Introduction to Machine Learning and AI

Prof. Douglas Kunda, PhD

Vice Chancellor – ZCAS University

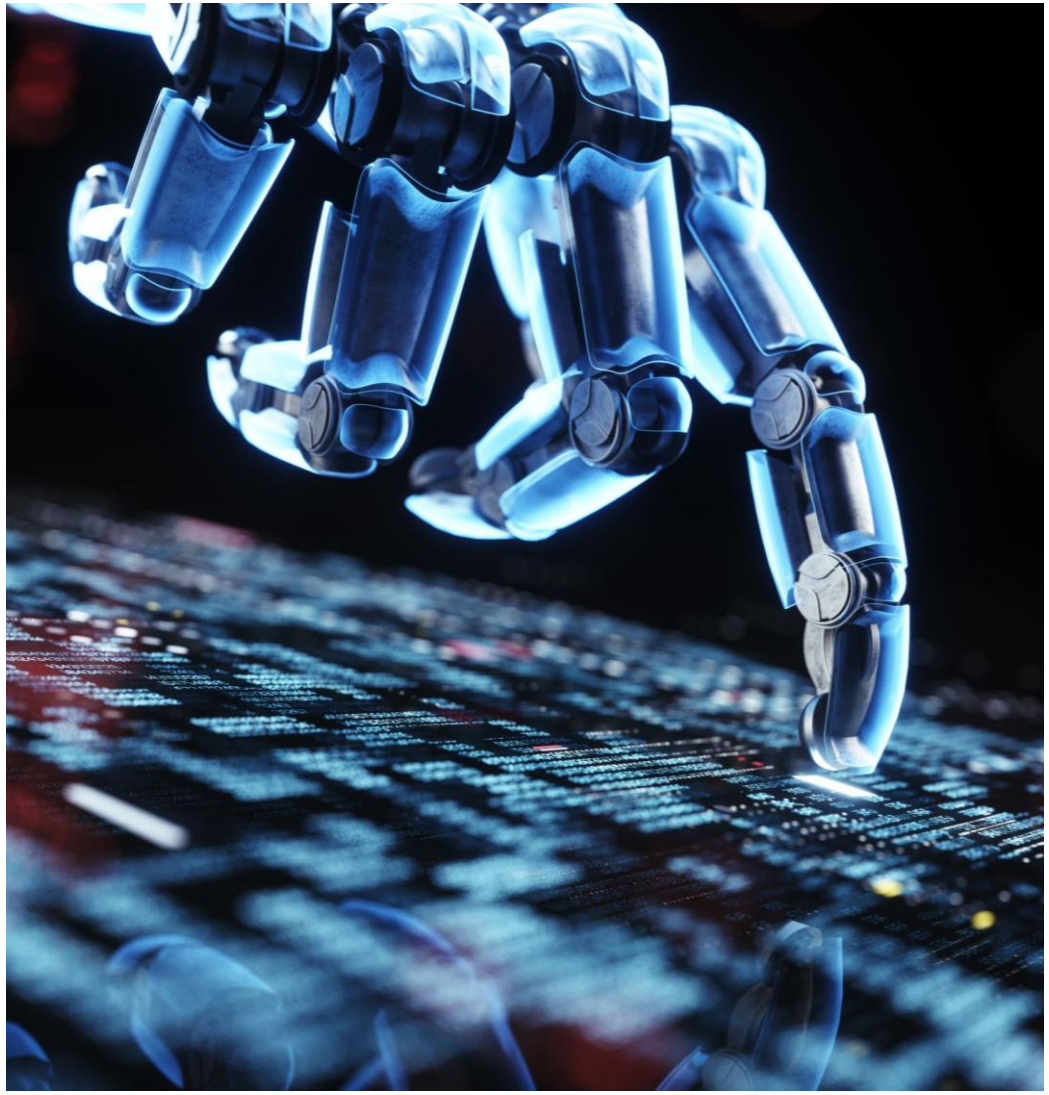
vc@zcasu.edu.zm

Contents

- Pre-requisites
- Approach
- AI, ML, Data Mining, Business Analytics, Data Science, Big Data
- Application of AI and Machine Learning
- Python example
- Exercise

Pre-requisite and learning outcomes

Pre-quisite Skills	Expected Learning outcomes
• Linear Algebra • Basic Statistics • Probability theory • Optimization • Basic python	• Develop new ML and AI algorithms • Develop new and efficient ML packages for implementation • Application of existing ML and AI algorithms to solve new or existing problems • Use of existing ML and AI software to solve new or existing problems

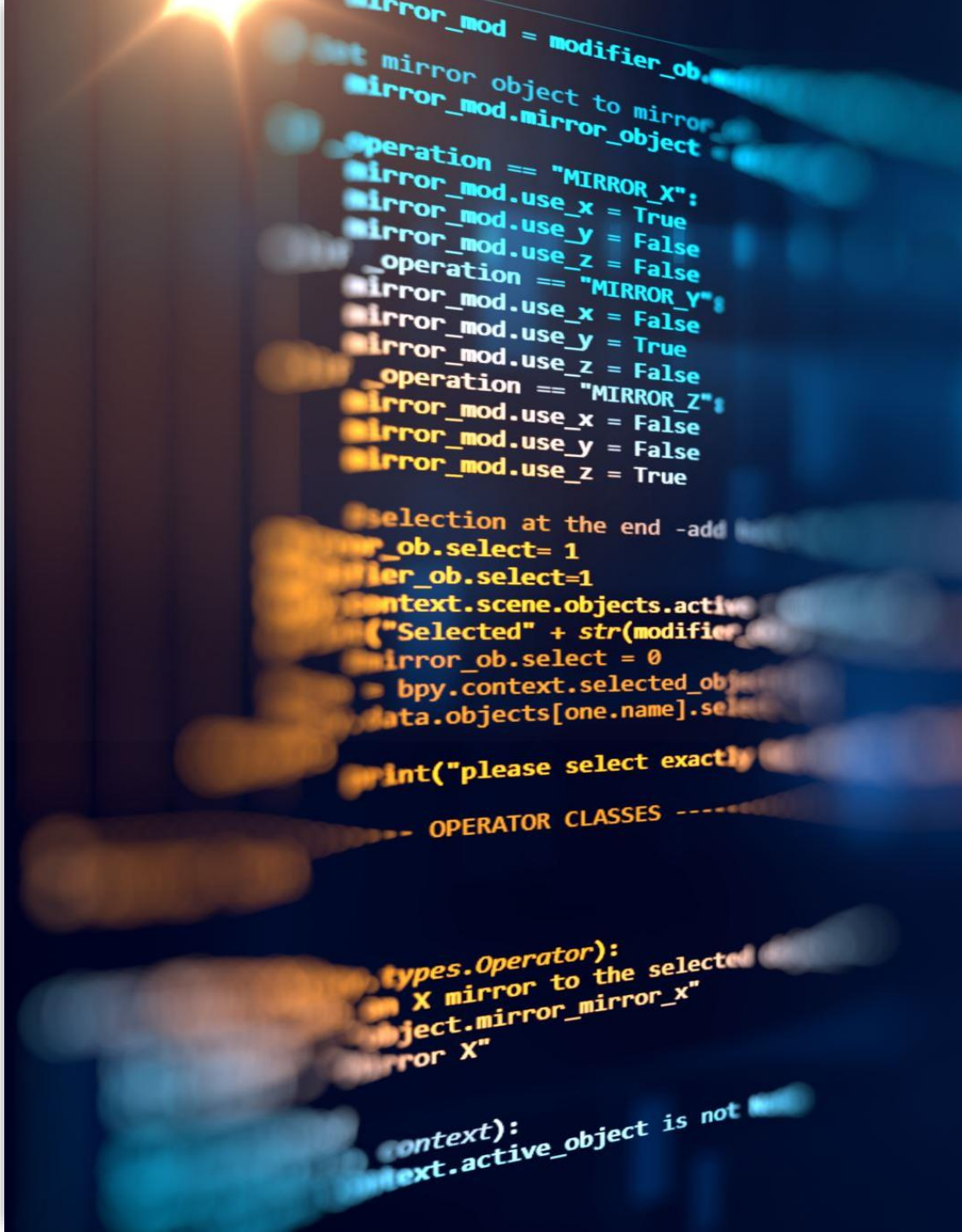


Topics covered

- Introduction to AI and Machine Learning
- Supervised Learning
- Unsupervised Learning
- Deep and reinforced Learning
- NLP and LLM
- Speech recognition and computer vision
- Robotics and AI
- Challenges of AI and ML

Approach

- Introduction to concepts
- AI and ML algorithms
- Pseudo code for the algorithms
- Python implementation example
- Efficiency and validation of AI and ML algorithms
- Exercise on application. Python examples shared in advance or use ChatGPT or BARD



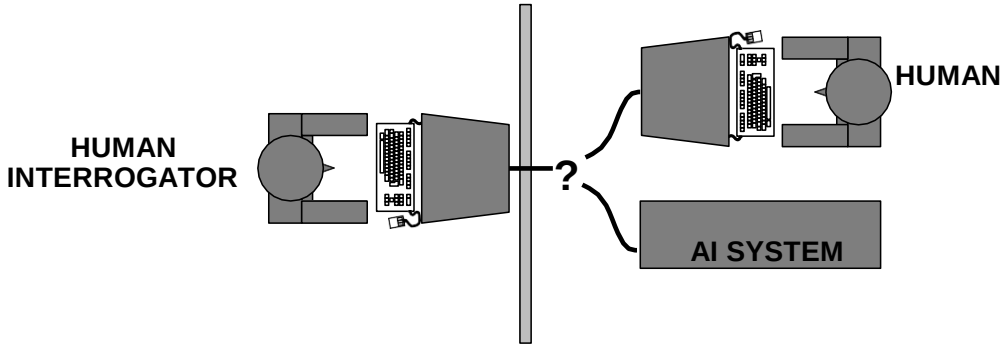
What is AI?

Systems that think like humans	**Systems that think rationally**
Systems that act like humans	Systems that act rationally

Acting humanly: the Turing test

Turing (1950) "Computing machinery and intelligence":

- "Can machines think?" $\longrightarrow$ "Can machines behave intelligently?"
- Operational test for intelligent behavior: the Imitation Game



- Predicted that by 2000, a machine might have a 30% chance of fooling a lay person for 5 minutes
- Anticipated all major arguments against AI in following 50 years
- Suggested major components of AI: knowledge, reasoning, language understanding, learning

Problem: Turing test is not reproducible, constructive, or amenable to mathematical analysis

Thinking humanly: Cognitive Science

1960s "cognitive revolution": information-processing psychology replaced prevailing orthodoxy of behaviorism

Requires scientific theories of internal activities of the brain

- What level of abstraction? "Knowledge" or "circuits"?
- How to validate? Requires
  1) Predicting and testing behavior of human subjects (top-down) or 2) Direct identification from neurological data (bottom-up)

Both approaches (roughly, Cognitive Science and Cognitive Neuroscience) are now distinct from AI

Both share with AI the following characteristic:

the available theories do not explain (or engender) anything resembling human-level general intelligence

Hence, all three fields share one principal direction!

Thinking rationally: Laws of Thought

Normative (or prescriptive) rather than descriptive

Aristotle: what are correct arguments/thought processes? Several

Greek schools developed various forms of logic:

notation and rules of derivation for thoughts;
may or may not have proceeded to the idea of mechanization

Direct line through mathematics and philosophy to modern AI

Problems:

1)Not all intelligent behavior is mediated by logical deliberation
2)What is the purpose of thinking? What thoughts should I have out of all the thoughts (logical or otherwise) that I could have?

Acting rationally

Rational behavior: doing the right thing

The right thing: that which is expected to maximize goal achievement, given the available information

Doesn't necessarily involve thinking—e.g., blinking reflex—but thinking should be in the service of rational action

Aristotle (Nicomachean Ethics):

> Every art and every inquiry, and similarly every action and pursuit, is thought to aim at some good

Rational agents

An agent is an entity that perceives and acts

This course is about designing rational agents

Abstractly, an agent is a function from percept histories to actions:

$$f : P^* \rightarrow A$$

For any given class of environments and tasks, we seek the agent (or class of agents) with the best performance

Caveat: computational limitations make perfect rationality unachievable

→ design best program for given machine resources

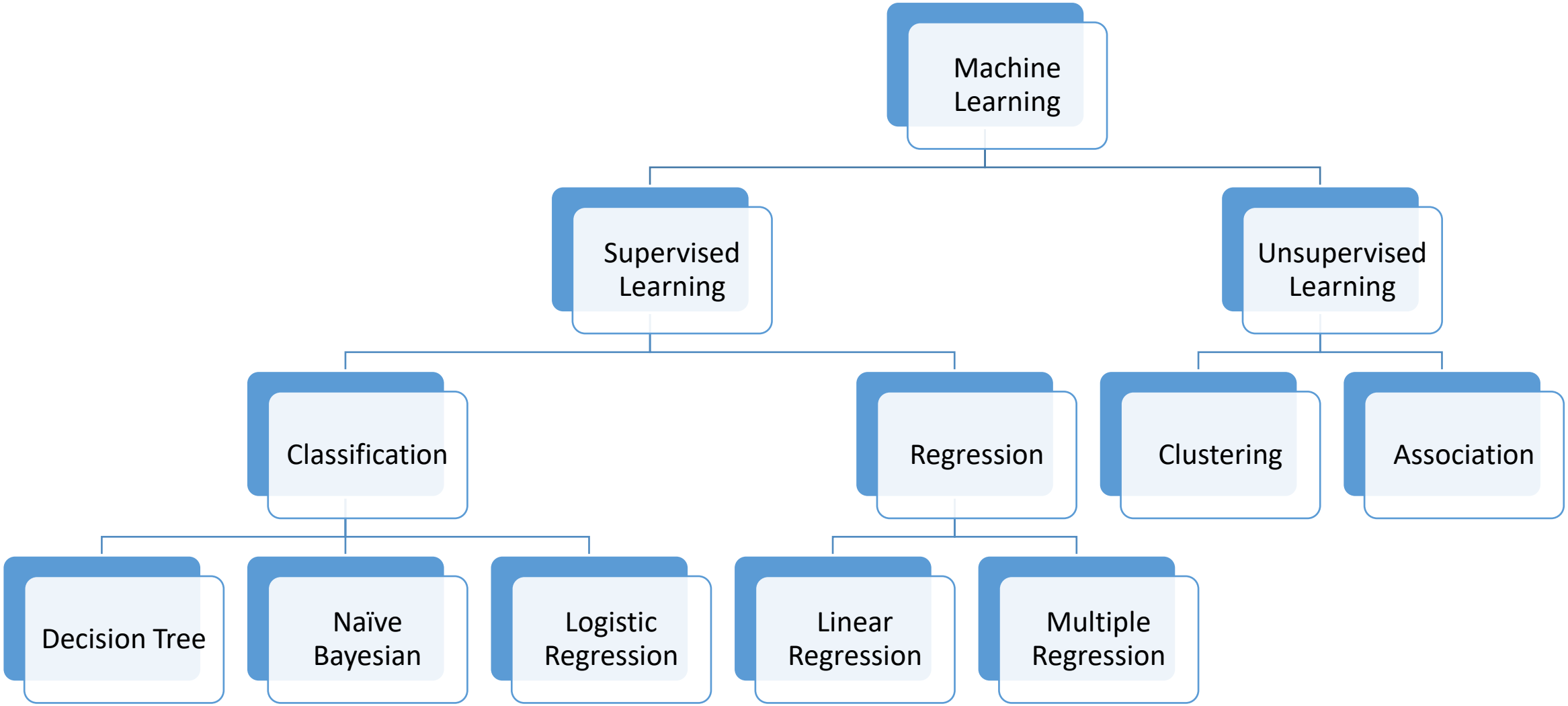
Artificial intelligence

- Artificial intelligence (AI) is a branch of computer science that deals with the creation of intelligent agents, which are systems that can reason, learn, and act autonomously.
- AI research has been highly successful in developing effective techniques for solving a wide range of problems, from flying a plane to medical diagnosis.
- Some of the most common AI techniques include machine learning, natural language processing, and computer vision.
- However, AI has the potential to revolutionize many aspects of our lives, from the way we work to the way we interact with the world around us.

Machine learning

- Machine Learning (ML) is a subset of artificial intelligence (AI) that involves the development of algorithms and models that enable computers to learn from and make predictions or decisions based on data.
- ML algorithms learn patterns from historical data and use these patterns to make predictions or decisions about new, unseen data.
- Supervised Learning: Learns from labeled data, predicting outcomes based on existing examples.
- Unsupervised Learning: Extracts patterns and relationships from unlabeled data, such as clustering similar data points.
- Reinforcement Learning: Involves training models to make sequential decisions through interaction with an environment.

Machine Learning Categories

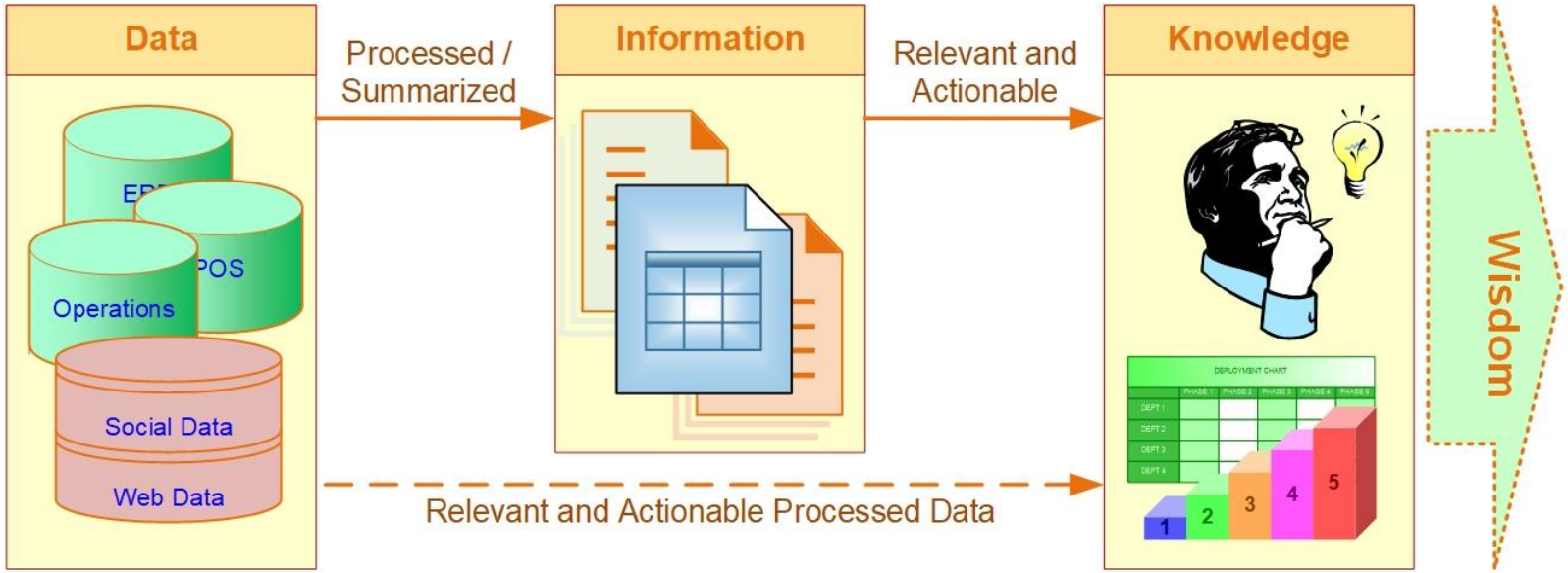


Data Mining

- It is the foundation of data/business analytics
- It is the process of discovering previously unknown patterns/knowledge in data
- Discovery of new knowledge improves customer service, and improved customer service derives Return on investment (ROI), profitability, and competitive posture
- Depending on who is defining it, data mining can be narrow or broad definition
  - Software vendors and consultancy companies have been the force behind these varying definitions

Data Mining & Knowledge Management

- Data mining is the process of converting data into information and then to knowledge …



- What is wisdom?

Drivers Behind Data Mining & ML

- More intense competition at the global scale
- Changing needs/wants of the customer
- Recognition of the value of data
- Changing culture of business management
- Improved data capture/storage techniques
- Emergence of data warehousing
- Hardware and software enhancements
- Cost of ownership
- Availability of data

What is Data Mining?

- The process of discovering (i.e., mining) knowledge (i.e., actionable information) from large amounts of data.
- Is “data mining” a misnomer?
  - Knowledge discovery in databases (KDD)?
  - Pattern mining, knowledge mining, insight mining, …
- The process involves statistical, mathematical, and artificial intelligence (machine learning) techniques and algorithms to extract patterns
  - e.g., business rules, affiliations, correlations, groupings, trends, predictions, etc.

What is Data Mining?

- Original definition: “data mining is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data stored in structured databases”

Fayyad *et al.* (1996)

- Keywords:
  - Process…
  - Nontrivial…
  - Valid…
  - Novel…
  - Useful and understandable…

Is Data Mining a New Discipline?

- Data mining is not a new discipline!
- It is a new philosophy for creating knowledge from secondary data sources

What Data Mining Is Not

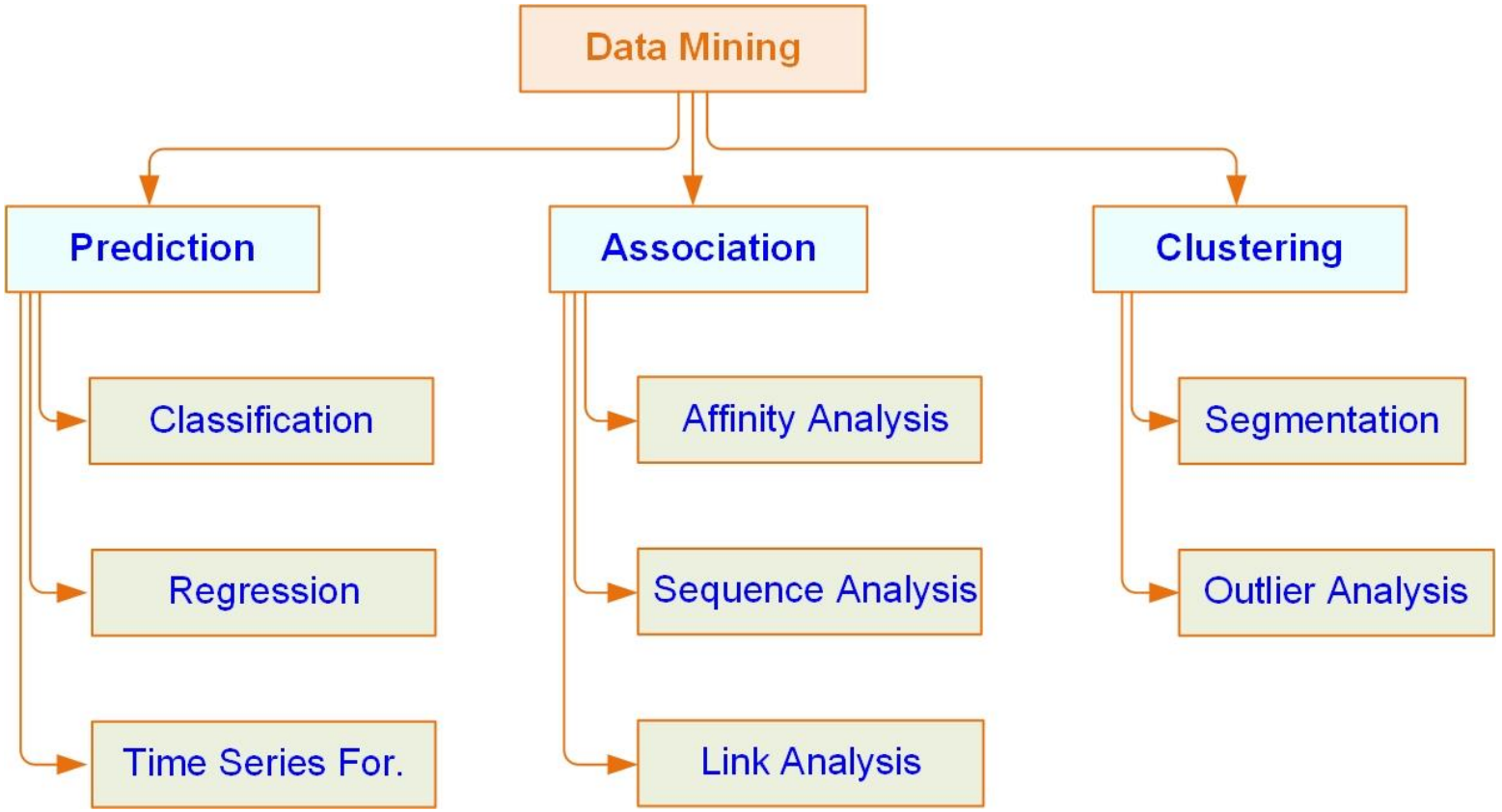
- Because of its popularity, its definition and scope has been a moving target
- Data mining is not
  - Internet searching…
  - Online analytics processing (OLAP)…
    - Data visualization… Dashboarding…
  - Statistics…
    - Hypothesis testing…
    - Causality analysis…
- Rather, it is the discovery of new knowledge from large collections of data sources

Most Common Data Mining Applications

- Marketing and CRM
- Banking, finance, insurance, security trading
- Retail, inventory, and logistics
- Manufacturing
- Health, healthcare, medicine
- Entertainment and sport
- Computer hardware and software
- Homeland security and law enforcement
- Education and professional training …

What Kind of Patterns?

- Data mining is used to discover a wide variety of patterns hidden in large data bases



Difference between ML and DM

- Data mining and machine learning are both subfields of artificial intelligence (AI) that deal with extracting knowledge from data.
- Data mining is the process of extracting patterns and trends from data. This can be done using a variety of techniques, such as statistical analysis, machine learning, and natural language processing. Data mining is often used to find hidden insights in data that would not be obvious to humans.
- Machine learning is a type of artificial intelligence that allows computers to learn without being explicitly programmed. Machine learning algorithms are trained on data, and they learn to make predictions or decisions based on the patterns in the data. Machine learning is often used for tasks such as classification, regression, and clustering.

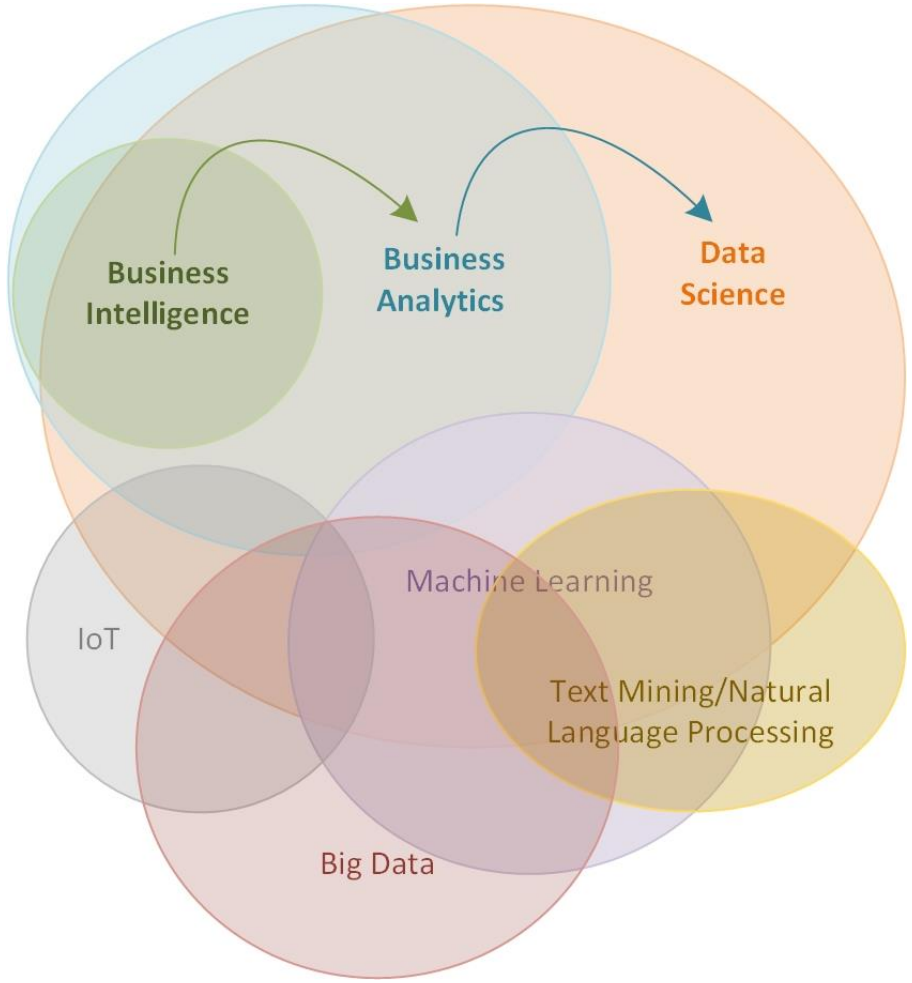
Feature	Data mining	Machine learning
Goal	To extract patterns and trends from data	To learn from data and make predictions or decisions
Techniques	Statistical analysis, machine learning, natural language processing	Machine learning algorithms
Applications	Customer segmentation, fraud detection, risk assessment	Spam filtering, image recognition, natural language processing

Business Use of Analytics

- Help in improving relationships with customers, suppliers, employees and stakeholders
- Detecting fraudulent transactions
- Enhance product and service features
- Optimize marketing and advertising campaign outcomes
- Optimize inventories and other business operations
- Empower employees with relevant and timely insight for better decision making

Business Analytics

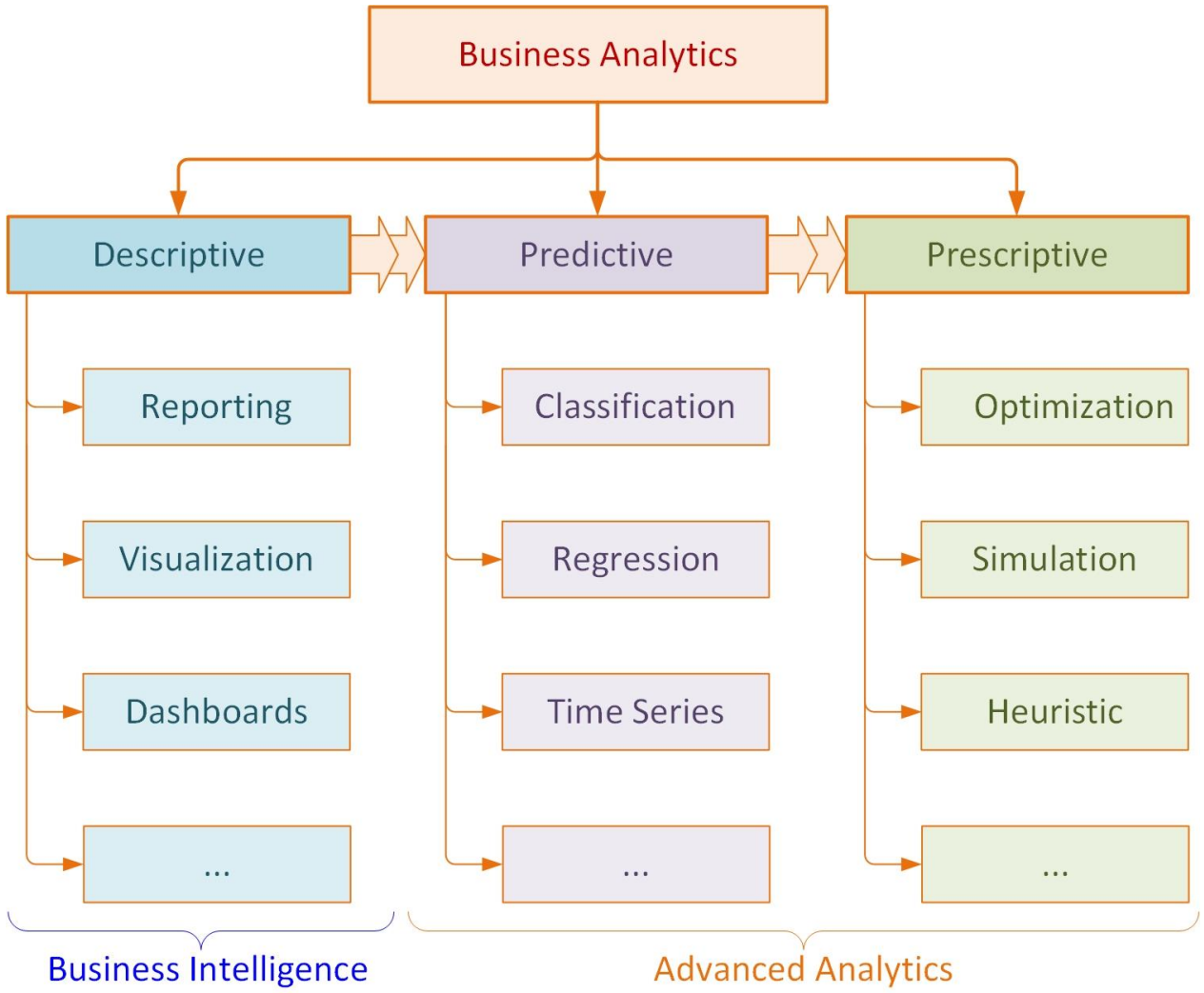
- Confusing of terms and popular buzzwords
- There are significant overlaps among these popular terms
- Business Intelligence vs Business Analytics vs Data Science



Business Analytics

- Business Intelligence → Reporting/Visualization
- Business Analytics → Forecasting/Foreseeing
- Data Science → Algorithmic/Mathematical
- Use of these terms in higher education
- As the data increase in volume, variety, and velocity (i.e., Big Data), so does the
  - Software – new algorithms and methods
  - Hardware – computing resources (CPU, GPU, etc.)
- Analytics versus Analysis
- Where does Data Mining fit in?

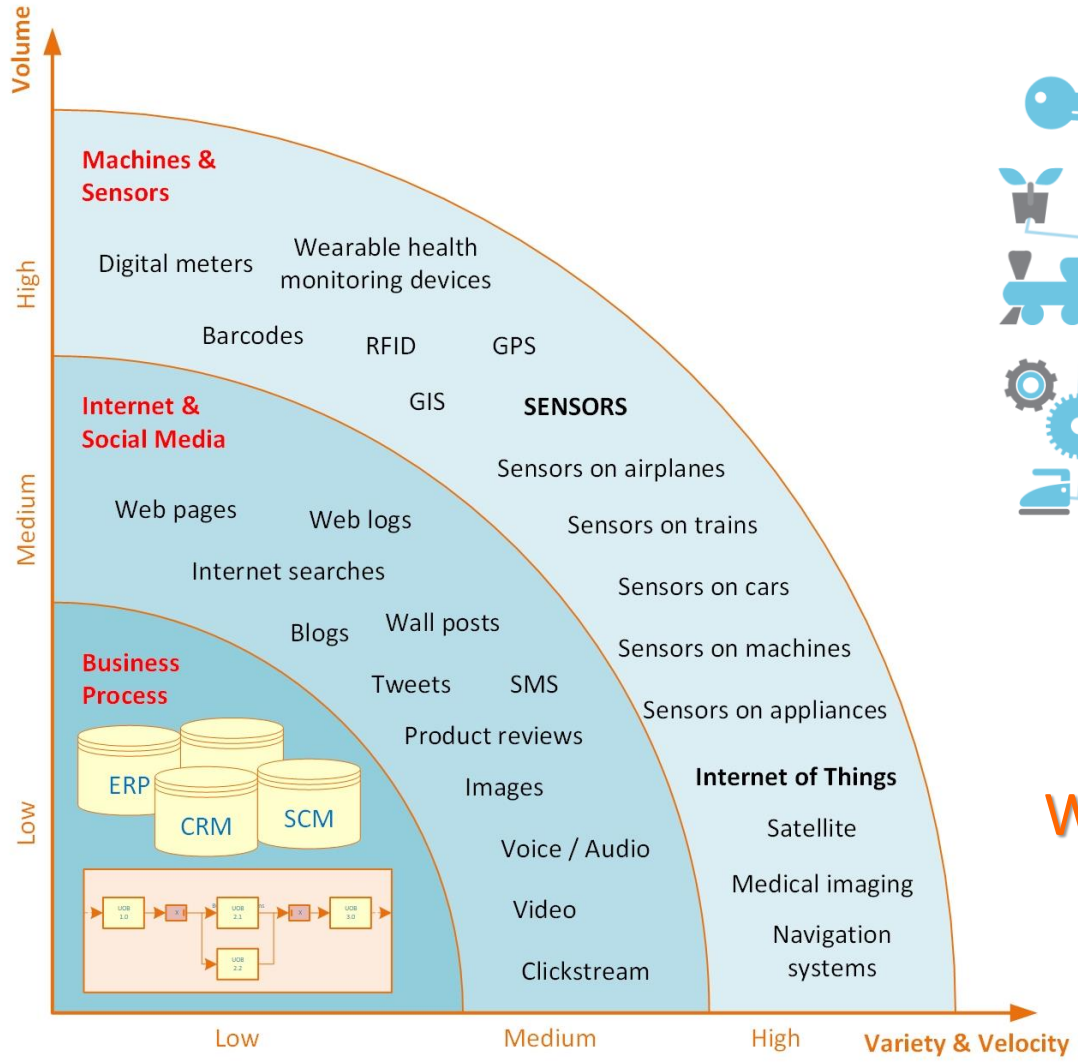
A Simple Taxonomy for Analytics

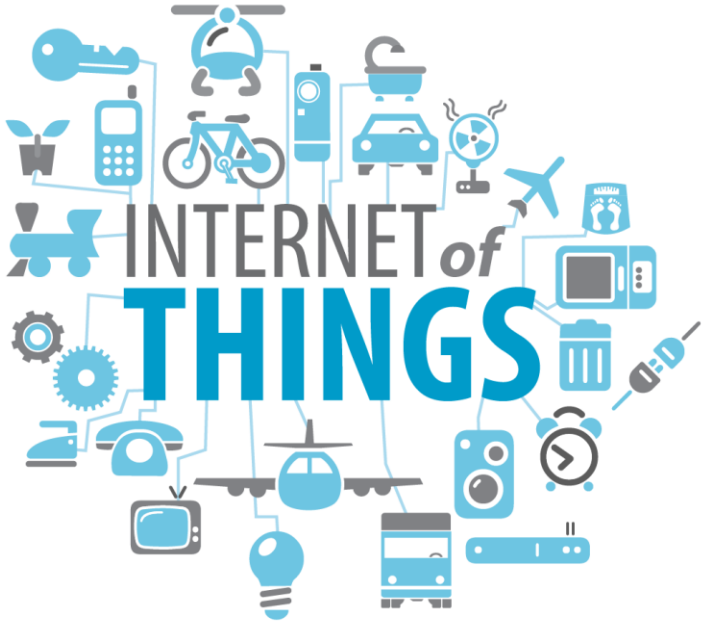


Big Data Concepts and Definition

- Big Data means different things to people with different backgrounds and interests
- Traditionally, “Big Data” = massive volumes of data
  - e.g., volume of data at CERN, NASA, Google, …
- Where does the Big Data come from?
  - Everywhere! Web logs, RFID, GPS systems, sensor networks, social networks, Internet-based text documents, Internet search indexes, detail call records, astronomy, atmospheric science, biology, genomics, nuclear physics, biochemical experiments, medical records, scientific research, military surveillance, multimedia archives, …

Where Does Big Data Come From?





Why is it so popular?

The Vs That Define Big Data

- Is Big Data a misnomer?
- The reality: Big Data is more than just “big”
- The words that start with “V” to define Big Data
  - Volume
  - Variety
  - Velocity
  - Veracity
  - Variability
  - Value
  - …

“Volume” of Big Data

- Hadron collider - 1 PB/sec
- Boeing jet - 20 TB/hr
- Facebook - 500 TB/day
- YouTube – 1 TB/4 min
- …

Name	Symbol	Value
Kilobyte	kB	10^3
Megabyte	MB	10^6
Gigabyte	GB	10^9
Terabyte	TB	10^{12}
Petabyte	PB	10^{15}
Exabyte	EB	10^{18}
Zettabyte	ZB	10^{21}
Yottabyte	YB	10^{24}
Brontobyte*	BB	10^{27}
Gegobyte*	GeB	10^{30}

*Not an official SI (International System of Units) name/symbol, yet.

Fundamental Concepts of Big Data

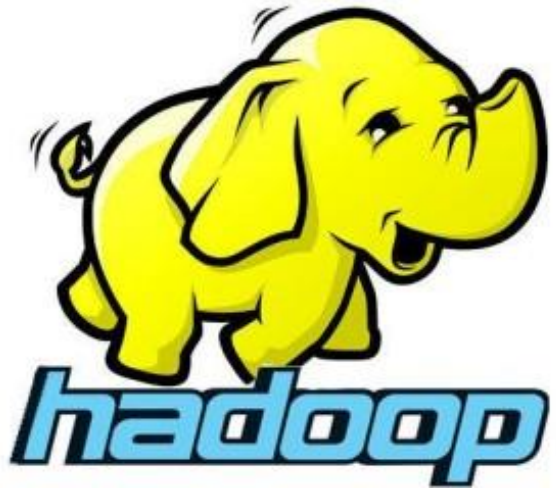
- Big Data by itself, regardless of the size, type, or speed, is worthless
- Big Data + “big” analytics = value
- With the value proposition, Big Data also brought about big challenges
  - Effectively and efficiently capturing, storing, and analyzing Big Data
  - New breed of technologies needed (developed or purchased or hired or outsourced …)
  - New type of talents/skills needed (under the title of “Data Scientist”)

When to Consider Big Data?

- You can’t process the amount of data that you want because of the limitations of your current platform
- You can’t include new data sources (e.g., social media, RFID, Sensory, Web, GPS, textual data) because it does not comply with the data storage schema
- You need to (or want to) integrate data as quickly as possible to be current on your analysis.
- You want to work with a schema-on-demand data storage paradigm because of the variety of data types involved
- The data is arriving so fast at your organization that your traditional analytics platform cannot handle it
- Consider: Data Storage, Data Mining, Data Analytics, Data Visualization

Big Data Technologies

- MapReduce …
- Hadoop …
- HDFS
- Spark
- Hive
- Hbase
- Flume
- Oozie
- Ambari
- Avro
- Mahout, Sqoop, Hcatalog, ….



Big Data Applications

- Process efficiency and cost reduction
- Brand management
- Revenue maximization, cross-selling, and upselling
- Enhanced customer experience
- Churn identification and customer recruiting
- Improved customer service
- Identify new products and market opportunities
- Risk management
- Regulatory compliance
- Enhanced security capabilities

Machine learning application

- **Image recognition:** Machine learning is used to identify objects in images. This is used in a variety of applications, such as facial recognition, medical image analysis, and self-driving cars.
- **Natural language processing:** Machine learning is used to understand natural language. This is used in a variety of applications, such as speech recognition, machine translation, and text summarization.
- **Recommendation systems:** Machine learning is used to recommend products or services to users. This is used in a variety of applications, such as online shopping, streaming media, and social media.
- **Fraud detection:** Machine learning is used to detect fraudulent transactions. This is used in a variety of applications, such as financial services and e-commerce.
- **Risk assessment:** Machine learning is used to assess risk. This is used in a variety of applications, such as insurance, lending, and healthcare.

Artificial intelligence application

- **Virtual assistants:** Virtual assistants, such as Amazon Alexa and Google Assistant, use AI to understand natural language and respond to user requests.
- **Self-driving cars:** Self-driving cars use AI to navigate roads and avoid obstacles.
- **Medical diagnosis:** AI is being used to develop tools that can help doctors diagnose diseases more accurately.
- **Fraud detection:** AI is being used to detect fraudulent transactions in financial services.
- **Customer service:** AI is being used to automate customer service tasks, such as answering FAQs and resolving complaints.
- **Computing service:** AI is being used to automate software development and cybersecurity eg Malware & phishing detection, Knowledge consolidation, Detection & prioritizing new threats, Breach risk prediction, Task automation.
- **AI in now in Everything:** One who leads in AI and ML development will lead the world

Python and package installation

Linux distro

- Python comes preinstalled on most of the Linux distributions like Debian. To check installed Python Version enter on terminal:
- `python -V` or `python -version`
- To install enter the following commands
- `sudo apt-get install python3`
- #or
- `sudo apt-get install python3.9`

Windows

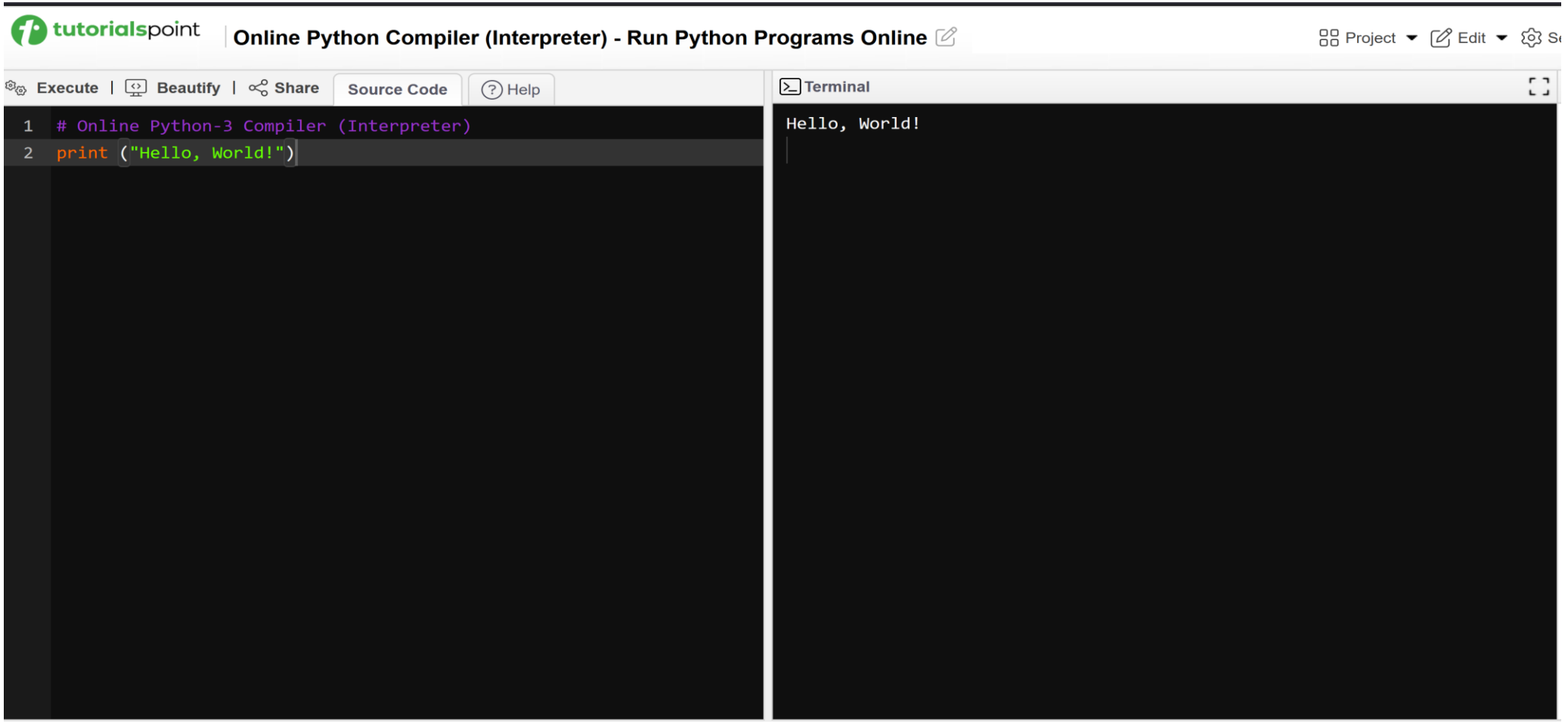
- Go to the official Python download page for Windows and find stable version and download the installer
- Install and follow instructions

To install ML packages enter the following in CMD

- `pip install -U scikit-learn`

Python online

- https://www.tutorialspoint.com/online_python_compiler.php
- https://www.programiz.com/python-programming/online-compiler/#google_vignette



Python example

```
1. import numpy
2. from scipy import stats
3. import matplotlib.pyplot as plt
4. ages=[13, 20, 20, 20, 21, 21, 22, 22, 23,
   24, 24, 25, 25, 25, 25, 28, 30, 33, 33,
   35, 35, 35, 35, 36, 40, 45, 46, 52, 66,
   70]
5. print("Age Values: ",ages)
6. #Compute mean, median, mode
7. ageMean = numpy.mean(ages)
8. print("Age Mean =",ageMean)
9. ageMedian = numpy.median(ages)
10.print("Age Median =",ageMedian)
11.ageMode = stats.mode(ages)
12.print("Age Mode =",ageMode)
13.#compute midRange
14.ageMax = numpy.max(ages)
15.ageMin = numpy.min(ages)
16.ageMidRange=(ageMax+ageMin)/2
17.print("Age MidRange =",ageMidRange)

18.#first and third quartile
19.print("Q1 quantile of age = ",
   numpy.quantile(ages, .25))
20.print("Q3 quantile of age = ",
   numpy.quantile(ages, .75))
21.#five-number summary of the data.
22.print("five-number summary : ","Minimum
   =",ageMin," , Q1 =",numpy.quantile(ages,
   .25),", Median =",
23.ageMedian," Q3 = ",numpy.quantile(ages,
   .75), " Maximum =",ageMax)
24.# Creating boxplot for dataset
25.fig = plt.figure(figsize =(10, 7))
26.plt.boxplot(ages)
27.# show plot for dataset
28.plt.show()
```

Exercise 1

1. Develop new ML and AI algorithms,
- N/A
2. Develop new and efficient ML packages and implementation,
- Install python on your laptop or use online compiler. Write a python program Suppose that the data for analysis includes the attribute age . The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. Using Python solve the following problems (a) What is the mean, median and mode of the data? (b) What is the midrange of the data? (c) what is first and third quartile of the data? (d) Give the five-number summary of the data.(e) Show a boxplot of the data.
3. Application of existing ML and AI algorithms to solve new or existing problems,
- List 3 possible application of ML and AI in your organization that you can implement after this training
4. Use of existing ML and AI software to solve new or existing problems,
- List 3 software in your organization that uses or implements ML and AI

References and further reading

- Stuart Russell (2021), Artificial Intelligence: A Modern Approach, 4th edition. Pearson publishing
- Dursun Delen (2021), Predictive Analytics: Data Mining, Machine Learning and Data Science for Practitioners, 2nd edition, Pearson FT Press
- Paul Deitel and Harvey M. Deitel (2020), Intro to Python for Computer Science and Data Science: Learning to Program with AI, Big Data and The Cloud, 1st edition, Pearson Publishing
- Mark E. Fenner (2020), Machine Learning with Python for Everyone, Addison-Wesley Publishing
- Pang-Ning Tan, Michael Steinbach, Vipin Kumar (2018), Introduction to Data Mining (2nd Edition), Pearson publishing
- Data mining: Concepts and Techniques, by Jiawei Han and Micheline Kamber, Morgan Kaufmann Publishers, ISBN 1-55860-489-8.
- Machine Learning, by Tom M. Mitchell, McGraw-Hill, ISBN 0-07-042807-7
- Modern Information Retrieval, by Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Addison Wesley, ISBN 0-201-39829-X