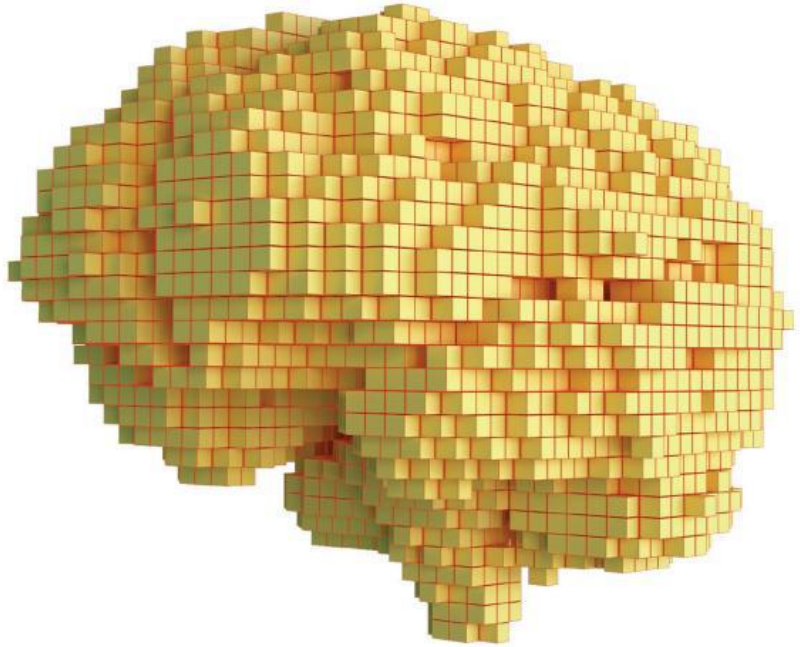
ICDS Innovation Center for Data Science

Meetup Mashhad Data Science

آزمایشگاه علم داده
دانشکده علوم ریاضی
دانشگاه فردوسی مشهد

دانشگاه فردوسی مشهد

دانشگاه فردوسی مشهد
دانشکده علوم ریاضی



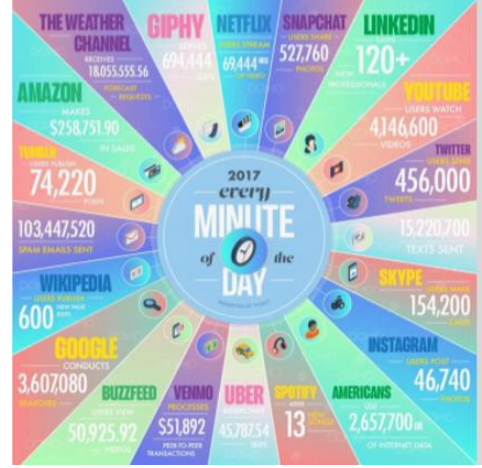
Becoming a DATA HEAD

looking at a business and scientific problems through the lens of data

Masoud Mazloom

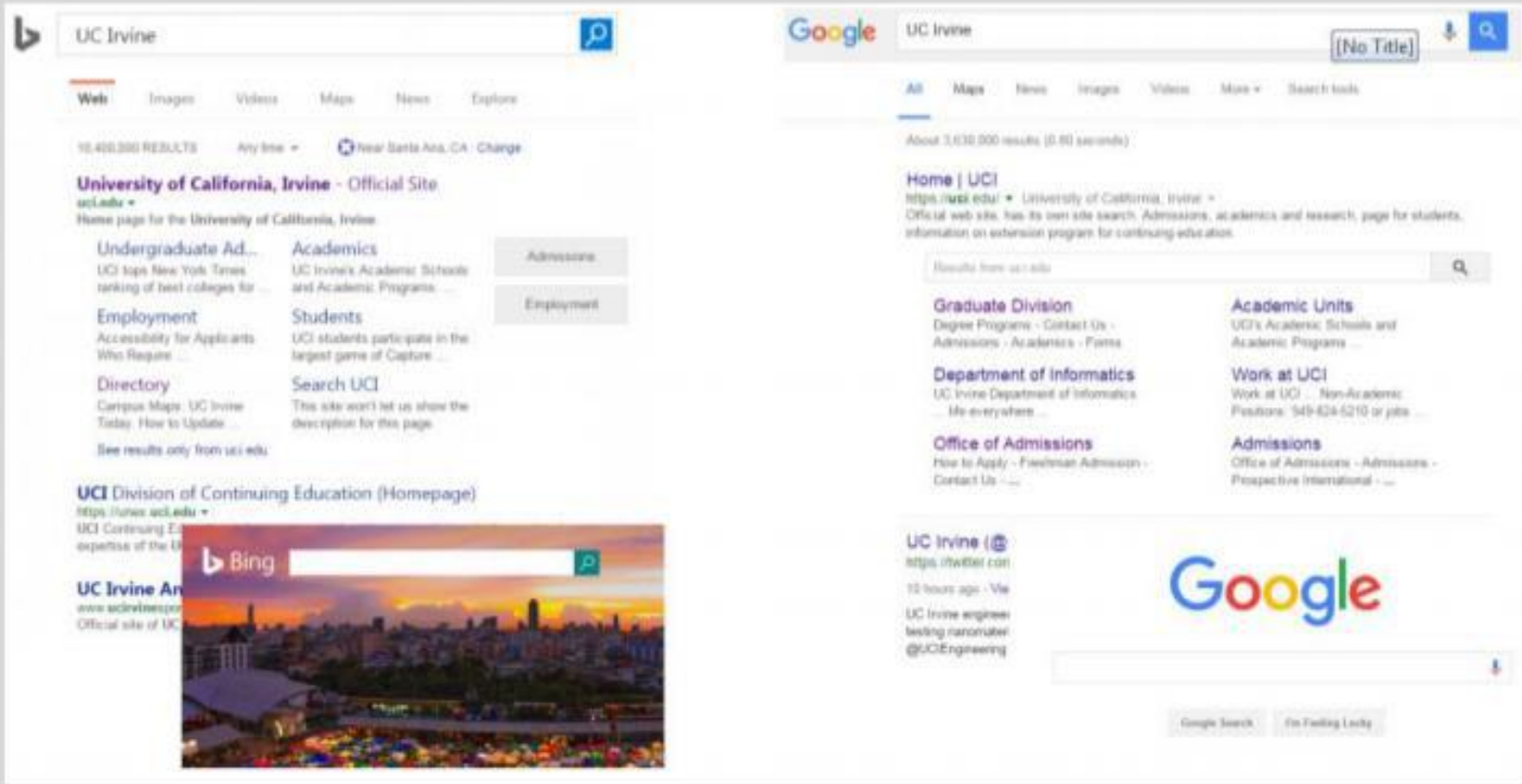
19.01.2022

Why DATA HEAD?

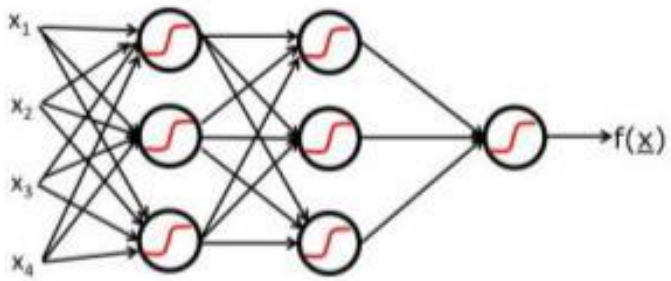


- Data is perhaps the single most important aspect to your job
- Without a doubt, we are in the age of data:
  - We create and consume more information than ever before
  - Big Data, AI, Data science, ML, and DNN everything is becoming data-field
  - Data is eating the world
  - Data has created an entire industry of promises, and products
  - Data is the core of science in the AI age
    - All science will become data science
- Becoming a Data Head is well-timed for the current state of data and analytics within organizations

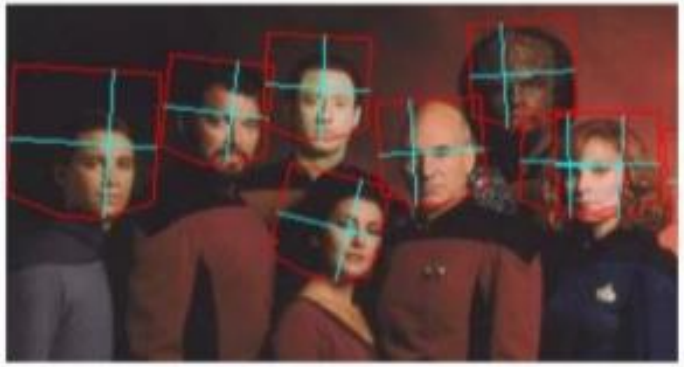
Web search: How do search engines like Google or Bing rank search results?



Social media: How does Facebook recognize people in images?



facebook



From Le Cun and Ranzato 2013

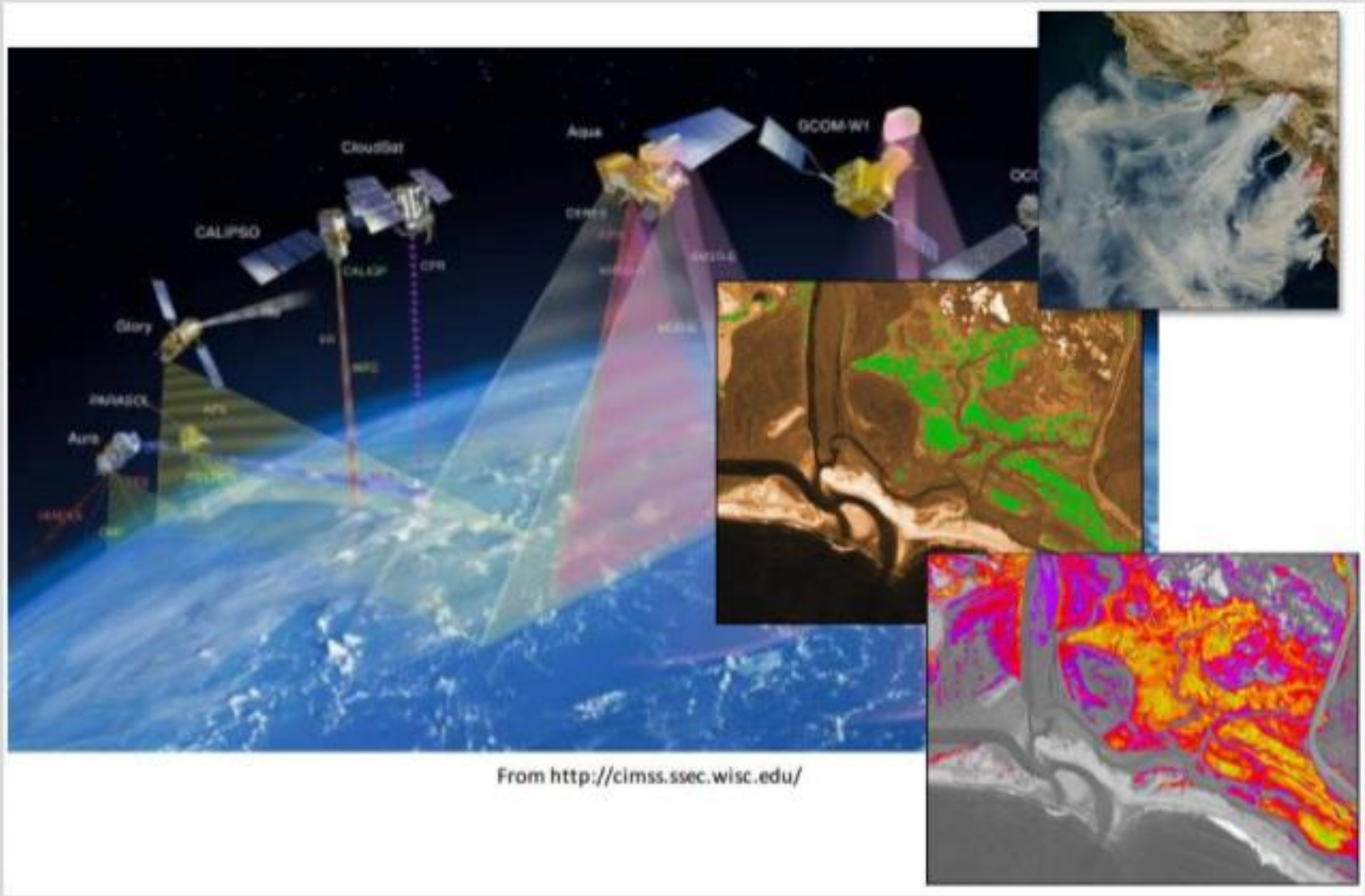
Shopping: How does Amazon forecast how many items it needs to store in its warehouses?

From dailymail.co.uk

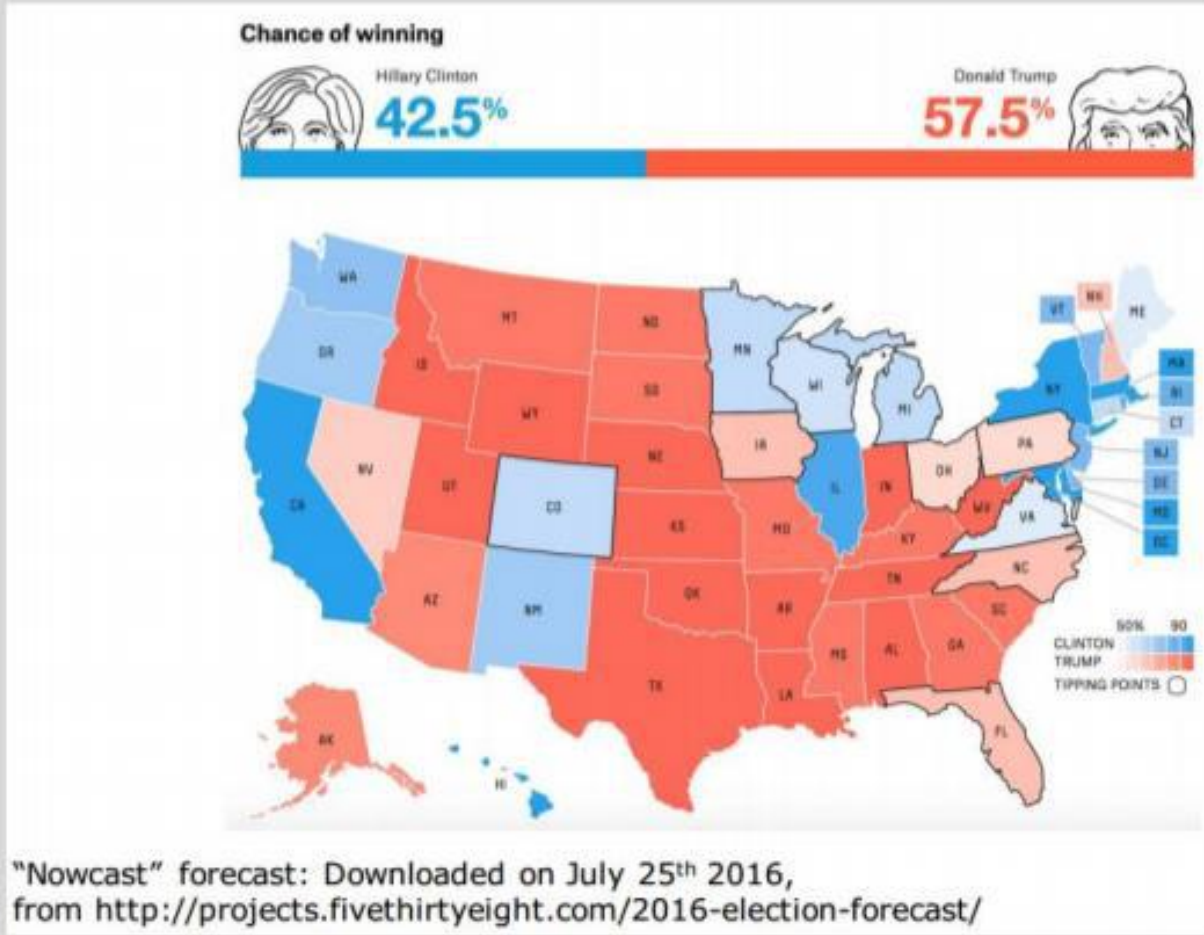
From www.formaspace.com

From linkedin.com

Climate: How can NASA automatically detect land changes using satellite image data?

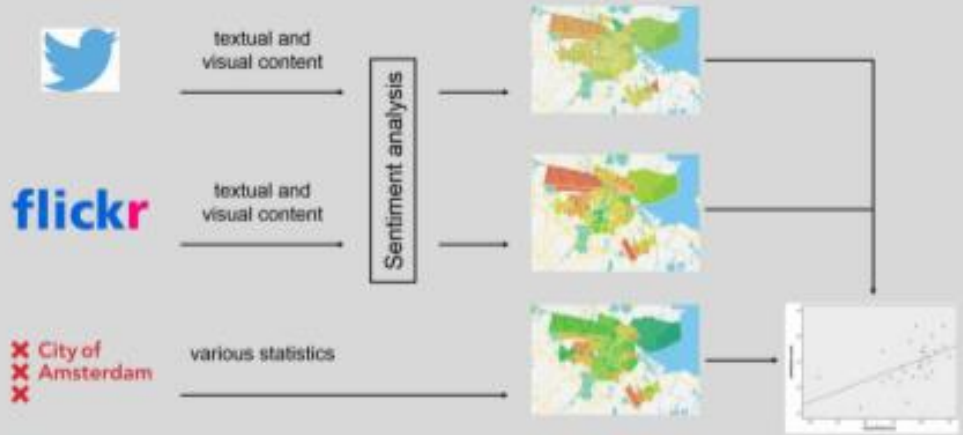


Politics: How can we reliably predict events like elections?

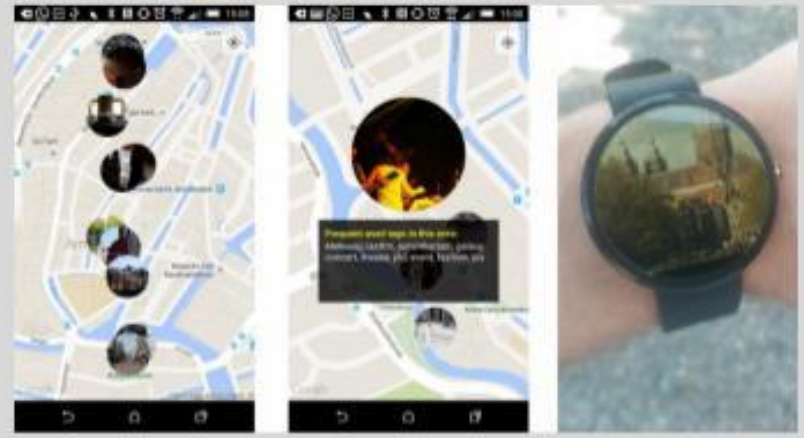


"Nowcast" forecast: Downloaded on July 25th 2016,
from http://projects.fivethirtyeight.com/2016-election-forecast/

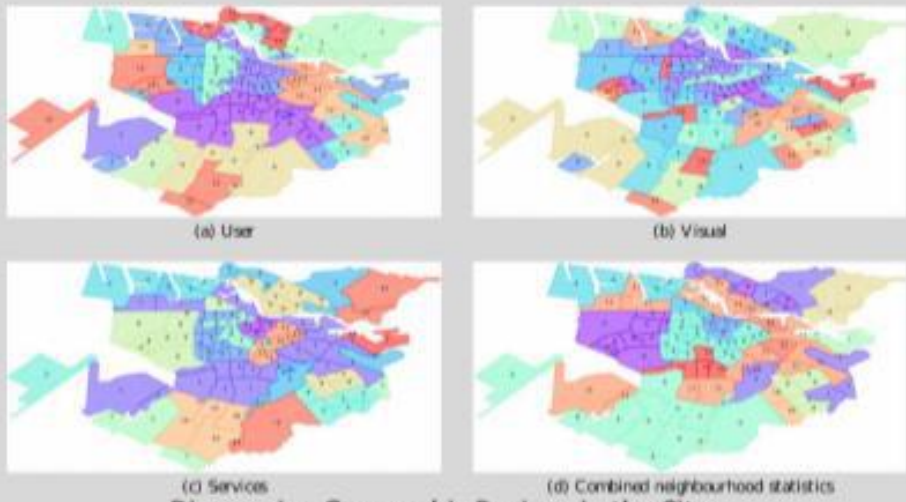
Smart city: How can we process multimedia data for having a smart city?



City Liveability



summarization of tourist routes



Discovering Geographic Regions in the City

How?

- All of these applications use **Data Science**
- The applications are built on combinations of ideas from:
  - Database systems
  - Algorithms
  - Machine learnings
  - Probabilistic models
  - Statistics
  - Data Visualization
  - And more…

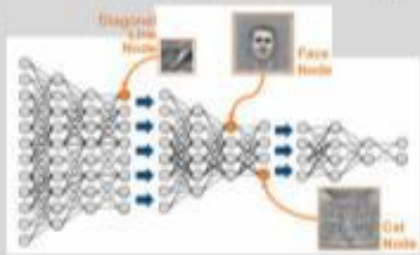
The success of AI

Moore's Law

Doubling compute power every 18 months

Big Data

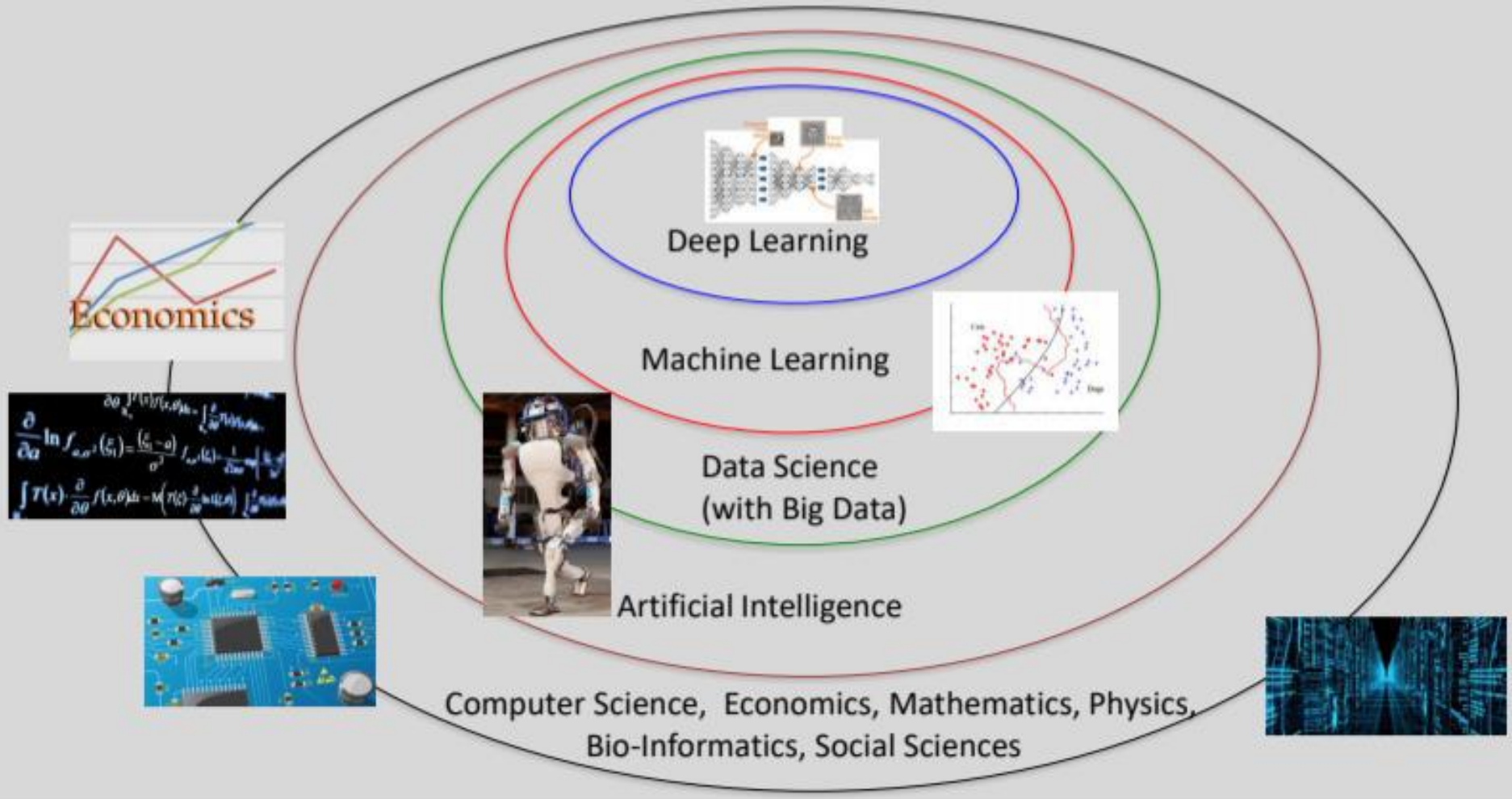
Deep Learning



Big Models: billions of parameters

[Slide credit by Max Welling]

The Second Wave of AI: Data Science



You'll become a Data Head

- Construct a mental model of data science, statistics, and machine learning

- What is a mental model?
  - It's "a simplified representation of the most important parts of some problem domain that is good enough to enable problem solving."
  - Think of it as a new storage room in your brain where you can put information

- Constructed mental model, you will learn how to ***think, speak, and understand*** data

You will be able to:

- **Think** statistically and understand the role variation plays in your life and decision making
- **Speak** intelligently and ask the right questions about the statistics and results
- **Understand** what's really going on with ML, text analytics, DL, and AI
- **Avoid common pitfalls** when working with and interpreting data

think critically about the data and results you come across

speak intelligently about all things data

understand data and its challenges at a deeper level

Assumptions about data and analytics

- Assumption 1: Analytics, big data, and AI are wholly different phenomena:
  - Becoming a Data Head, however, correctly adopts the view that they are highly interrelated

- Assumption 2: Data scientists are the only people who can play in this sandbox:
  - Automated ML tools make it easier to create models that do an excellent job of predicting

- Assumption 3: Data scientists are “unicorns” who have all the skills:
  - Data Heads understand:
    - **Rudiments of data science**
    - Know the **business and ask good questions**
    - **Can manage projects effectively**
    - **Are excellent at building business relationships and** increase the likelihood that **data science projects will lead to business value**

Forget these false assumptions

- Data science isn’t the magic unicorn that will solve all of your company’s problems
- Despite the claims and proliferation of data promises and products, data science projects are failing at alarming rates (“78% of data science projects failing”)
- Highlighting when data science is (and isn’t) the right approach

The 2016 United States General Election

- On November 8, 2016, Donald J. Trump, won the general election of the US
    - For the political pollsters this came as a shock. Their models hadn't predicted his win
    - In 2008, Nate Silver had done a fantastic job predicting Barack Obama's win
    - In 2012, once again, Nate Silver was predicting another win for Barack Obama
    - This successful prediction only reinforced the importance of forecasting with data
    - Articles warned executives to adopt data or be swallowed by a data-driven competitor
    - By 2016, news had invested in a prediction algorithm to forecast the election outcome
    - The vast majority of them suggested an overwhelming victory for Hillary Clinton

Pollsters wrongly assuming independence between variables

What happened?

- Why do data problems like this occur?
  - Hard problems:
    - Many data problems are fundamentally difficult
    - Even with lots of data, the right tools and techniques, and the smartest analysts, **mistakes** happen
  - Lack of critical thinking:
    - Some analysts and stakeholders **stopped thinking critically** about data problems
    - The Data Science Industrial painted a picture of **certainty** and **simplicity**
  - Poor communication between data scientists and decision makers:
    - Even with the best intentions, results are often lost in **translation**
    - Decision makers don't speak the language because no one bothered to teach data literacy
    - And, frankly, data workers **don't always explain things well**
    - There's a communication gap

Example

Us: "Based on our supervised learning analysis of the binary response variable using multiple logistic regression, we found an out-of-sample performance of 0.76 specificity and several statistically significant independent variables using alpha equal to 0.05."

*Business Professional: *awkward silence**

Us: "Does that make sense?"

*Business Professional: *more silence**

Us: "Any questions?"

Business Professional: "No questions at the moment."

Business Professional's internal monologue: "What the hell are they talking about?"

- In hindsight, our presentations were too technical.
- We started practicing explaining complex statistical concepts to other audiences.
- Our audiences would listen to us. Audiences couldn't think critically about our work because they didn't understand what we said.
- Discovered a middle ground between data workers and business professionals where honest discussions about data can take place without being too technical or too simplified. It involves both sides ***thinking more critically about data problems, large or small***.

Understand the big picture

- To become better at ***understanding*** and ***working*** with data you will need to be open to learning seemingly complicated data concepts

- How to translate them to your audience of stakeholders

- Develop a mental framework of data science, statistics, and machine learning

Construct your mental model to succeed as a Data Head

- **Thinking Like a Data Head:**
  - Think critically and ask the right questions about the data projects your organization takes on; what data is and the right lingo to use; and, how to view the world through a statistical lens.
- **Speaking Like a Data Head:**
  - Data Heads are active participants in important data conversations. This part will teach you how to "argue" with data and what questions to ask to make sense of the statistics you encounter. You'll be exposed to basic statistics and probability concepts required to understand and challenge the results you see
- **Understanding the Data Scientist's Toolbox:**
  - Data Heads understand the fundamental concepts of how statistical and machine learning models work. You'll gain an intuitive understanding of unsupervised learning, regression, classification, text analytics, and deep learning.
- **Ensuring Success:**
  - Data Heads understand the common mistakes and traps when working with data. You'll learn about technical pitfalls that cause projects to fail, and you'll learn about the people and personalities involved in data projects.

Thinking Like a Data Head

- Companies rush to try the “next big thing” in data without ever pausing to ask the right business questions
- Learn basic data terminology
- Learn how to look at the world through a statistical lens
- “Thinking Like a Data Head,” prepares you for the road ahead and puts you in the right mindset to think about and understand data

- Main questions:
  - What Is the Problem?
  - What Is Data?
  - Prepare to Think Statistically

Thinking Like a Data Head cont.

- **What Is the Problem?** (the most important and sometimes hardest question)
    - Helpful questions a Data Head should ask before attacking a data problem:
        - Why Is This Problem Important?
        - Who Does This Problem Affect?
        - What If We Don't Have the Right Data?
        - When Is the Project Over?
        - What If We Don't Like the Results?
    - Understanding Why Data Projects Fail
    - Working on Problems That Matter

Not asking these questions leads to a project failure

Thinking Like a Data Head cont.

- **What Is Data?**
  - Data vs. Information
    - From the sensors on satellites to the neurons firing in our brains, information is continually created
    - Information is derived knowledge from activities: measuring a process, thinking about something
    - *data is encoded information*
  - Data Types
  - How Data Is Collected and Structured:
    - Observational vs. Experimental Data
    - Structured vs. Unstructured Data
  - Basic Summary Statistics

Thinking Like a Data Head cont.

- **Prepare to Think Statistically**
  - There Is Variation in All Things
  - Probabilities and Statistics
    - Probability vs. Intuition
    - Discovery with Statistics

- The stock market fluctuates daily
- political poll numbers change depending on the week
- gas prices move up and down
- your blood pressure spikes when the doctor is present but not the nurse
- A business's sales fluctuate daily, weekly, monthly, and yearly
- Customer satisfaction survey results can vary wildly one day to the next
- **If we accept the reality of variation in our lives, we don't need to explain every peak and valley**
- This is what businesses will attempt to do:
  - What was done differently the week of high sales? Let's repeat the good, reduce the bad, they'll say

"Statistical thinking is a different way of thinking that is part detective, skeptical, and involves alternate takes on a problem."
Frank Harrell, statistician and professor

Speaking Like a Data Head

- Ask intelligent questions about the data and analytics you encounter in the workplace
  - Argue with the Data
  - Explore the Data
  - Examine the Probabilities
  - Challenge the Statistics

speak intelligently about all things data

Speaking Like a Data Head cont.

- **Argue with the Data (**asking questions about the data used in a project**)**
  - What Would You Do?
    - Missing Data Disaster
  - Tell Me the Data Origin Story
    - Who Collected the Data?
    - How Was the Data Collected?
  - Is the Data Representative?
    - Is There Sampling Bias?
    - What Did You Do with Outliers?
  - What Data Am I Not Seeing?
    - How Did You Deal with Missing Values?
    - Can the Data Measure What You Want It to Measure?
  - Argue with Data of All Sizes

- **If the raw data is bad, no amount of data cleaning, statistical methodology, or machine learning can hide the stench**
- **Garbage in garbage out**

Speaking Like a Data Head cont.

- **Explore the Data**
  - Embracing the Exploratory Mindset
    - Questions to Guide You
      - Can the Data Answer the Question?
        - Set Expectations and Use Common Sense
        - Do the Values Make Intuitive Sense?
        - Watch Out: Outliers and Missing Values
      - Did You Discover Any Relationships?
        - Understanding Correlation
        - Watch Out: Misinterpreting Correlation
        - Watch Out: Correlation Does Not Imply Causation
      - Did You Find New Opportunities in the Data?

Speaking Like a Data Head cont.

- **Examine the Probabilities**
  - The Rules of the Game
    - Conditional Probability and Independent Events
    - The Probability of Multiple Events
      - Two Things That Happen Together
  - Be Careful Assuming Independence
  - All Probabilities Are Conditional
    - Don’t Swap Dependencies
    - Bayes’ Theorem
  - Ensure the Probabilities Have Meaning
    - Calibration
    - Rare Events Can, and Do, Happen

Speaking Like a Data Head cont.

- **Challenge the Statistics**
  - The Process of Statistical Inference
  - The Questions You Should Ask to Challenge the Statistics
    - What Is the Context for These Statistics?
    - What Is the Sample Size?
    - What Are You Testing?
    - What Is the Null Hypothesis?
    - What Is the Significance Level?
    - How Many Tests Are You Doing?
    - Can I See the Confidence Intervals?
    - Is This Practically Significant?

Understanding the Data Scientist's Toolbox

- Search for Hidden Groups
- Understand the Regression Model
- Understand the Classification Model
- Understand Text Analytics
- Conceptualize Deep Learning

understand data and its challenges at a deeper level

Ensuring Success

- Discover how to make the most out of your Data Head journey by learning from others' mistakes, both technical and human:
  - Watch Out for Pitfalls
  - Know the People and Personalities

Becoming a DATA HEAD

is a way to understand data-driven problem

And become a GREAT DATA SCIENTIST

DATA BRIDGE

A way of connecting science to business

Knowledge

Enthusiasm

Hope and desire

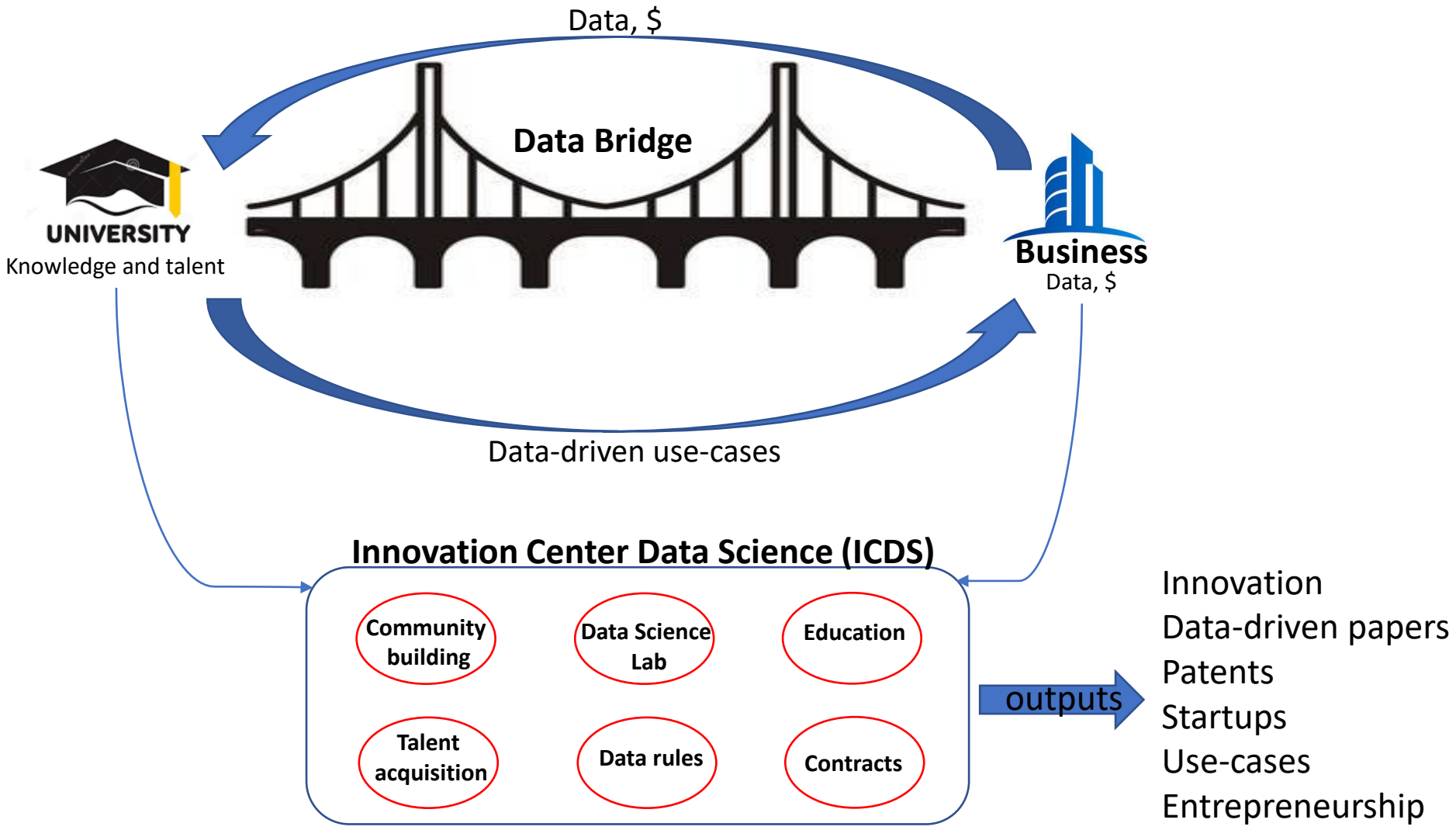
Empowerment

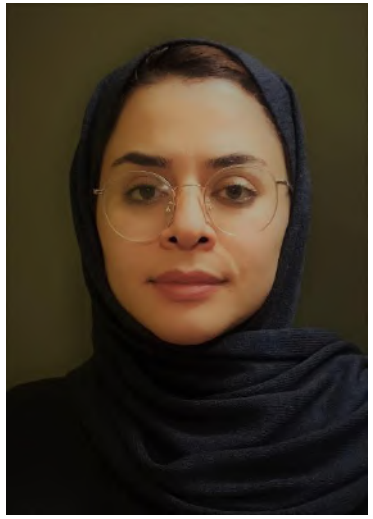
Motivation

Society

Business

What is and How to make a DATA Bridge?



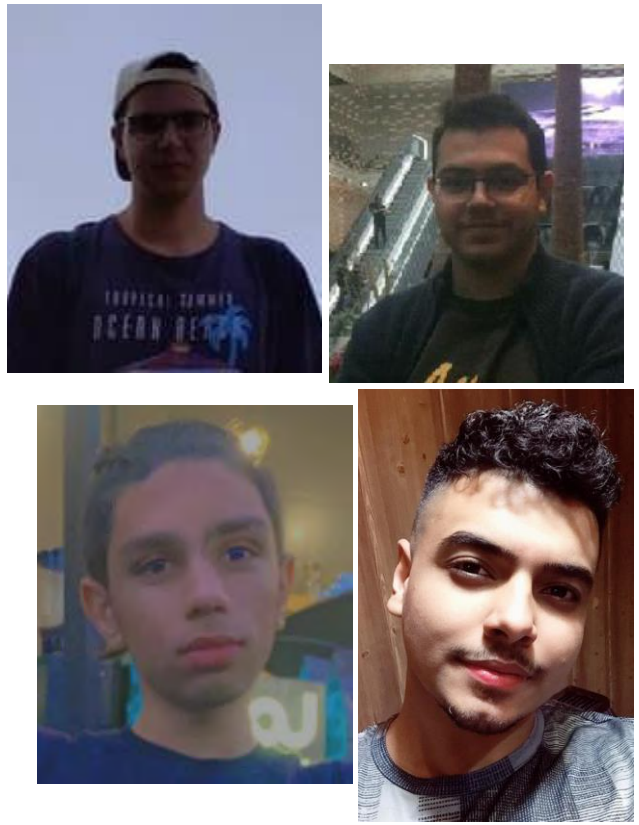


https://mashhad-icds.ir

https://twitter.com/mashhad_icds

https://www.instagram.com/mashhad_icds/

https://www.linkedin.com/company/mashhad-icds/



Email: ma.mazloom@um.ac.ir
masoud.mazloom@gmail.com