12,467	6,836	246	62	10,868	1,764	1,503	1,384	17,848	4,018,454
Finance and Insurance	40,261	3,871	26	32	7,809	7	4,569	35	98,700	5,787,861
Health Care and Social Assistance	102,401	34,512	158	612	114,965	734	28,477	1,793	481,770	38,700,000
Information	32,613	4,483	189	26	6,213	85	17,079	2	130,353	2,229,564
Management of Companies and Enterprises	348	302	0	4	160	2	108	0	1,156	80,359
Manufacturing	37,215	7,738	257	144	73,053	582	24,472	1,733	96,368	6,307,396
Mining	674	1,098	0	828	489	2	257	0	1,736	216,880
Other Services (except Public Administration)	1,614	1,695	0	77	2,314	12	181	5	5,786	1,120,323
Professional, Scientific, & Technical Services	97,704	39,478	5,403	395	36,715	1,550	33,912	1,683	435,035	7,333,834
Public Administration	4,838	21,175	41	29	3,467	70	680	46	13,446	2,448,184
Real Estate Rental and Leasing	2,086	3,302	5	16	3,815	0	500	1	8,000	1,293,048
Retail Trade	21,705	1,220	14	132	11,214	21	5,917	2	54,247	8,720,486
Transportation and Warehousing	3,411	2,639	1	398	2,803	6	1,236	2	11,433	4,668,538
Utilities	801	3,896	0	51	886	6	493	1	2,877	333,522
Wholesale Trade	1,311	2,680	10	21	2,185	2	1,842	18	6,071	620,793
Total	370,572	142,645	6,355	2,957	284,474	4,883	124,506	6,716	1,415,289	93,237,194

Number of jobs in data by industry and technology (2015-2018)

Research by Goldfarb et al (2019)

ML's Promise as a GPT (cont.)

→ Jobs widespread across industries

Industry NAICS2	AI	GIS	Quantum	Fracking	Robotics	Nanotech	IoT	CRISPR	Cloud	Total
Accommodation and Food Services	5,389	883	4	5	1,516	5	828	2	15,090	5,594,113
Admin & Support & Waste Mgmt & Remediation Srvcs.	4,142	2,099	0	75	2,363	6	1,836	9	26,763	2,237,225
Agriculture, Forestry, Fishing and Hunting	820	1,388	0	3	384	0	146	0	2,310	97,644
Arts, Entertainment and Recreation	437	917	0	2	925	7	109	0	2,591	578,889
Construction	335	2,406	1	45	2,330	22	361	0	3,709	850,081
Educational Services	12,467	6,836	246	62	10,868	1,764	1,503	1,384	17,848	4,018,454
Finance and Insurance	40,261	3,871	26	32	7,809	7	4,569	35	98,700	5,787,861
Health Care and Social Assistance	102,401	34,512	158	612	114,965	734	28,477	1,793	481,770	38,700,000
Information	32,613	4,483	189	26	6,213	85	17,079	2	130,353	2,229,564
Management of Companies and Enterprises	348	302	0	4	160	2	108	0	1,156	80,359
Manufacturing	37,215	7,738	257	144	73,053	582	24,472	1,733	96,368	6,307,396
Mining	674	1,098	0	828	489	2	257	0	1,736	216,880
Other Services (except Public Administration)	1,614	1,695	0	77	2,314	12	181	5	5,786	1,120,323
Professional, Scientific, & Technical Services	97,704	39,478	5,403	395	36,715	1,550	33,912	1,683	435,035	7,333,834
Public Administration	4,838	21,175	41	29	3,467	70	680	46	13,446	2,448,184
Real Estate Rental and Leasing	2,086	3,302	5	16	3,815	0	500	1	8,000	1,293,048
Retail Trade	21,705	1,220	14	132	11,214	21	5,917	2	54,247	8,720,486
Transportation and Warehousing	3,411	2,639	1	398	2,803	6	1,236	2	11,433	4,668,538
Utilities	801	3,896	0	51	886	6	493	1	2,877	333,522
Wholesale Trade	1,311	2,680	10	21	2,185	2	1,842	18	6,071	620,793
Total	370,572	142,645	6,355	2,957	284,474	4,883	124,506	6,716	1,415,289	93,237,194

Number of jobs in data by industry and technology (2015-2018)

Research by Goldfarb et al (2019)

ML's Promise as a GPT (cont.)

➔ Research jobs also widespread across industries

Industry NAICS2	AI	GIS	Quantum	Fracking	Robotics	Nanotech	IoT	CRISPR	Cloud	Total
Accommodation and Food Services	438	24	1	0	43	2	32	2	214	13,410
Admin & Support & Waste Mgmt & Remediation Srvcs.	506	56	0	3	69	1	55	9	472	16,768
Agriculture, Forestry, Fishing and Hunting	243	220	0	0	24	0	6	0	130	9,847
Arts, Entertainment and Recreation	51	110	0	0	12	4	10	0	57	7,823
Construction	53	76	0	0	68	1	13	0	69	6,204
Educational Services	3,991	1,785	71	14	1,979	682	139	1,095	1,023	267,403
Finance and Insurance	6,409	271	15	3	309	2	249	29	2,132	114,244
Health Care and Social Assistance	11,974	1,905	63	21	8,186	366	895	1,393	7,299	1,078,322
Information	6,148	144	19	0	584	21	404	1	3,383	53,490
Management of Companies and Enterprises	42	7	0	0	6	0	9	0	19	1,061
Manufacturing	7,682	382	103	35	6,024	248	1,598	1,363	3,306	357,069
Mining	108	51	0	25	34	1	70	0	37	4,251
Other Services (except Public Administration)	273	162	0	3	102	2	3	0	104	15,440
Professional, Scientific, & Technical Services	13,595	1,423	1,048	23	3,704	1,134	1,798	1,305	8,845	391,943
Public Administration	728	1,544	17	5	401	33	55	37	287	149,862
Real Estate Rental and Leasing	371	544	2	1	83	0	30	0	193	9,976
Retail Trade	3,599	129	9	2	570	14	290	2	1,470	35,200
Transportation and Warehousing	643	98	0	1	95	1	27	1	188	13,639
Utilities	99	28	0	5	29	5	19	0	36	5,762
Wholesale Trade	197	34	0	0	126	2	48	13	128	16,492
Total	57,150	8,993	1,348	141	22,448	2,519	5,750	5,250	29,392	2,568,206

Number of research jobs in data by industry and technology (2015-2018)

Research by Goldfarb et al (2019)

AI and Business

- AI (specifically ML) shows early indicators of being a general-purpose technology
- Implications for firms if ML is a GPT:
  - Realize that most industries are likely to change
  - Be patient: the transformative impact may come with a lag
- To effectively leverage the opportunities, managers need to understand the technology and its applications, and make changes to their "business models, technology infrastructure, organizational processes, and culture"

Content from: "Could Machine Learning Be a General-Purpose Technology? Evidence from Online Job Postings" by Goldfarb, Taska & Teodoridis

University Of Danang
University Of Economics
THE FACUTY OF ELECTRONIC COMMERCE

AI in Business

Big Data Overview

Tran Van Loc

Agenda for Module 1

- Big Data Overview
  - What big data is, how it is being generated, and why it matters
- Big Data Analysis
  - Approach to analysis, analytics competencies, and broad skillset needs
- Big Data Tools
  - Data management tools and data analysis tools
- Extracting Intelligence from Big Data
  - Predictive analytics and implications for business strategy

> “Data is the new Oil. Data is just like crude. It’s valuable, but if unrefined it cannot really be used.”
>
> **— CLIVE HUMBY**

> “We have for the first time an economy based on a key resource [Information] that is not only renewable, but self-generating. Running out of it is not a problem, but drowning in it is.”
>
> **— JOHN NAISBITT**

What is Big Data?

- Data has been very important to businesses for multiple decades, but the emphasis on big data is relatively new
- Data that “exceeds the capacity or capability of…conventional methods and systems” (National Institute of Standards and Technology)
- Big data is not only about volume of data, it is also about:
  - The structure of the data set
  - The speed at which it is created
  - The tools you need to analyze it
  - What you can do with the data set

Source: https://www.technologyreview.com/s/519851/the-big-data-conundrum-how-to-define-it/

source: “Big Data” by Victor Mayer- Schonberger and Kenneth Cukier and https://www.sas.com/en_us/insights/big-data/what-is-big-data.html

Big Data Characteristics

New data characteristics created by today's digitized marketplace

"Terabytes to petabytes of existing data to process"

"Structured, unstructured, text, multimedia"

"Streaming data, milliseconds to seconds to respond"

"Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations"

source: https://www.cbpr.me/big-data-and-public-relations/4vs/
https://www.datasciencecentral.com/profiles/blogs/data-veracity

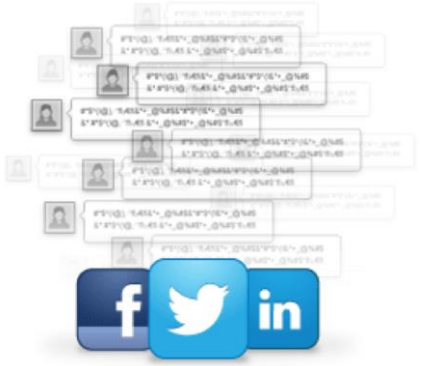
Drivers of Big Data

- Computing capacity: the capacity to store data has increased and the associated cost has decreased
- Data generation: the world is going digital, more people and things are connected than ever before

Drivers of Big Data

Big data allows you to:

1. Ask new questions



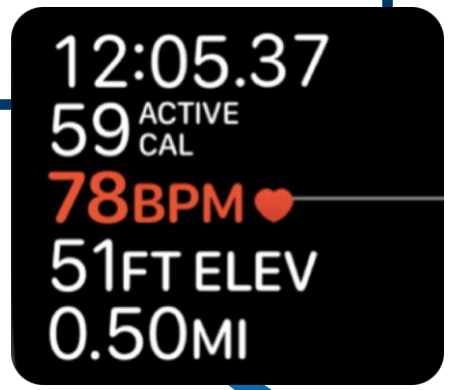
2. Answer same questions better

- This can be done across industries
  - Healthcare, education, transportation, and more

Big Data in Healthcare

Big Data

- Data about heart rate, sleep quality, exercise and more

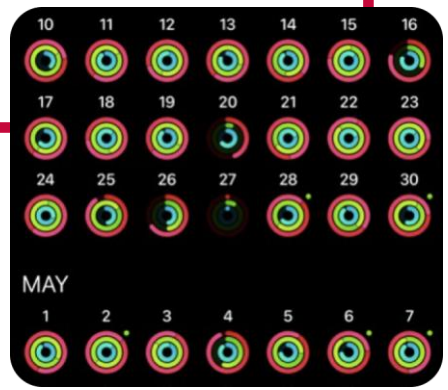


- Gathered continuously and available in real time

Availability

Result

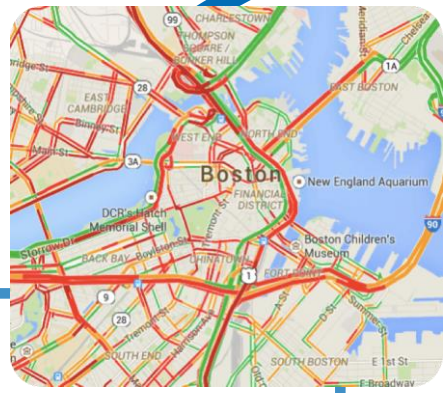
- Improves tracking of health patterns



Big Data in Transportation

Big Data

- Data about traffic, road closures, accidents, etc.

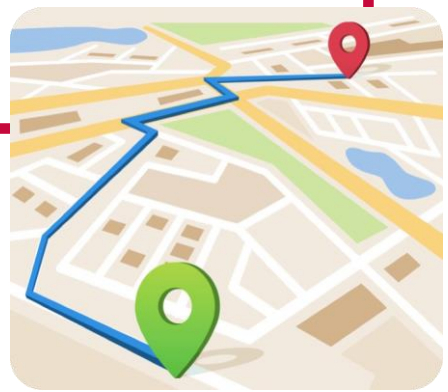


Availability

- Available in a visual format in real time

Result

- Allows for better route planning and scheduling



University Of Danang
University Of Economics
THE FACUTY OF ELECTRONIC COMMERCE

AI in Business

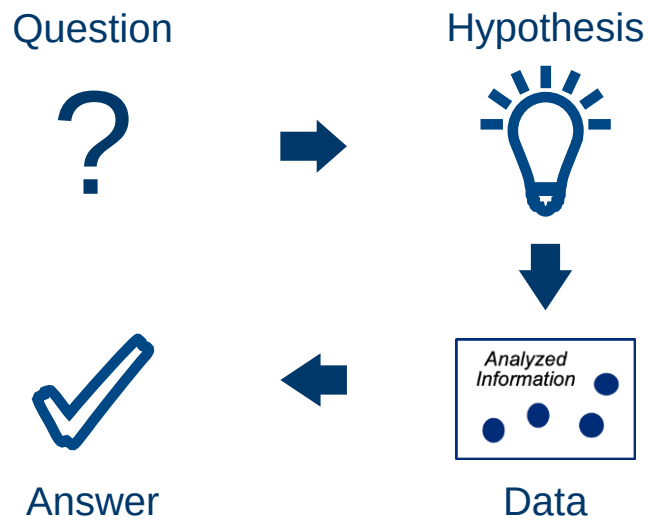
Big Data Analysis

Tran Van Loc

Big Data Approach to Analysis

Traditional Analytics

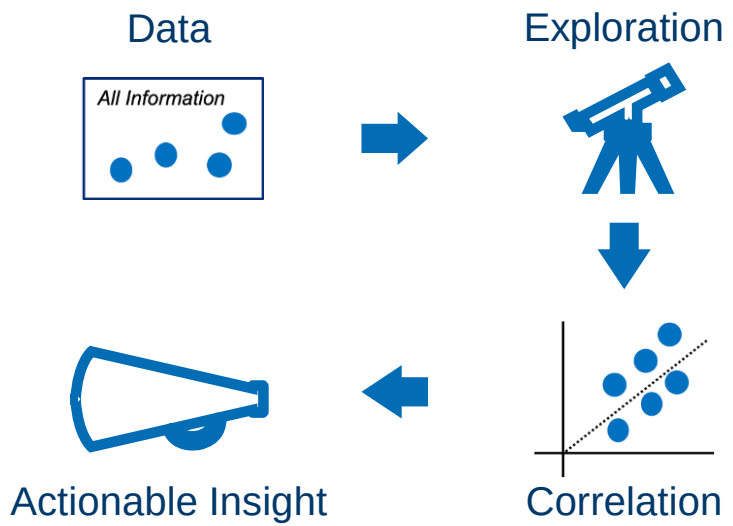
Structured & Repeatable



Start with hypothesis
Test against selected data

Big Data Analytics

Iterative & Exploratory



Data leads the way
Explore *all* data, identify correlations

Big Data Requires a Broad Skillset

Manage the data

Tool Developers

Data Experts: Data architects, governance, policy

Understand the data

Data Science: Statistics, computer science

Visualization Expertise: Interpret data, graph them in meaningful ways

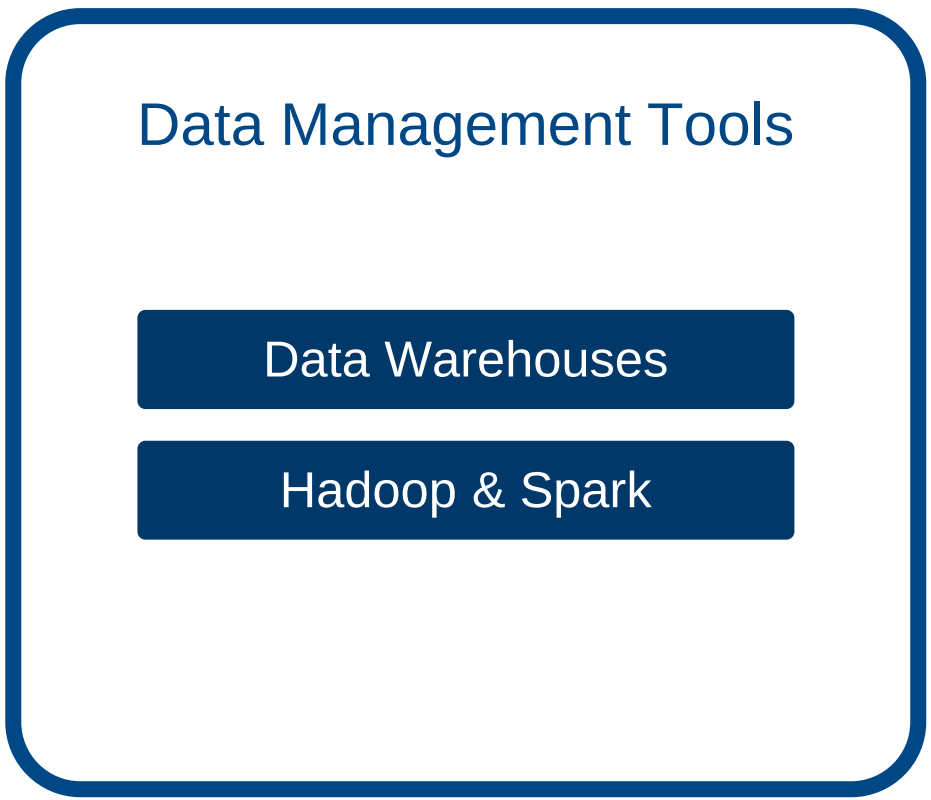
Act on the data

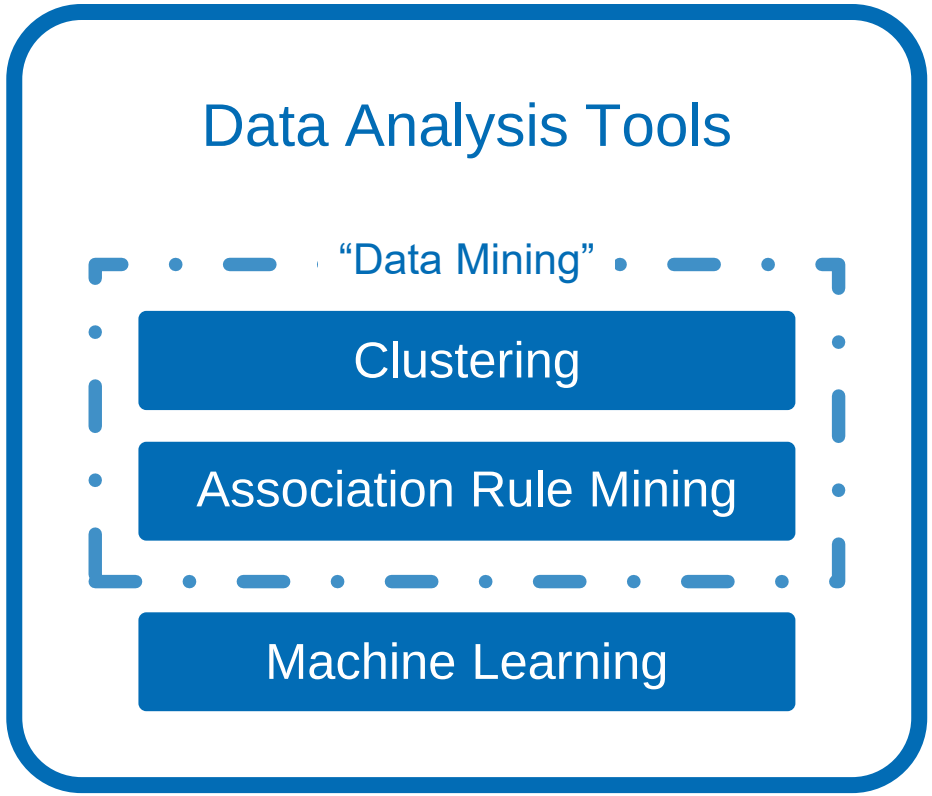
Decision Making - Exec. & Management: Apply data to solve business issues

Industry Vertical Domain Expertise: Identify relevant business issues, ask the right questions

Choosing a Big Data Tool

- Two broad categories of big data tools to choose from, depending on whether you are trying to manage the data or analyze it





University Of Danang
University Of Economics
THE FACUTY OF ELECTRONIC COMMERCE

AI in Business

Data Management Tools

Tran Van Loc

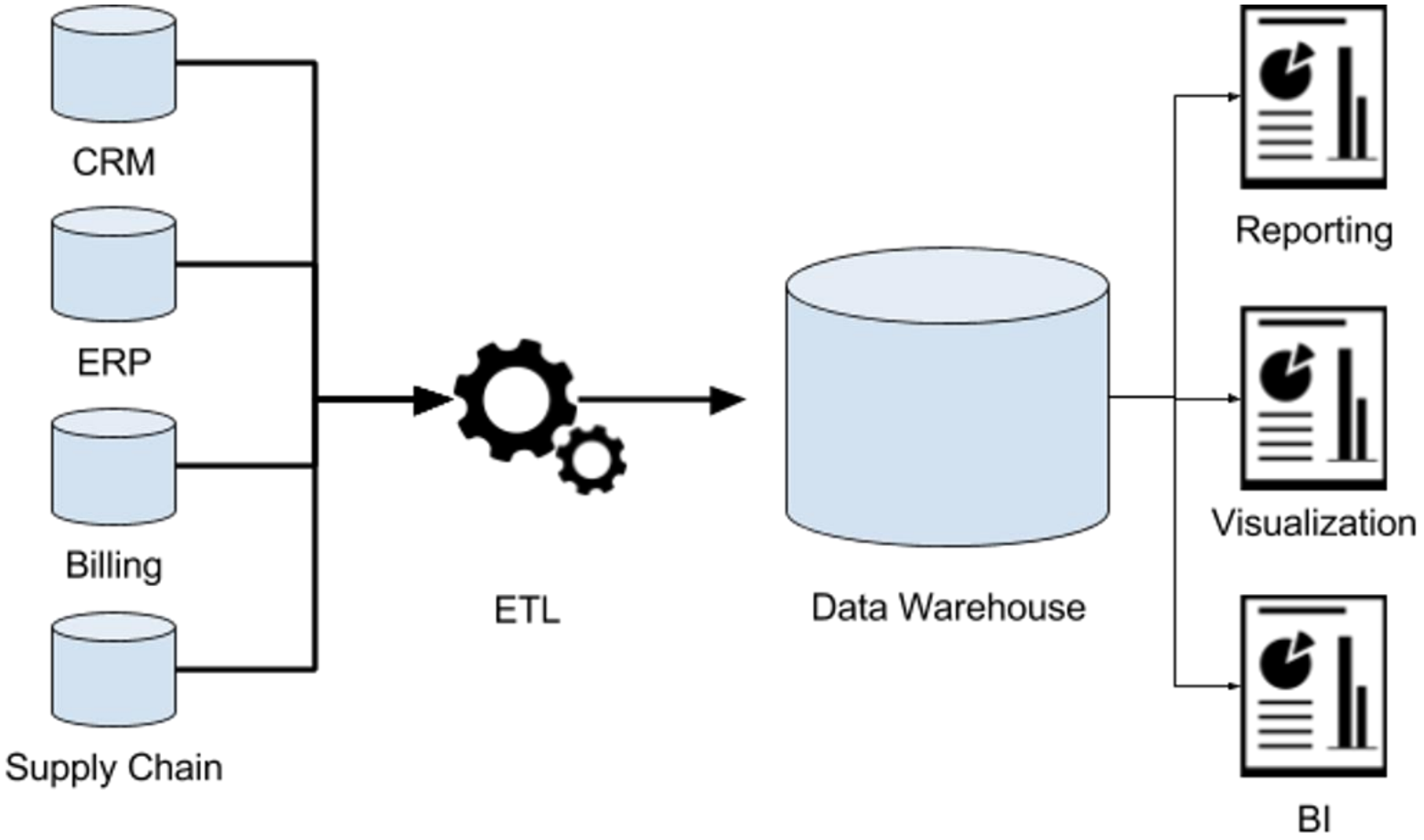
What is a Data Warehouse?

- Database
  - A structured collection of **data**
  - Example: an Excel spreadsheet
- Database Management System (DBMS)
  - Allow users to access and manage the database
  - Example: Excel, Microsoft, Oracle, MongoDB
  - Often referred to as a database

What is a Data Warehouse?

- Data warehouse
  - A particular kind of DBMS
  - Specialized in the type of data it stores - historic data from many sources
  - Specialized in the purpose it serves - analytics
  - Example: Azure SQL Data Warehouse, Google BigQuery, Snowflake, Amazon Redshift

How Do Data Warehouses Work?



source: https://www.dremio.com/what-is-a-data-warehouse/

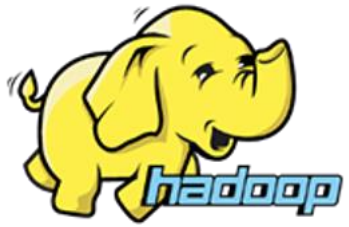
Value of a Data Warehouse

- A single point of access for data
- A history of all the data is stored
- Separates operations from analytics
- Allows answers to questions like: “How much revenue has each product line brought in per month over the last 10 years, broken out by month, city, and state?”

source: https://www.dremio.com/what-is-a-data-warehouse/

Open Source Big Data Tools: Hadoop and Spark

- Big data tools like Hadoop serve two main purposes, storage and processing
  - Storage - lots of data (usually in a distributed fashion)
  - Processing - involves distributed processing across nodes, parallelizing the data process as much as possible
- Popular Hadoop distribution: Cloudera



source: https://hci.stanford.edu/courses/cs448g/a2/files/map_reduce_tutorial.pdf

Open Source Big Data Tools: Hadoop and Spark

- Spark is a more recent version of Hadoop
- Databricks is the dominant company built around Spark

University Of Danang
University Of Economics
THE FACUTY OF ELECTRONIC COMMERCE

AI in Business Introduction

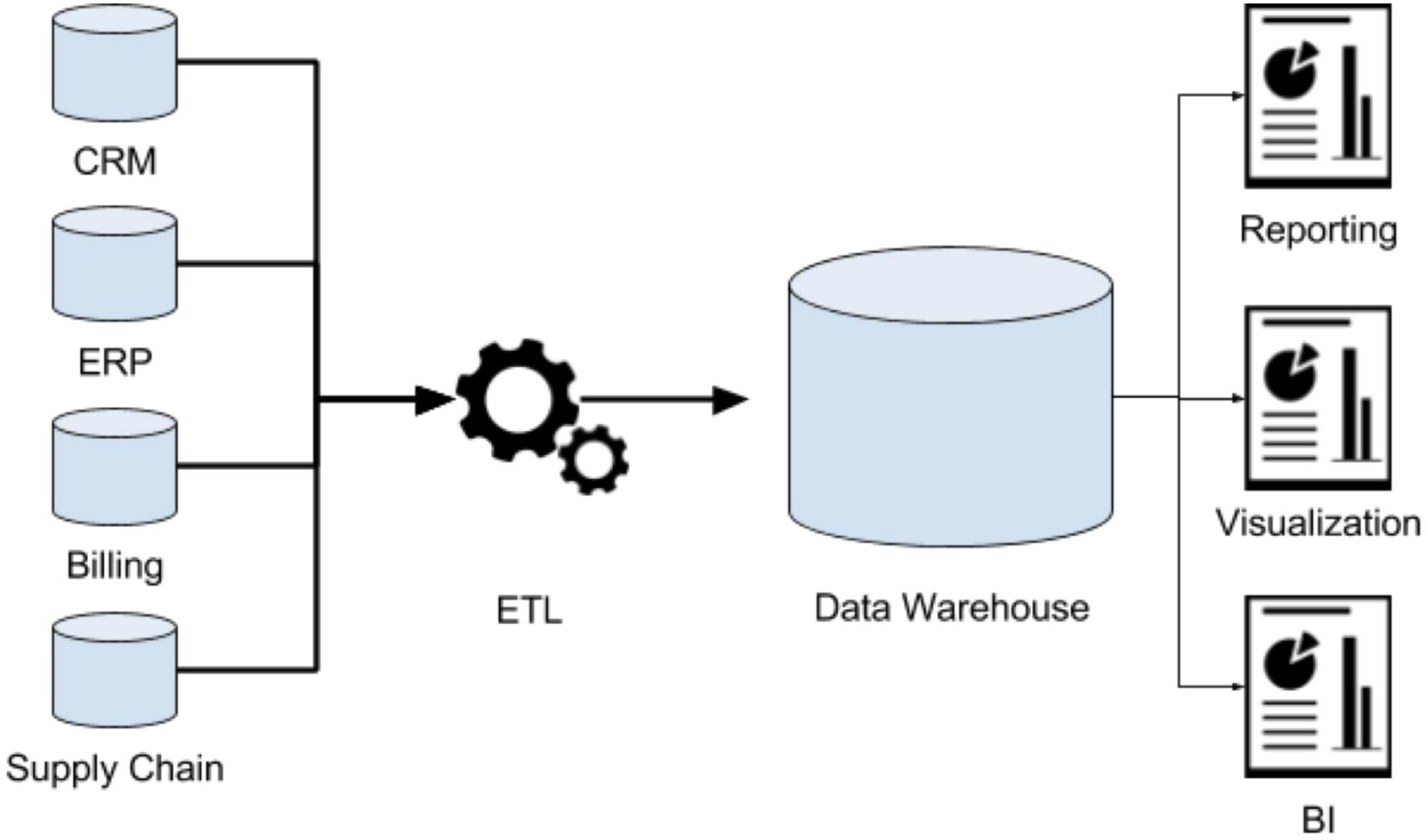
Data Management Infrastructure

Tran Van Loc

What is a Data Warehouse?

- Database:
  - “A structured collection of data”
  - Example: an Excel spreadsheet
- Database Management System (DBMS):
  - “Allows user to access and manage the database”
  - Examples: Excel, Oracle, MongoDB
  - Often also referred to as a database
- Data Warehouse:
  - A particular kind of DBMS, “specialized in the data it stores - historic data from many sources - and the purpose it serves - analytics”
  - Azure SQL Data Warehouse, Google BigQuery, Snowflake, Amazon Redshift

How Do Data Warehouses Work?



source: https://www.dremio.com/what-is-a-data-warehouse/

Value of Data Warehouses

Data warehouses provide:

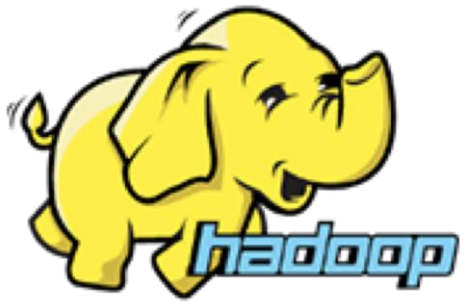
- "A single point of access for data"
- "An assurance of data quality"
- "A history of the data they store"
- "A separation of operations and analytics"
- "A standard set of semantics around data"

Allowing answers to questions like:

- "How much revenue has each product line brought in per month over the last 10 years, broken out by city & state?"
- "What is the average transaction size at our ATMs, broken out by time of day & customer assets?"
- "What is the % employee turnover for the past year in stores that have been open for at least 3 years. How many hours did those employees work per week?"

What is a Data Warehouse?

- Big data tools like Hadoop solve 2 problems:
  - Storage of lots of data (usually in a distributed manner) - E.g. HDFS
  - Processing big data, which usually involves (i) distributed processing (1000s of servers), (ii) parallelization - E.g. MapReduce
- Example - when you have text from emails, webpages, etc., and you want to count the number of times particular words appear



- In this case, Hadoop with MapReduce will:
  1. Split the text into groups sent to different machines
  2. Map each group into pairs (<key, value>)
  3. Shuffle the pairs to group pairs with the same key
  4. Reduce each group of pairs with the same key

University Of Danang
University Of Economics
THE FACUTY OF ELECTRONIC COMMERCE

AI in Business Introduction

Data Analysis: Extracting Intelligence from Big Data

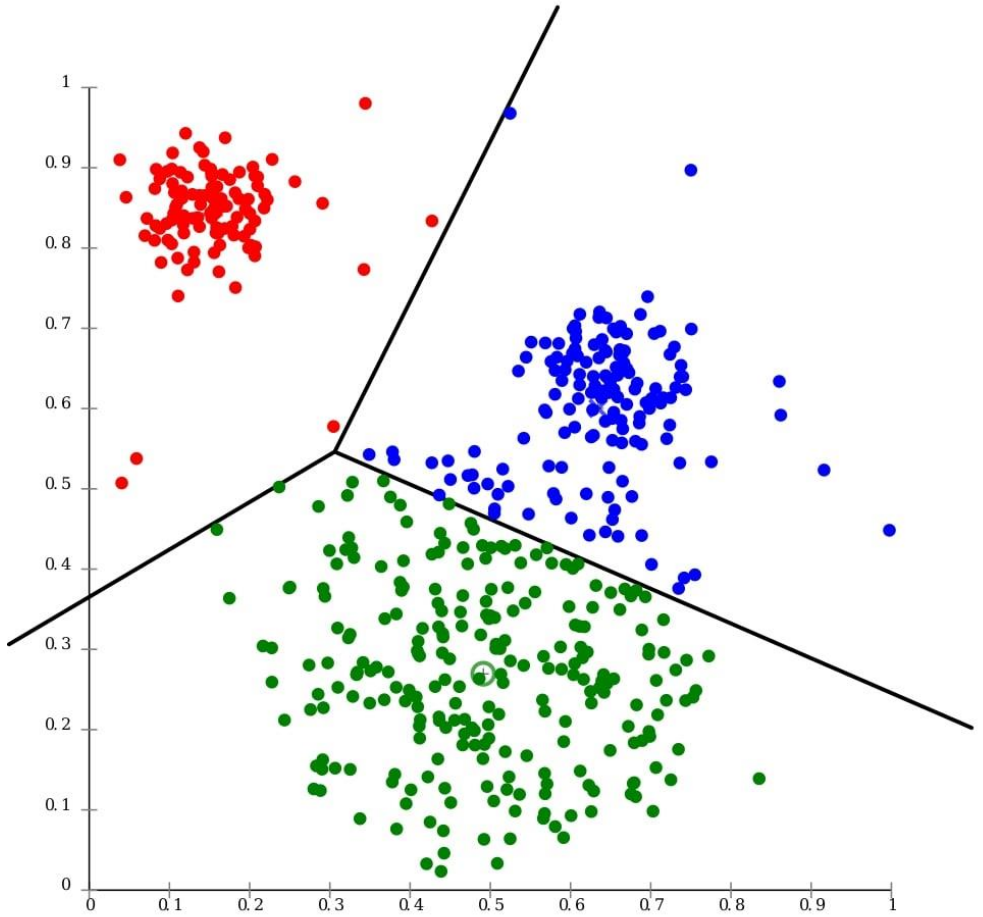
Tran Van Loc

Data Mining

- Term encompassing tools for discovering patterns in large datasets
- Data mining vs. regressions?
  - Representation
    - Risk = 0.93*prior_default + 0.23*nun_cards – 1.3*employed – 0.734
- How did the technique find the relevant variables and coefficients?
  - Regression: analyst had a hypothesis
  - Data mining: data-driven exploration
- Data mining techniques/tools
  - Clustering
  - Association rule mining

Clustering

- Clustering: Grouping data such that intra-group similarity is maximized while inter-group similarity is minimized
- Application: Data-driven customer segmentation



Association Rule Mining

Tìm sự xuất hiện phổ biến trong dữ liệu

- Association rule mining - finding common co-occurrences in data
- E
  - {bread, butter} ➜ {milk}

Chẩn đoán y khoa: Dầu cá và Raynaud's bệnh

- Medical diagnosis: Fish oil and Raynaud's disease

•"Độ nhớt máu tăng cục bộ trong quá trình Hiện tượng Raynaud do lạnh gây ra"

  - "Local increase of blood viscosity during cold-induced Raynaud's phenomenon"
  - "Reduction in blood viscosity by eicosapentaenoic acid"

Giảm độ nhớt của máu bằng cách axit eicosapentaenoic

Từ mô tả đến dự đoán

From Description to Prediction

(bao gồm phân cụm và khai thác quy tắc kết hợp) cuối cùng về việc khám phá và mô tả những mẫu tồn tại trong dữ liệ

- Data mining (including clustering and association rule mining) is ultimately about discovering & describing what patterns exist in data

- The next step, and the core business opportunity, involves using data to make predictions about the future, called predictive analytics
  - Can we predict demand?
  - Can we predict fraud?

Bước tiếp theo và cơ hội kinh doanh cốt lõi liên quan đến việc sử dụng dữ liệu để đưa ra dự đoán về tương lai, được gọi là phân tích dự đoán

- Chúng ta có thể dự đoán nhu cầu không?
- Chúng ta có thể dự đoán gian lận không?

Predictive Analytics: Amazon Recommendations

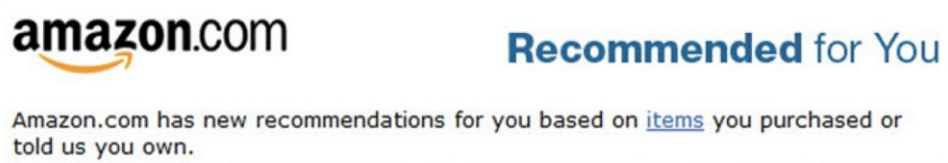
Visit website

Shop and place in cart

Pay for items

Items shipped

Amazon's Recommendations

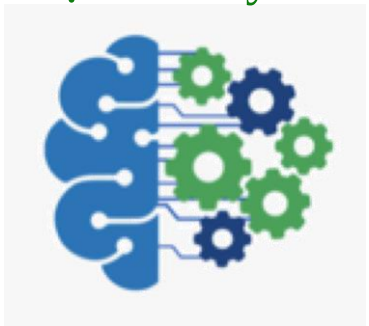


Phân tích dự đoán: Thanh toán Amazon

Predictive Analytics: Amazon Payments

Thuật toán dự đoán gian lận đã xảy ra

Phương thức thanh toán vật bị loại bỏ



Items purchased*

Algorithm predicts fraud has occurred

Payment method rejected

*Dữ liệu trước đây về hành vi của khách hàng trên trang web, cùng với biết được giao dịch mua hàng nào trước đây là gian lận, giúp đào tạo hệ thống để đưa ra dự đoán gian lận theo thời gian thực dựa trên hành vi của khách hàng (điều này sẽ được thảo luận thêm trong Mô-đun 3)

*Past data about customer behavior on the site, coupled with knowledge of which previous purchases were fraudulent, helps train the system to make real-time fraud predictions based on customer behavior (this will be discussed further in Module 3)

Ý nghĩa của dữ liệu lớn và phân tích dự đoán

Implications of Big Data & Predictive Analytics

Dữ liệu lớn và phân tích dự đoán mang lại lợi ích trực tiếp cho Amazon

- Big data & predictive analytics have direct benefits for Amazon

đề xuất hữu ích -> nhiều người mua sắm và mua hàng hơn

  - Helpful recommendations → more shoppers and purchases
  - Accurate fraud predictions → fewer losses & happier customers

Có lẽ quan trọng nhất, dự đoán có thể là một vòng tuần hoàn tốt

- Perhaps most importantly, prediction can be a virtuous cycle
  - Better predictions → more shoppers → more data → better predictions

University Of Danang
University Of Economics
THE FACUTY OF ELECTRONIC COMMERCE

AI in Business Introduction

Introduction to Artificial Intelligence

Tran Van Loc

Agenda

- Artificial Intelligence (AI) Overview
  - What AI is, types of AI, history of AI and the expert systems approach
- Machine Learning (ML) Overview
  - How ML differs from AI and the three types of ML (supervised, unsupervised, and reinforcement learning)
- Detailed View of ML
  - ML at 30,000 feet and factors that influence accuracy in ML
- Specific ML Methods: A Deep Dive
  - Logistic regression, decision trees & random forests, and neural networks

What is Artificial Intelligence (AI)?

Definition of AI

- “The theory and development of computer systems able to perform tasks that normally require human intelligence” - Merriam Webster
- Understanding language, reasoning, speech recognition, decision-making, navigating the visual world, manipulating physical objects, etc.

At its core, AI is about getting computers to do things that require human intelligence.

source: https://www.securitymagazine.com/articles/90871-whats-the-real-role-of-ai-and-ml-in-cybersecurity

Intelligence on a Spectrum

AI can refer to many different types of intelligence:

Artificial Narrow Intelligence (weak AI):

- AI that is very good at one specific task
  - A chess algorithm that can beat any human
  - Amazon's book recommendations

Artificial General Intelligence (strong AI):

- A computer program that could do all intelligent things a human could do, just as quickly and easily
  - Artificial Neural Networks are the closest thing to reaching this level

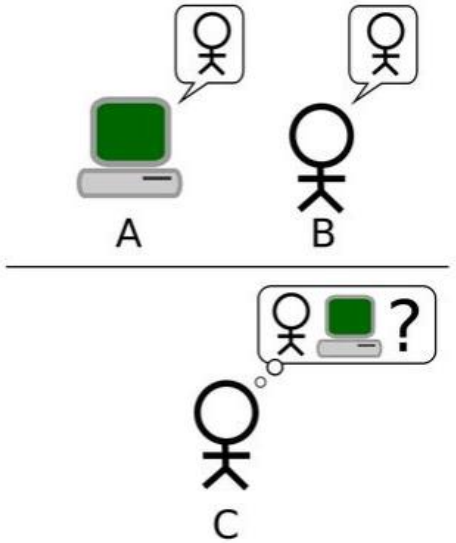
Artificial Super Intelligence:

- A computer program that can rapidly improve itself and do all things a human could do at a significant increase in speed and competency

History of AI

Origins

- Alan Turing, Cryptography
  - Can machines think?
  - “My contention is that machines can be constructed which will simulate the behavior of the human mind very closely.”
  - Proposed the Turing test for machine intelligence (1950) - Can machines do well in imitation games?
- Soon after, a workshop called the Dartmouth Summer Research Project on Artificial Intelligence was held, which historians believe is what coined the term AI



source: “A Human’s Guide to Machine Intelligence” by Kartik Hosanagar

> “Calling it AI made it extremely ambitious, and it inspired many people to enter the field, which has been responsible for a lot of the progress.”
>
> **— PEDRO DOMINGOS**

AI in the Press

Much of what the general public knows about AI is about AI playing popular games, which has been widely covered in the press

The New York Times

Computer Wins on 'Jeopardy!': Trivial, It's Not

By JOHN MARKOFF FEB. 16, 2011

Progression of AI in Games

IBM's Deep Blue

- May 11, 1997 - IBM's Chess Playing computer beat #1 in the world, Garry Kasparov
- No ML in system - Ability to analyze > 200,000 moves per second

IBM's Watson

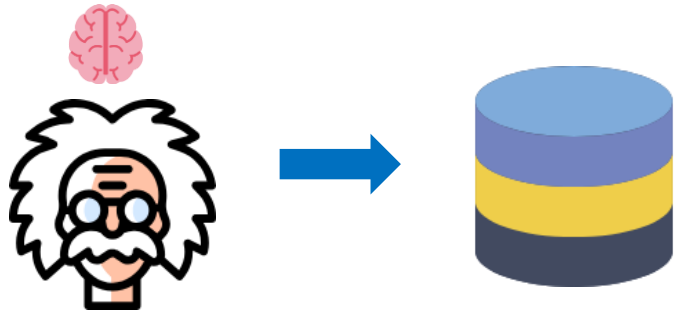
- 2011 - IBM Watson beats Ken Jennings and Brad Rutter (two of Jeopardy's best)
- Using Natural Language Processing (NLP) and Question Answer (QA) information retrieval

Google + DeepMind's AlphaGo

- Wins 4:1 vs. World Champion Lee Sedol in GO
- Techniques - Machine Learning (Neural Nets & Reinforcement Learning)

How to Build AI: Expert Systems Approach

- Knowledge engineering, or expert systems
  - Capturing and transferring knowledge from experts to a computer system



- Examples:
  - Software to diagnose diseases - interview doctors and codify the rules they use to diagnose diseases
  - Software for self-driving cars - interview drivers and codify the rules they use to drive cars
- Can create reasonably intelligent systems, but over time expert systems are often not able to beat human beings at complex tasks that require intelligence

How to Build AI: Expert Systems Approach (cont.)

Limitations of expert systems

- Polanyi's Paradox - Tacit knowledge
  - Idea that we have a lot of knowledge that we are not aware of
- Asking people to give us all the knowledge they have gets us a good amount of information, but because of tacit knowledge it doesn't give us *all* the information needed
  - Doctors arrive at diagnoses in seconds based off pattern matching, not rules
  - Drivers cannot articulate all the rules they use to drive cars
- The limitations of the expert systems approach has led to the emergence of a newer approach known as machine learning (ML)

University Of Danang
University Of Economics
THE FACUTY OF ELECTRONIC COMMERCE

AI in Business Introduction

Machine Learning Overview

Tran Van Loc

What is Machine Learning?

Machine learning (ML) is a subfield of AI

- ML methods are characterized by their ability to learn from data without being explicitly programmed
- ML is often used for making predictions

PREDICTION TASKS

Structured Data: “Is a transaction fraudulent?”

Text: “Is an email spam?”

Images: Image recognition in driverless cars

Audio: Speech recognition

APPLICATIONS OF ML

Healthcare: Automated medical diagnosis

HR: Which applicants are best suited for a job

Tech: Voice interfaces, Autonomous cars, personalization

Finance: Investing

Three Types of Machine Learning

Supervised Learning

Develop predictive model based on input and output data

- Classification
- Regression

Unsupervised Learning

Group and interpret observations based only on input data

- Anomaly Detection
- Clustering

Reinforcement Learning

Acquire new data by taking actions and receiving ad hoc feedback

- Bandit Algorithms
- Q-learning

Supervised Learning

- Idea: Learning from data where you have cleanly labeled outputs and inputs
  - Inputs can be referred to as features/covariates
  - Outputs are often called targets (this is what we are trying to predict)
- Example - Email data
  - The output we are trying to predict is whether an email is spam or not
  - The inputs, or features/covariates, are the actual text in the email
  - Supervised learning is used when there is cleanly labeled past data, which have a “correct” answer
    - Certain data have been labeled as “spam”
    - Certain other data have been labeled as “not spam”

Supervised Learning Example

Concrete example: Tagging text content on Facebook

- 106,316 Facebook posts submitted by large brands; wanted to identify what types of posts were associated with highest engagement
- Manually tagging content is expensive, so we wanted to automate it with a supervised machine learning technique
- Solution - have a random sample tagged manually on Amazon Mechanical Turk, and use as a training dataset for supervised ML
  - Took a sample of 5,000 posts and had human beings label each of these posts
  - These labels were then used as a training data set for a supervised machine learning algorithm that learned what kinds of words are predictive of whether a post is emotional, or humorous, etc.
  - That algorithm was used to make predictions for the remaining posts that hadn't been labeled by a human being

NLP Algorithm Performance

	Accuracy	Precision	Recall
REMFACT	0.998	0.998	0.998
EMOTION	0.996	0.992	0.999
HUMOR	0.999	0.999	1
PHILANTHROPIC	0.999	0.999	1
FRIENDLIKELY	0.997	0.996	0.998
SMALLTALK	0.858	0.884	0.803
DEAL	0.996	0.999	0.994
PRICECOMPARE	0.999	0.999	1
TARGETING	0.999	0.998	1
PRODAVAILABILITY	0.999	0.998	1
PRODLOCATION	0.970	0.999	0.901

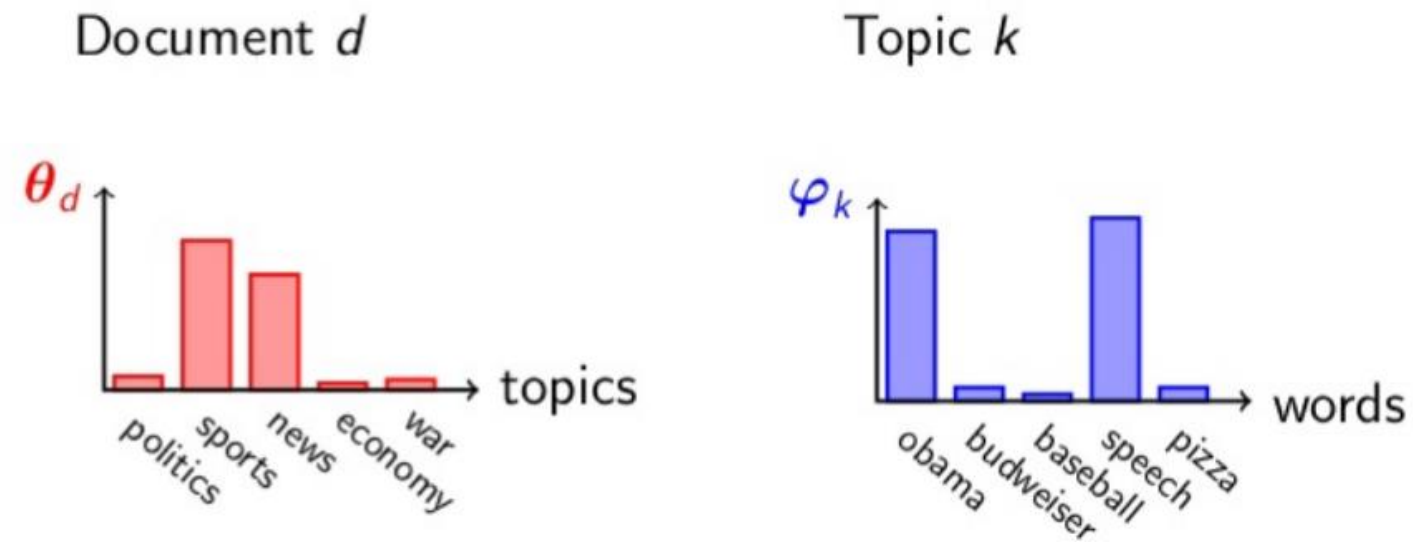
Detail at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2290802

Unsupervised Learning

- No fixed set of outputs provided; must be learned from inputs
- Goal is to cluster and identify important features

Unsupervised Learning

- Latent Dirichlet Allocation (LDA)
  - Takes as input a set of documents
  - Identifies both common topics across set of documents
  - Labels documents with corresponding topics



Unsupervised Learning

"Arts"	"Budgets"	"Children"	"Education"
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

The William Randolph Hearst Foundation will give $1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services," Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center's share will be $200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive $400,000 each. The Juilliard School, where music and the performing arts are taught, will get $250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual $100,000 donation, too.

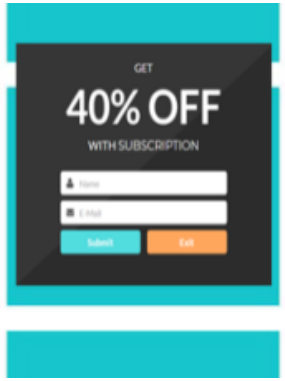
Reinforcement Learning

- Idea: Let algorithms learn by testing various actions/strategies and deciding which one works best
- Very powerful method for simulation and robotics-based applications
  - At the heart of AlphaGo and other state-of-the-art gaming AI
- In many applications experimentation isn’t free
  - You often want to balance both exploration and exploitation
  - Multi-armed bandit algorithms

Reinforcement Learning

- Suppose you have two ad copies and you don’t know which will attract more clicks (and therefore visitors to your website)
- Traditional A/B testing involves showing ad A 50% of the time and ad B 50% of the time, and then assessing which ad performed better

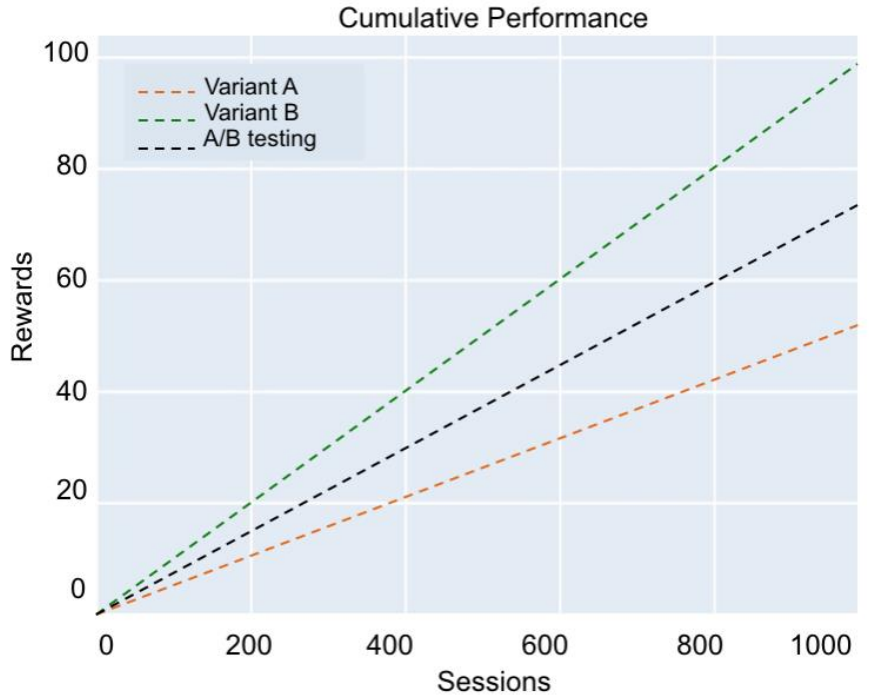
Ad A



Ad B

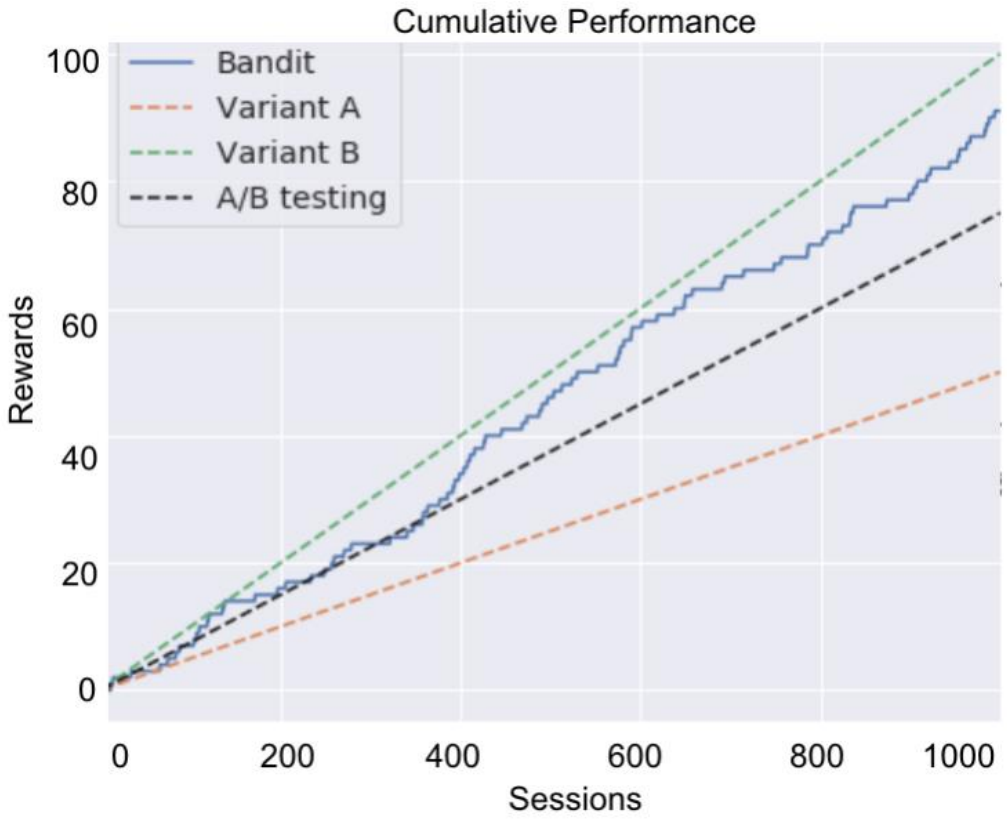
B performs better than A

The A/B test line is midway between the A and B lines as it is 50% A and 50% B



Reinforcement Learning

- Machine learning can improve upon A/B testing through bandit algorithms
- Bandit algorithms update beliefs based upon performance
  - They spend more time on best performers early on while still learning and improving over time
  - The bandit begins by showing 50% A & 50% B, but slowly starts allocating more and more traffic to the higher-performing ad as it learns and confirms which one is better



University Of Danang
University Of Economics
THE FACUTY OF ELECTRONIC COMMERCE

AI in Business

Reinforcement Learning

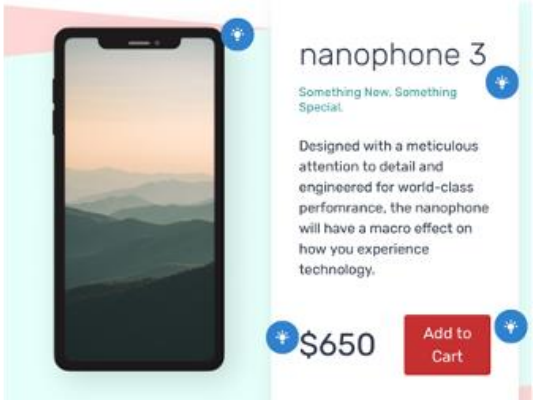
Tran Van Loc

Bandit Algorithms for Website Optimization

Website

nanophone

“Action” space



Objective

- Conversion rate
- Revenue per session

E-commerce retailer

Design of product page

Outcome of interest

Exploration and Exploitation

- Exploration - gathering more information about the decision environment
- Exploitation - making the best decision given the current information
- Example: Restaurant selection
  - Exploration - going to a new restaurant
  - Exploitation - going to your favorite restaurant
- How do we balance exploration vs. exploitation?
  - This tradeoff is handled by algorithms like multi-armed bandit, a classical reinforcement learning approach

Content/quotes from: https://www.davidsilver.uk/wp-content/uploads/2020/03/intro_RL.pdf

Multi-Armed Bandit

A finite set of resources must be allocated among multiple choices

- Gambler has finite amount of time
- Try new slot machines to try to maximize reward (exploration)
- Stick with the slot machine producing reasonable returns (exploitation)

Content/quotes from: https://www.amazon.com/RecZone-Jumbo-Slot-Machine-Bank/dp/B000H0KBDY

Bandit Algorithms

Many algorithms can be used to balance exploration and exploitation

- **Epsilon First** - Heuristic in which we experiment (explore) early, once we've learned, start exploiting
  - Allocate 100% of traffic to the best performing variant
- **Thompson Sampling** - Allocate traffic probabilistically to each arm, with higher probabilities given to arms that have a higher chance of being the best

Summary

- Reinforcement learning offers an alternative to ML based on having large datasets
- Reinforcement learning has found many applications in gaming and in online personalization
- Reinforcement learning is not as widely used as supervised machine learning

University Of Danang
University Of Economics
THE FACUTY OF ELECTRONIC COMMERCE

AI in Business

A Detailed View of Machine Learning

Tran Van Loc

ML

Focus on supervised learning:

- At its core, supervised machine learning is about using “a set of variables” to predict “an outcome”
- You do this all the time
  - Dark clouds + strong winds → rain

ML at

Focus on supervised learning:

- At its core, supervised machine learning is about using “a set of variables” to predict “an outcome”

1	3	19104	iOS
2	10	19141	Windows
…	…	…	…

Purchase
Yes
No
…

- Names for input data - predictors, features, “data”, variables, characteristics, covariates

ML at 30,000 Feet

Focus on supervised learning:

- At its core, supervised machine learning is about using “a set of variables” to predict “an outcome”
- You do this all the time
  - Dark clouds + strong winds ➜ rain
  - What someone is wearing, how they interact with you ➜ whether you’ll be friends
  - Resume (school, experience, skills) ➜ good employee
- Applications in business are abundant
  - Will someone buy your product?
  - Will someone click your ad?

ML at 30,000 Feet

Focus on supervised learning:

- At its core, supervised machine learning is about using “a set of variables” to predict “an outcome”

1	3	19104	iOS
2	10	19141	Windows
…	…	…	…

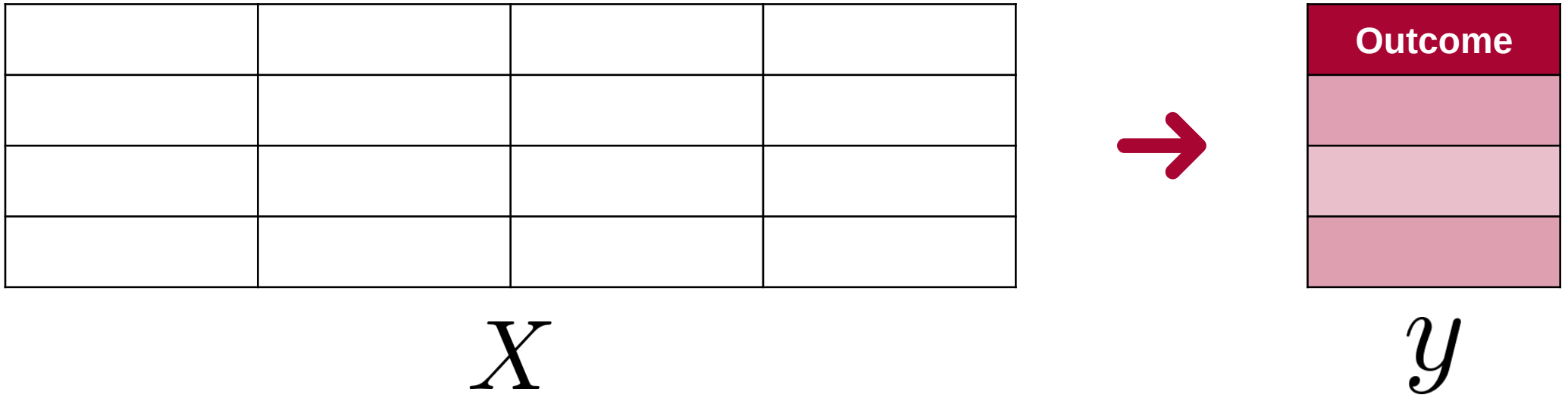
Purchase
Yes
No
…

- Names for input data - predictors, features, “data”, variables, characteristics, covariates

ML at 30,000 Feet

Focus on supervised learning:

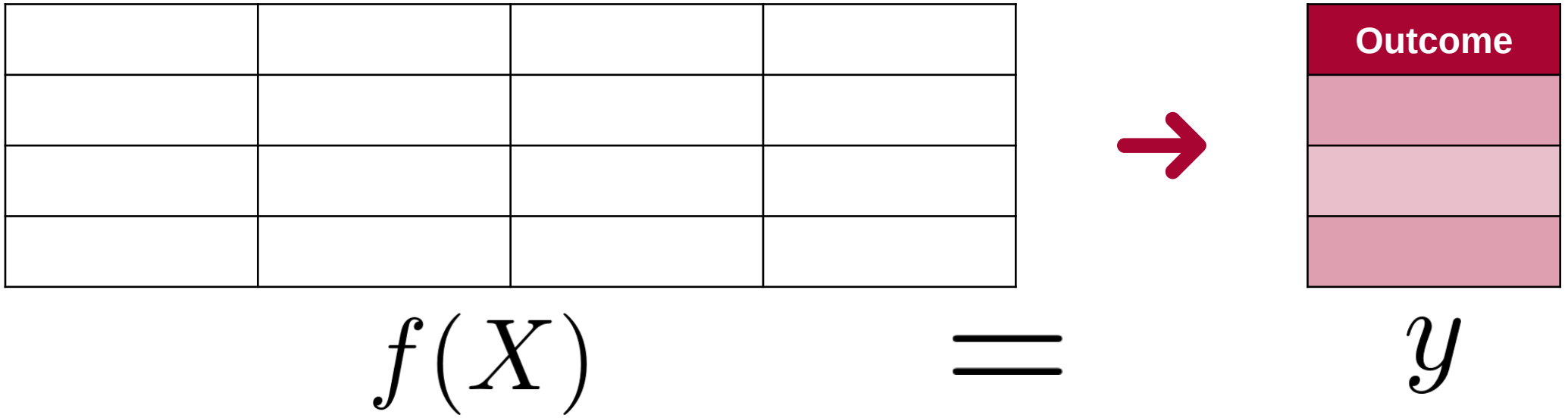
- At its core, supervised machine learning is about using “a set of variables” to predict “an outcome”



ML at 30,000 Feet

Focus on supervised learning:

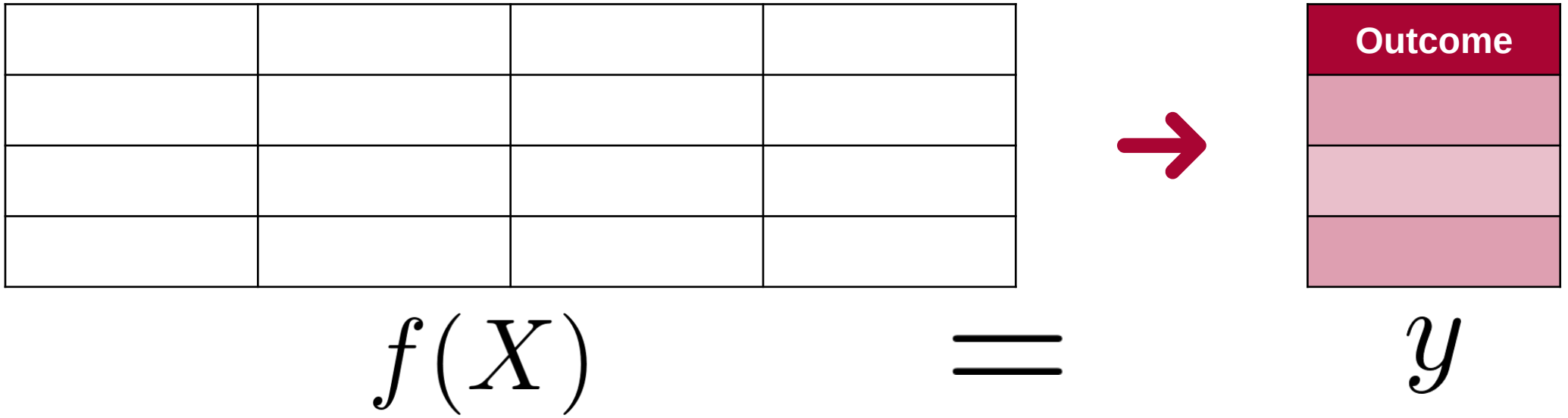
- At its core, supervised machine learning is about using “a set of variables” to predict “an outcome”



ML at 30,000 Feet

Focus on supervised learning:

- At its core, supervised machine learning is about using “a set of variables” to predict “an outcome”



All of supervised ML comes down to approximating this function with high fidelity

What Influences Accuracy in ML?

What is accuracy?

- Ability to make correct predictions on data you haven’t seen

Outcome

What Influences Accuracy in ML?

Quantity of data

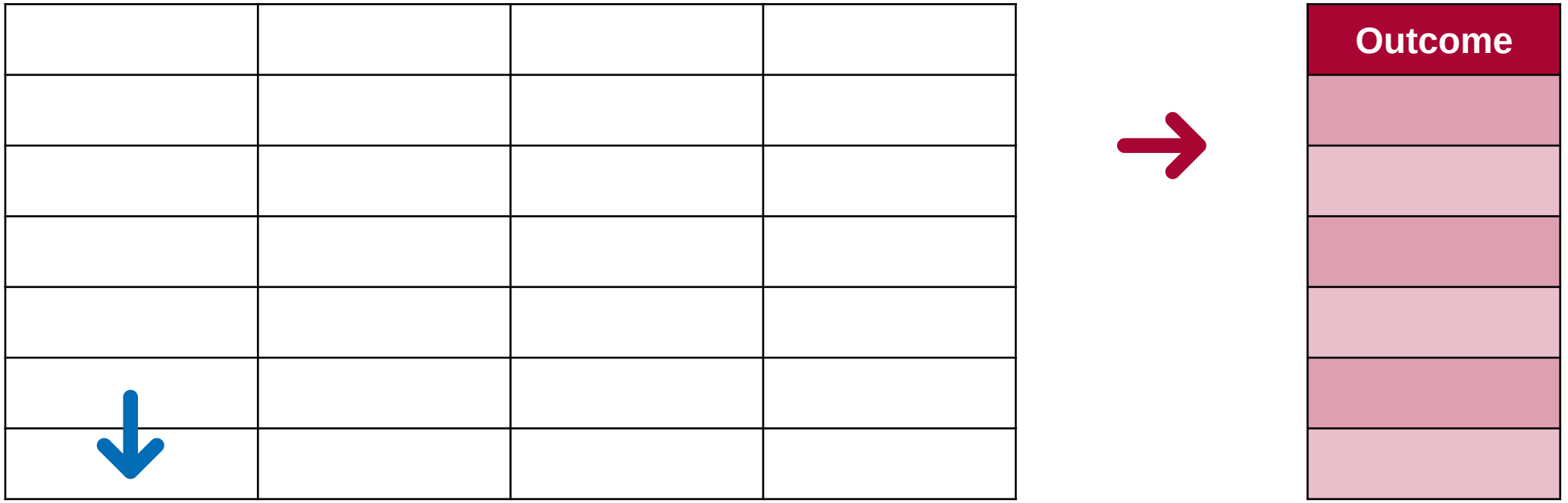
- Number of distinct observations

Outcome

What Influences Accuracy in ML?

Quantity of data

- Number of distinct observations



What Influences Accuracy in ML?

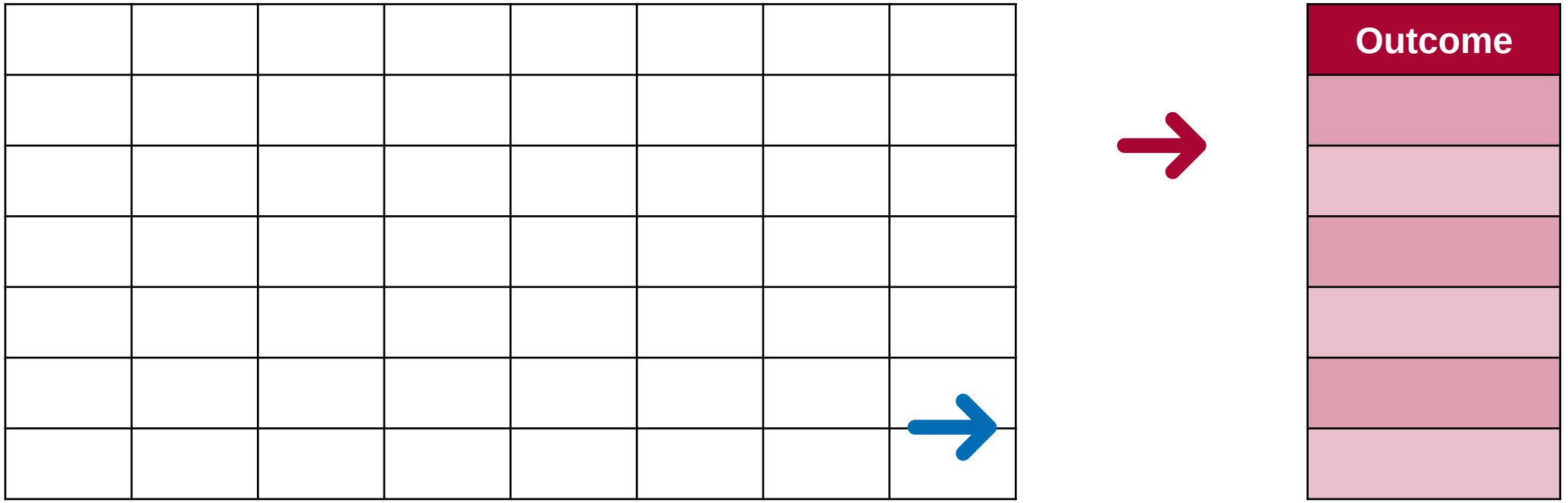
Quantity of data

- Number of distinct observations
- Different characteristics about your observations

What Influences Accuracy in ML?

Quantity of data

- Number of distinct observations
- Different characteristics about your observations



What Influences Accuracy in ML?

- Quantity of data (number of rows)
- Number of features (columns)

Lots of other things!

- Relevance of information to underlying phenomena
  - Umbrellas predict rain better than color of a person's dress
- Complexity of your model (how we approximate "*f*")
  - Modern developments in "deep learning" allow for very complex models that were historically very hard to train
- "Feature engineering"
  - Using domain knowledge to create new features from the input data (some of it can be automated but the analyst has a big role to play)