



UNIVERSITAS
INDONESIA

Veritas, Probatum, Sapientia

FACULTY OF
**COMPUTER
SCIENCE**



UNIVERSITAS
INDONESIA

Veritas, Probatum, Sapientia



pusilkom ui

DSNP DJPb Kementerian Keuangan RI

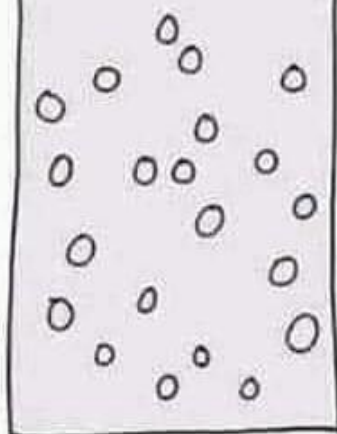
Intro to DS: Managing Projects

Instructor: Muhammad Hilman, Ph.D

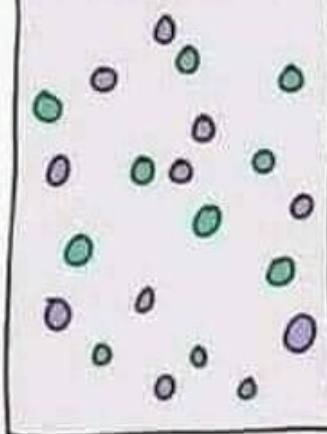
Slide by Fariz Darari, Ph.D.

WELCOME

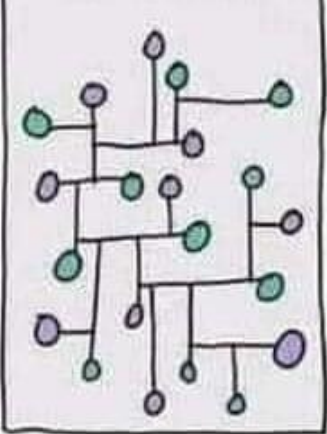
Data



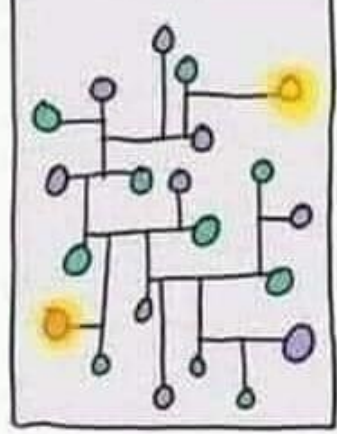
Information



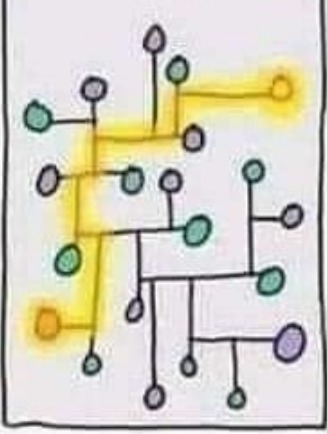
Knowledge



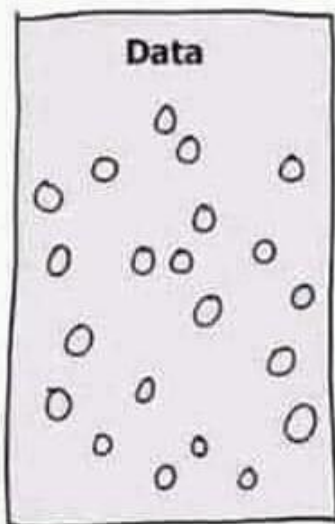
Insight



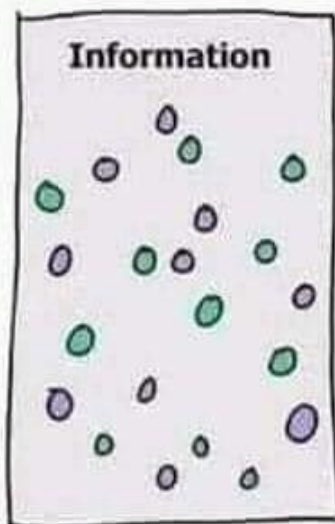
Wisdom



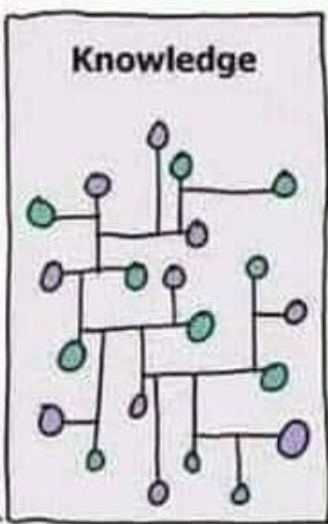
Data



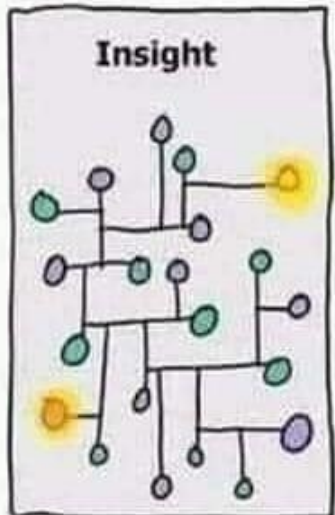
Information



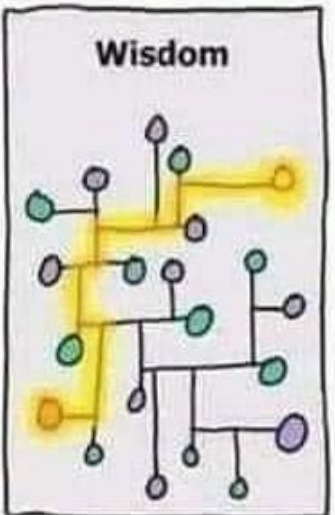
Knowledge



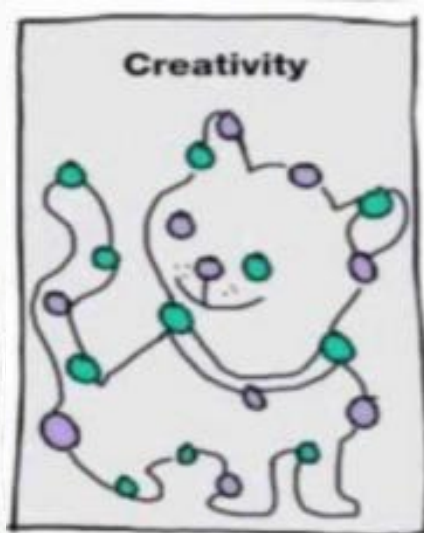
Insight

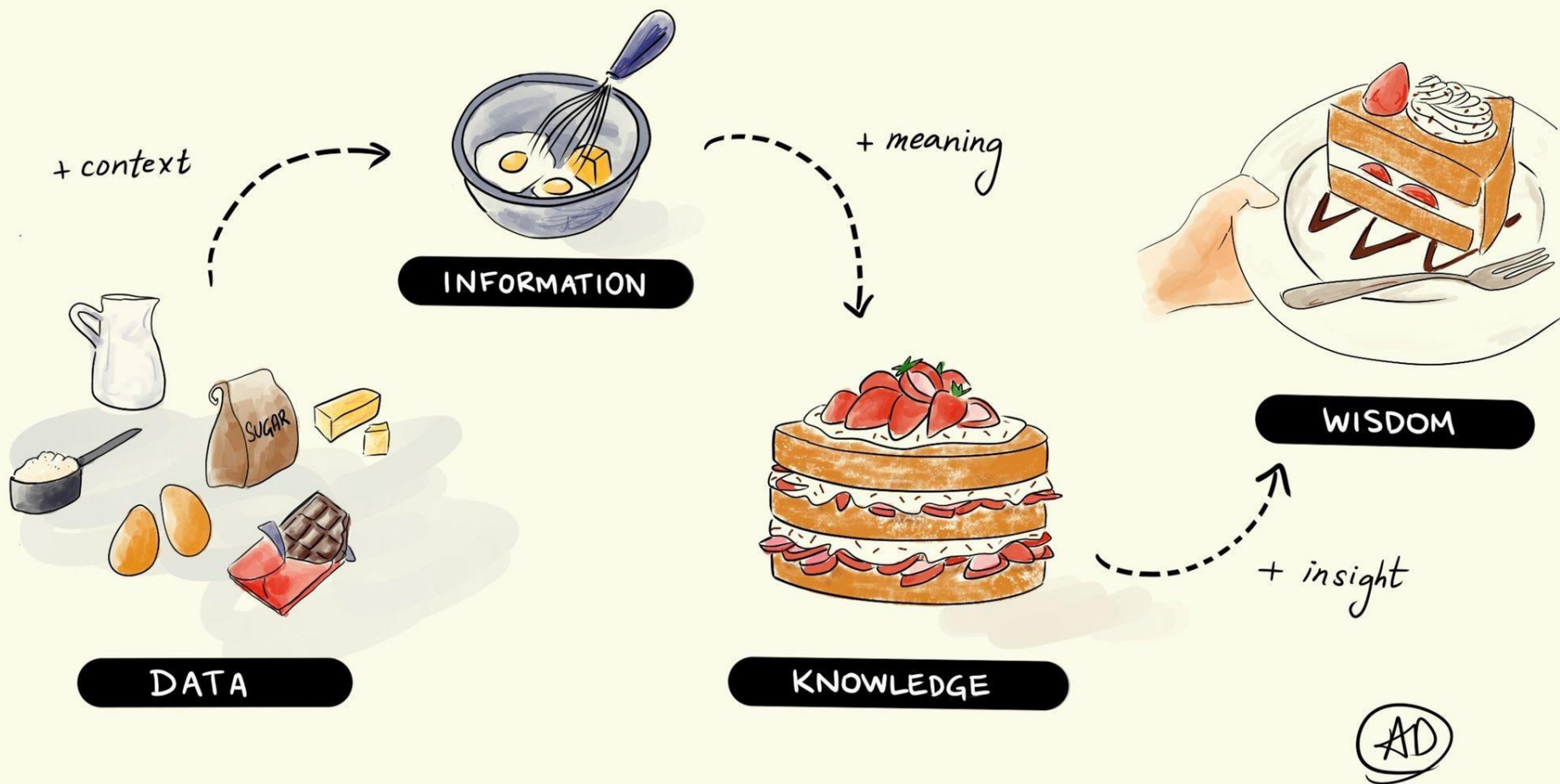


Wisdom



Creativity





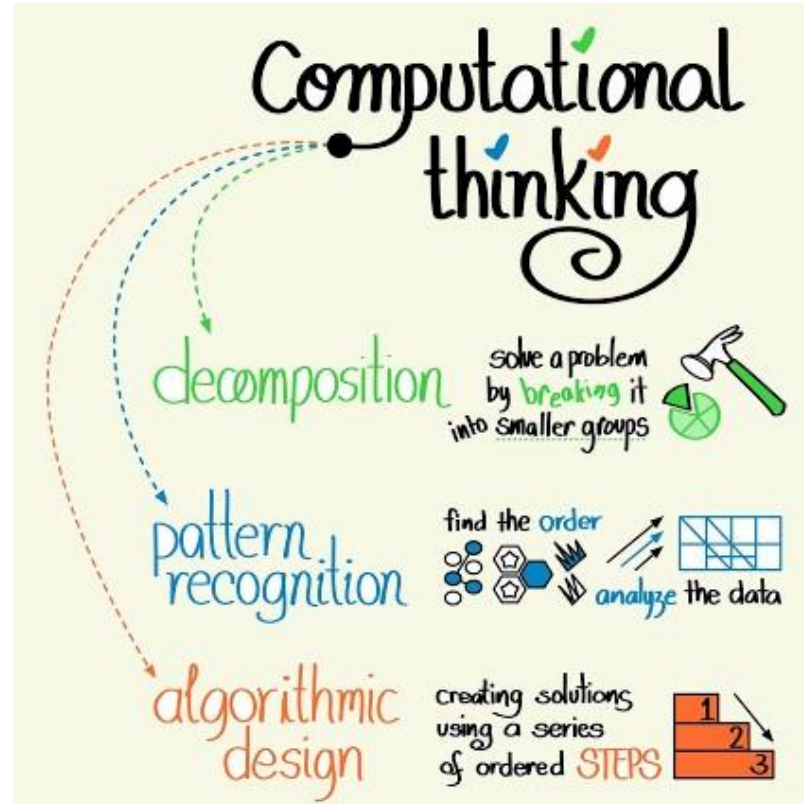
Data-driven Decision Making

A decision-making process which involves:

- **collecting** data
- **extracting patterns** and facts from that data
- **utilizing** those facts to make inferences that influence decision-making

It's making decisions based on hard data as opposed to intuition or guesswork

Computational Thinking



Decomposition



Breaking the problem into smaller, more manageable parts.

Pattern recognition



Recognising which parts are the same and the various attributes we can use to define them.

Algorithm design



Planning the step-by-step instructions that need to be carried out to achieve the goal.

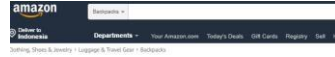
Data Science: Definitions

- Data science is a field that comprises **everything related to data**, that is, it is about how to: understand data, process data, extract value from data, visualize data, and communicate data. (UC Berkeley)

Data Science: Definitions

- Data science is a field that comprises **everything related to data**, that is, it is about how to: understand data, process data, extract value from data, visualize data, and communicate data. (UC Berkeley)
- Data science is a multidisciplinary approach to **extracting actionable insights** from the large and ever-increasing volumes of data collected and created by today's organizations. (IBM)

Data Science Showcase: Recommenders



Customers who viewed this item also viewed



KROSER Laptop Bag 15.6
Inch Briefcase Shoulder
Messenger Bag Water
Repellent Laptop Bag...
★★★★☆ 311
\$19.99



Laptop Case 15.6 inch,
Laptop Bag Briefcase for
Men Women, Slim
Business Portable...
★★★★☆ 12
\$19.99



KROSER Laptop Bag 15.6
Inch Laptop Briefcase
Laptop Messenger Bag
Water Repellent...
★★★★☆ 390
\$26.99



KROSER Laptop Bag 15.6
inch Briefcase Laptop
Messenger Bag Water
Repellent Computer...
★★★★☆ 146
\$19.99



KROSER Laptop Bag
Laptop Briefcase Fits Up to
16 Inch Laptop Water-
Repellent Light Weight...
★★★★☆ 103
\$21.99



Acer Wireless Optical
Mouse
★★★★☆ 168
\$12.00

Data Science Showcase: Sentiment Analysis

[uClassify](#) [Classifiers](#) [Translate](#) [Docs](#) [Pricing](#) [About](#)

Classify Text

Classify Url

Classify Text

Data science is super awesome!

Classify

👍 Success

Show [REST XML URL](#)

positive

99%

negative

<https://www.uclassify.com/browse/uclassify/sentiment?input=Text>

Data Science Showcase: Sentiment Analysis (cont.)

Classify Text

Classify Url

Classify Text

The food was great but I hate the waiters.

Classify

👍 Success Show REST XML URL

negative

60%

positive

40%

Data Science Showcase: Face Recognition



Data Science Showcase: Genuine/Fraud Transactions



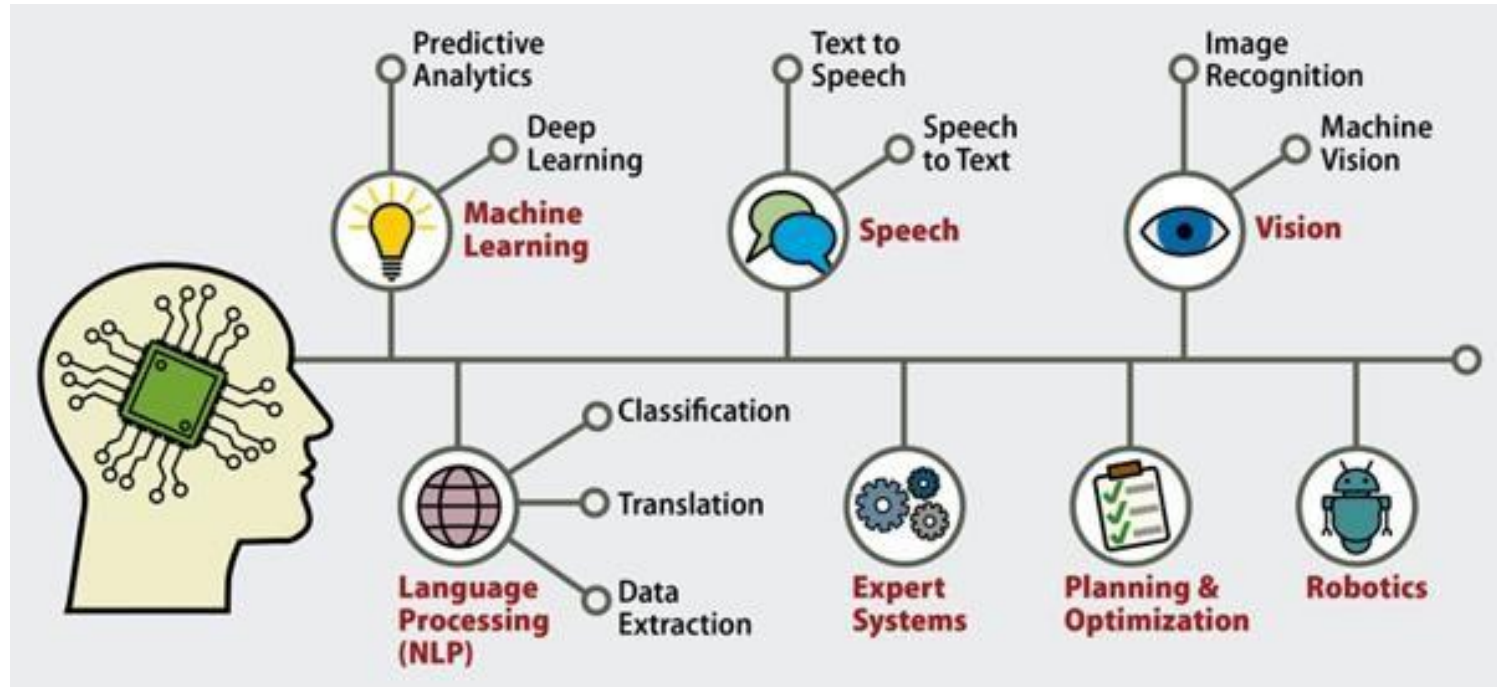
Data Science Showcase: Genuine/Fraud Transactions (cont.)



Data Science Showcase: Usage Pattern Analysis and Prediction



Data Science Areas



National AI Strategy of Indonesia



Sharing Time

Are there any data science applications you have encountered in life but not yet mentioned by the instructor?

Feel free to share your thoughts and experiences!

AI, ML, and DL

Artificial Intelligence

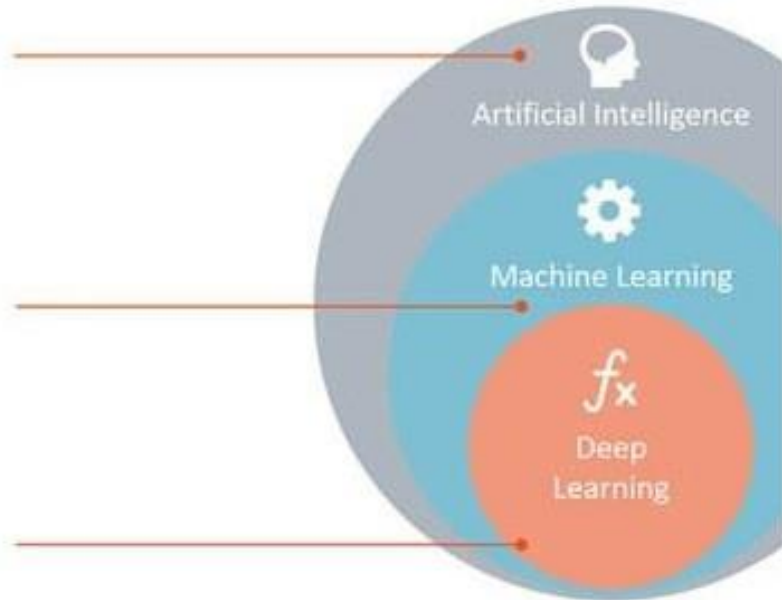
Any technique which enables computers to mimic human behavior.

Machine Learning

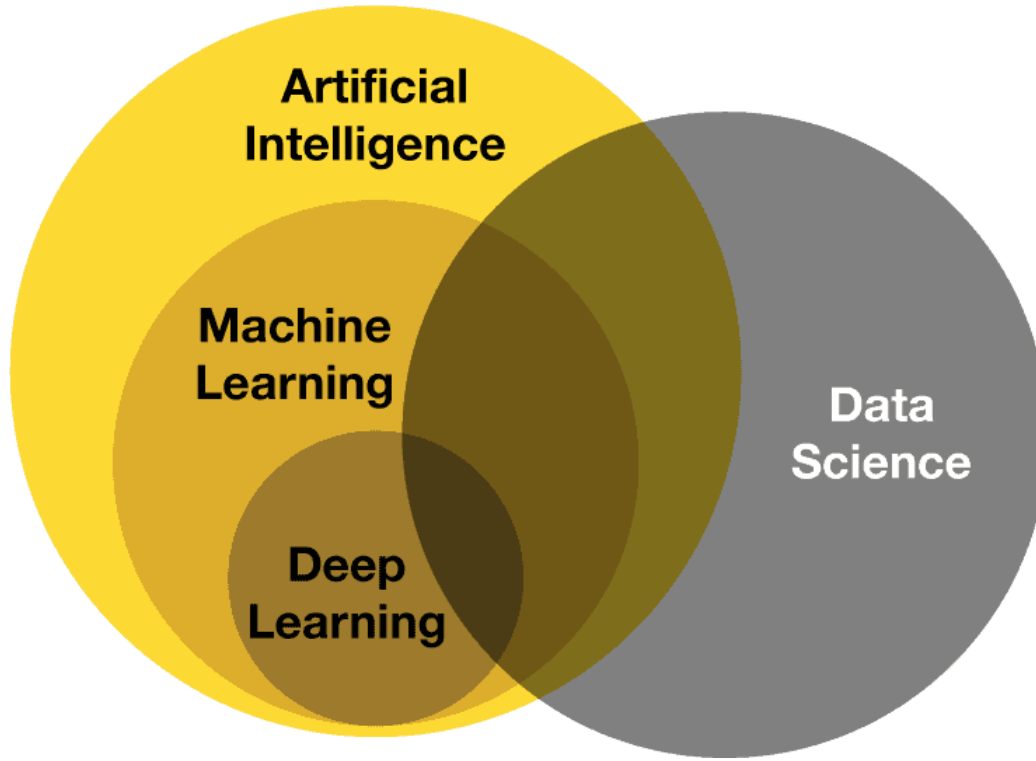
Subset of AI techniques which use statistical methods to enable machines to improve with experiences.

Deep Learning

Subset of ML which make the computation of multi-layer neural networks feasible.



AI, ML, DL, and DS



ML/DS Requirements

- Exists some underlying pattern to be learned, so performance can be measured
- No programmable (easy) definition, you won't need ML/DS to compute $1 + 1$
- There is data about the pattern, so machine learning/data science could have inputs to analyze and learn from

A bit more formal

Machine Learning (ML) is the study of algorithms that:
at some **task T**,
improve their **performance P**,
with **data or experience E**

A well-defined ML task is given by: **(T, P, E)**

ML Task Examples

Machine Learning (ML) is the study of algorithms that:
at some **task T**,
improve their **performance P**,
with **data or experience E**

A well-defined ML task is given by: **(T, P, E)**

T: Recognizing hand-written words

P: Percentage of words correctly classified

E: Database of human-labeled images of handwritten words

ML Task Examples

Machine Learning (ML) is the study of algorithms that:
at some **task T**,
improve their **performance P**,
with **data or experience E**

A well-defined ML task is given by: (T, P, E)

T: Recognizing hand-written words

P: Percentage of words correctly classified

E: Database of human-labeled images of handwritten words

T: Driving on four-lane highways using vision sensors

P: Average distance traveled before a human-judged error

E: A sequence of images and steering commands recorded while observing a human driver.

ML Task Examples

Machine Learning (ML) is the study of algorithms that:
at some **task T**,
improve their **performance P**,
with **data or experience E**

A well-defined ML task is given by: (T, P, E)

T: Recognizing hand-written words

P: Percentage of words correctly classified

E: Database of human-labeled images of handwritten words

T: Driving on four-lane highways using vision sensors

P: Average distance traveled before a human-judged error

E: A sequence of images and steering commands recorded while observing a human driver.

T: Categorize email messages as spam or legitimate.

P: Percentage of email messages correctly classified.

E: Database of emails, some with human-given labels

Types of Learning

→ **Supervised learning**

Given: Training data + desired outputs (labels)

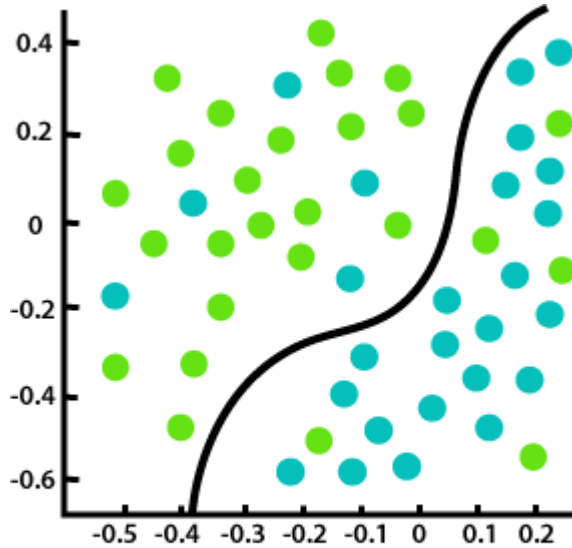
Learn: How to predict the label of new data

→ **Unsupervised learning**

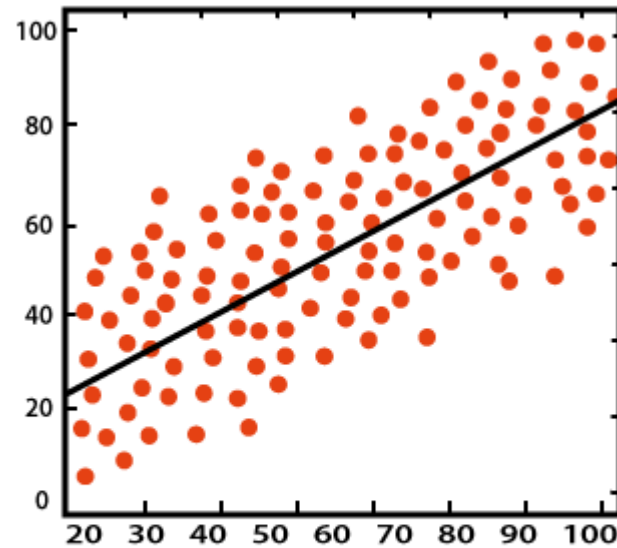
Given: Training data (without desired outputs)

Learn: The inherent pattern of data

Supervised Learning



Classification

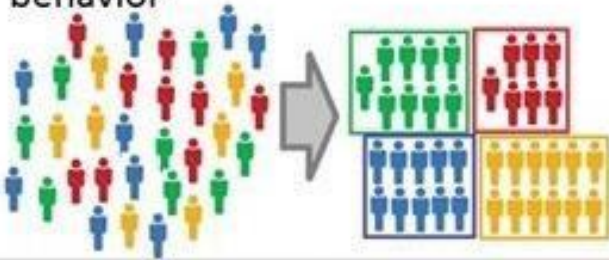


Regression

Unsupervised Learning

Clustering

Grouping customers by purchasing behavior

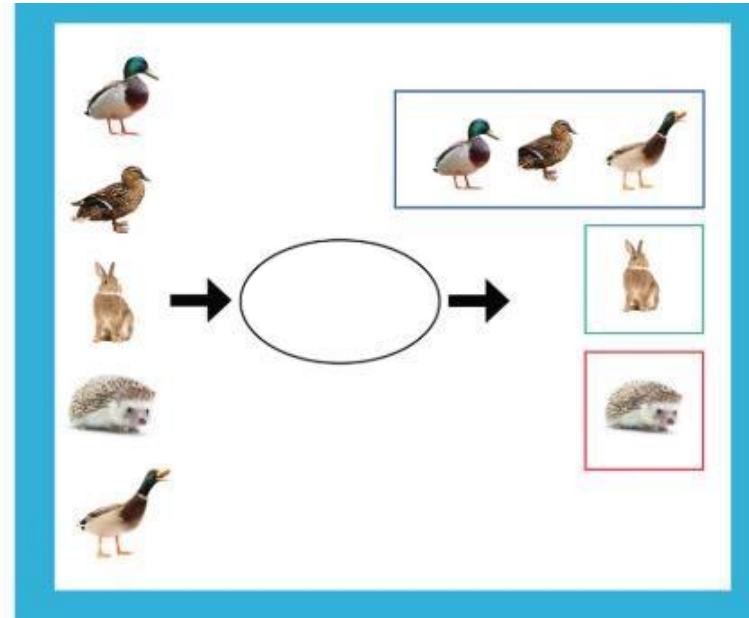
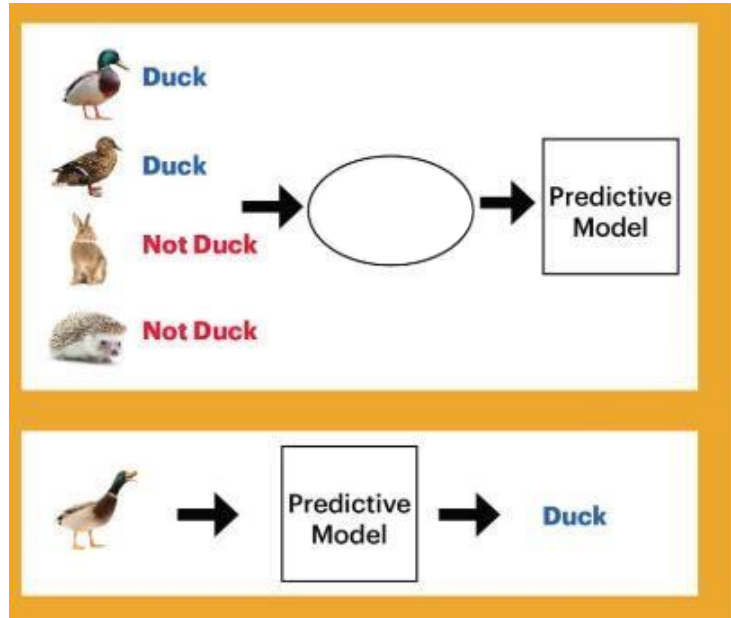


Association

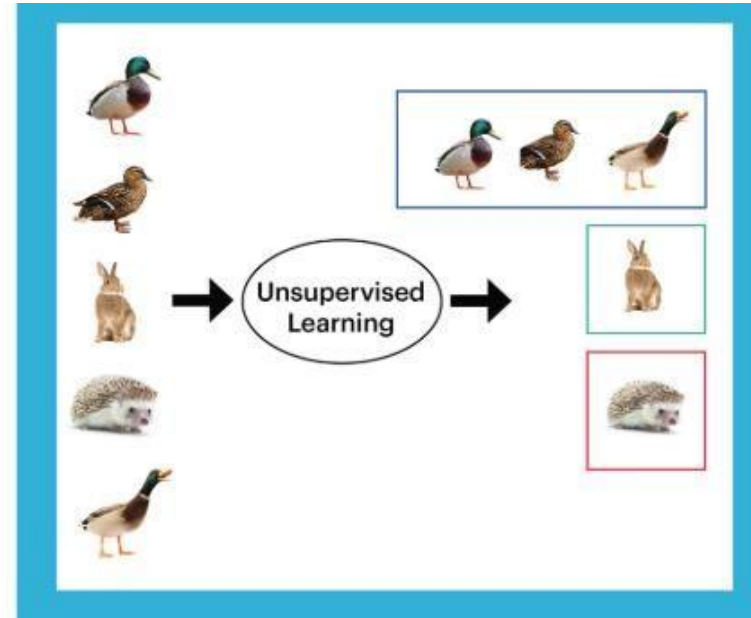
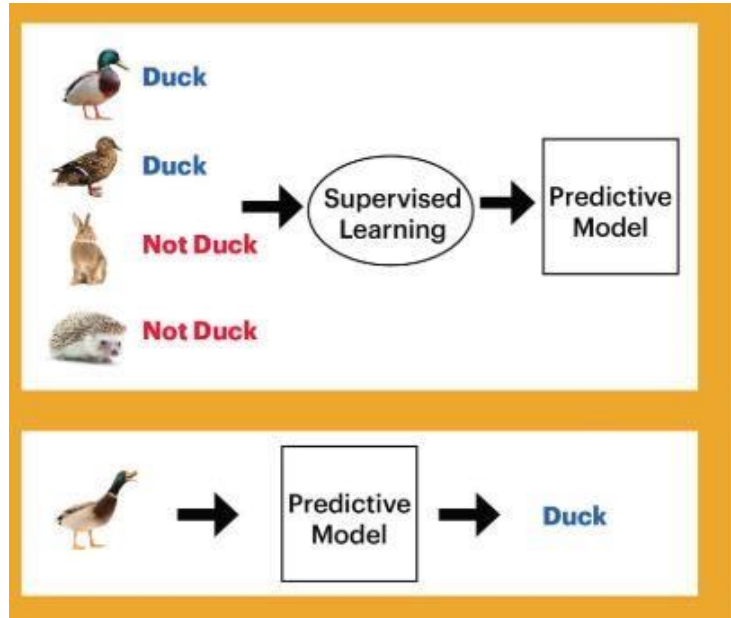
People that buy X tend to buy Y
People that buy A+B tend to buy C



Quiz Time: Which one is unsupervised vs. supervised?



Quiz Time: Which one is unsupervised vs. supervised?



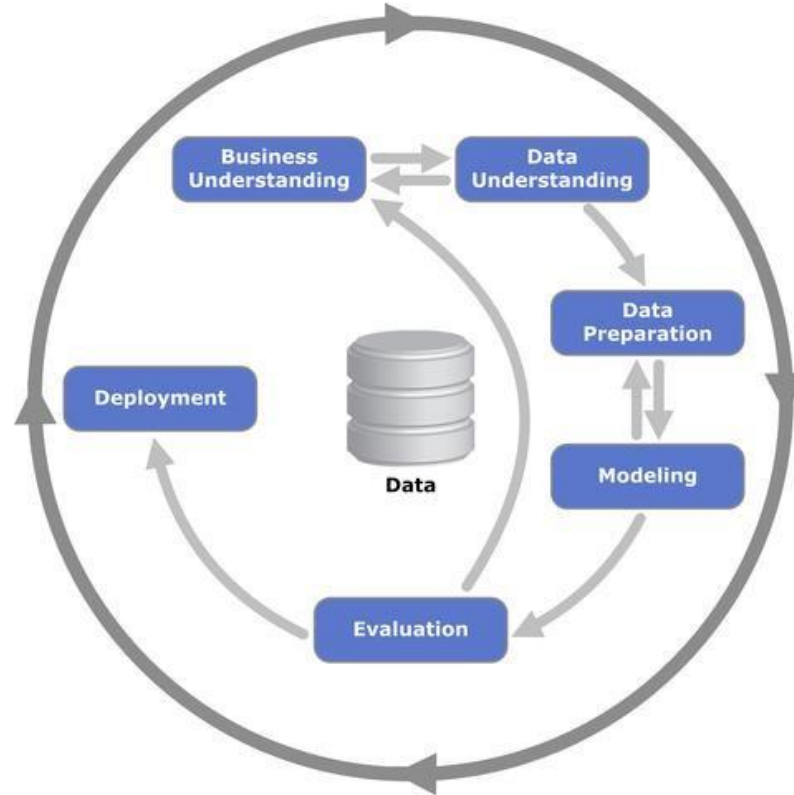
CRISP-DM Methodology

- Cross-Industry Standard Process for Data Mining
- European community funded effort to develop framework for data mining/data science tasks
- Has been around since 1999
- Goals:
 - > Encourage interoperability
 - > Clearer breakdown of data science tasks

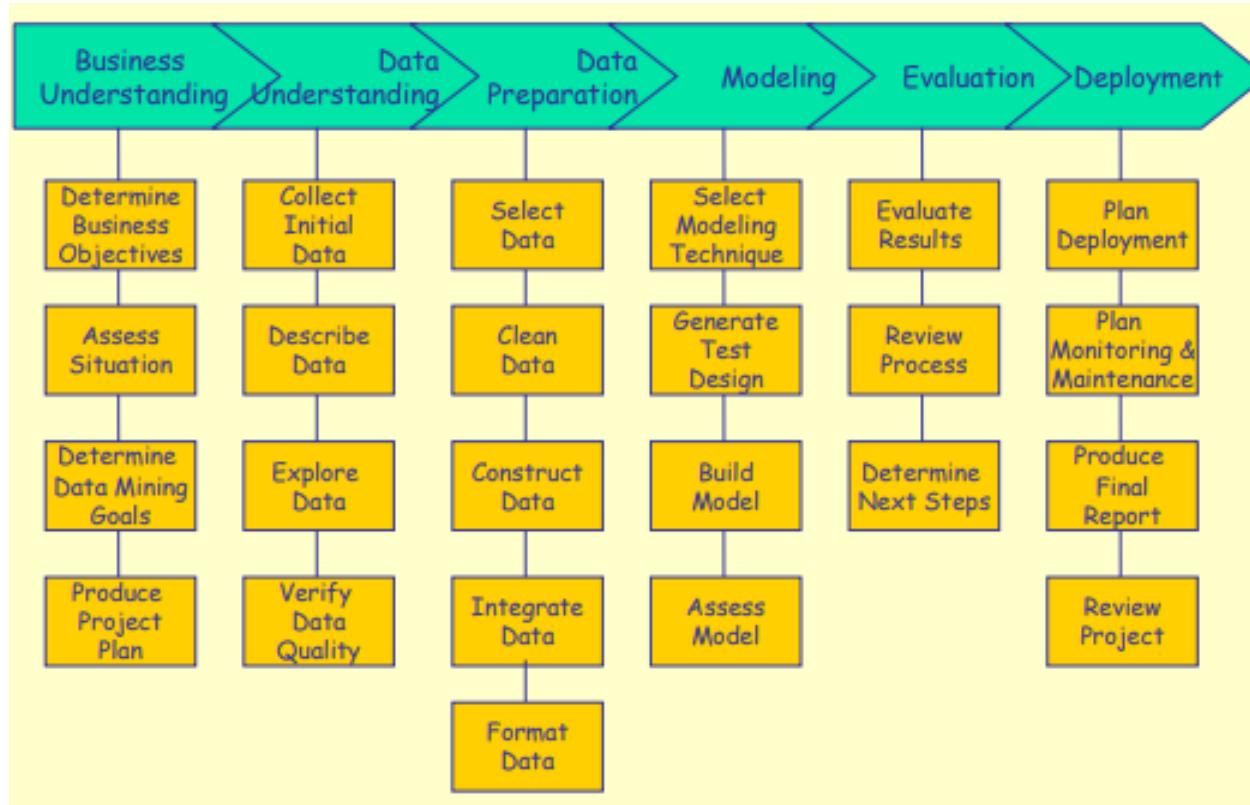
CRISP-DM Methodology Characteristics

- Non-proprietary
- Application/industry neutral
- Tool neutral
- Focus on business issues and technical analysis
- Industry-proven way to guide your data science efforts

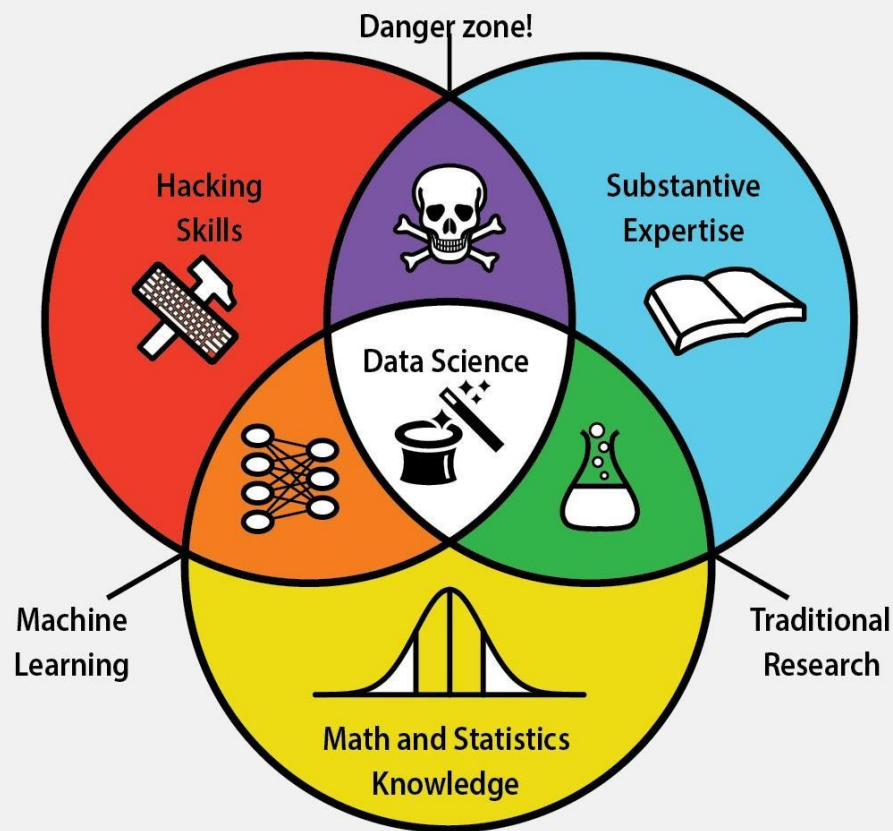
CRISP-DM Methodology Phases



CRISP-DM Methodology Phases and Tasks



DATA SCIENCE SKILLSET



Data science, due to its interdisciplinary nature, requires an intersection of abilities: **hacking skills**, **math and statistics knowledge**, and **substantive expertise** in a field of science.



Hacking skills are necessary for working with massive amounts of electronic data that must be acquired, cleaned, and manipulated.



Math and statistics knowledge allows a data scientist to choose appropriate methods and tools in order to extract insight from data.



Substantive expertise in a scientific field is crucial for generating motivating questions and hypotheses and interpreting results.



Traditional research lies at the intersection of knowledge of math and statistics with substantive expertise in a scientific field.



Machine learning stems from combining hacking skills with math and statistics knowledge, but does not require scientific motivation.




Danger zone! Hacking skills combined with substantive scientific expertise without rigorous methods can beget incorrect analyses.

Roles in Data Science




Data Scientist

also known as Data Managers, statisticians.



A data scientist will be able to take data science projects from end to end. They can help store large amounts of data, create predictive modelling processes and present the findings.

Skills: Mathematics, Programming, Communication



Will use programmes such as:
SQL, Python, R

Data Engineers

also known as database administrators and data architects.



They are versatile generalists who use computer science to help process large datasets. They typically focus on coding, cleaning up data sets, and implementing requests that come from data scientists.

Skills: Programming, Mathematics, Big data



Will use programmes such as:
Hadoop, NoSQL, and Python

Data Analysts

also known as business Analysts.



They typically help people from across the company understand specific queries with charts.

Skills: Statistics, Communication, Business knowledge



Will use programmes such as:
Excel, Tableau, SQL

5+1 Tips on Starting a Data Science Career

1. Data science is a big field: You can't know everything about everything
2. Solve for efficiency
3. Data is never clean: Deal with it
4. Data science is more than machine learning
5. Don't tell me your worth, prove it!
6. Be kind and help others



Thank You

4 005401 190081