DSC650: Data Technology and Future Emergence

Lecture 8 : Trend in Data Technology

Lecturer: Dr Khairul Anwar Sedek

Lecture 8: **Trend in Data Technology**

Automated Data Discovery
Deep Learning
The Next Frontier

At the end of the lecture, students should be able to;

- CLO1: Demonstrate an understanding on the basic **concepts and practices** of big data technology
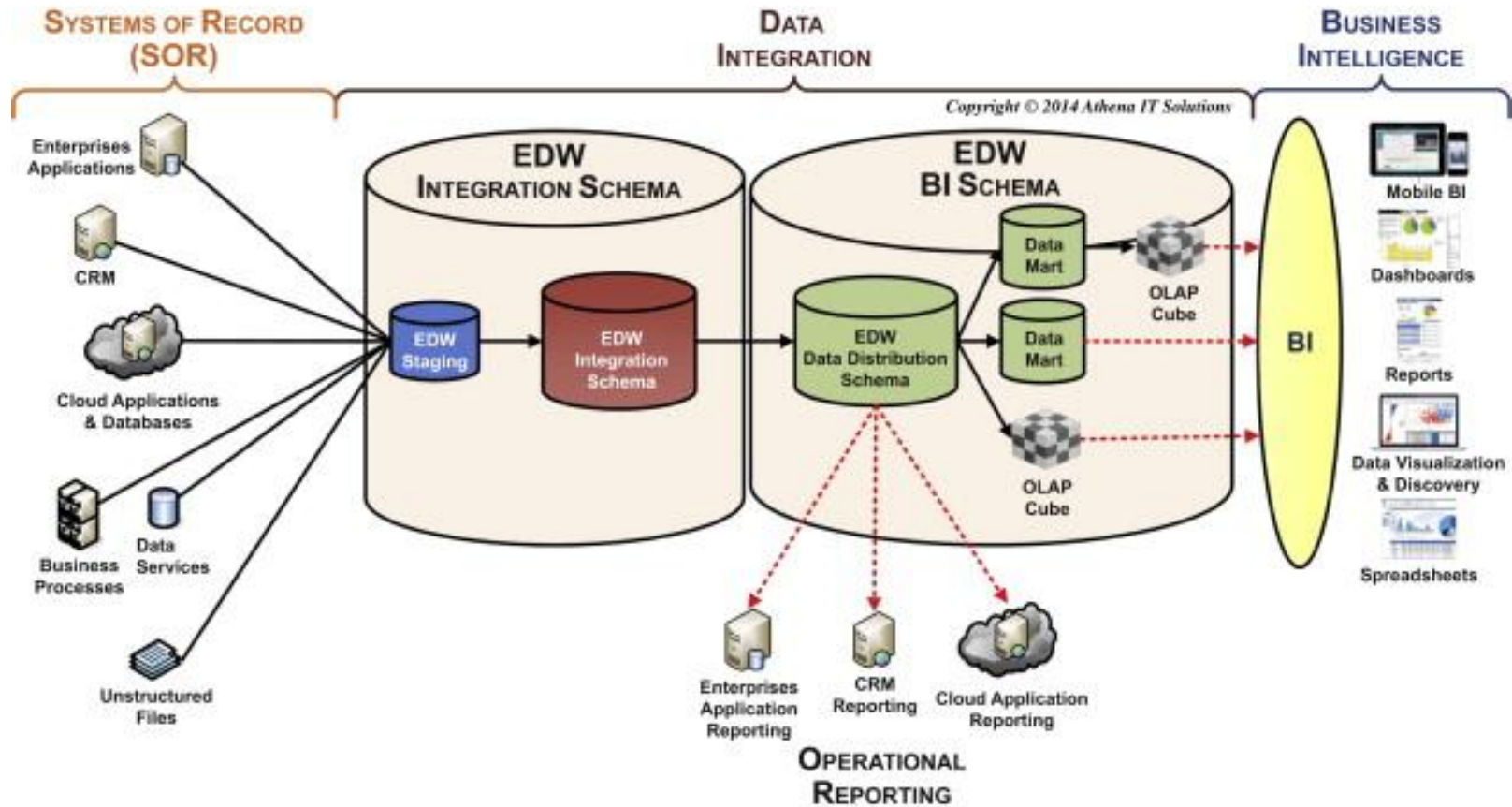
Source:

Kelleher, J.D. (2019) Deep Learning. MIT Press.

Mueller, J.P. & Massaron, L. (2019) Deep Learning for Dummies. John Wiley & Sons.

Data Discovery

- It is a business user-oriented process **for detecting patterns and outliers** by **visually navigating data or applying guided advanced analytics**.

- An iterative process that includes 3 categories:

  – data preparation;

  – visual analysis; and

  – guided advanced analytics.

- requires skills in understanding **data relationships and data modelling** as well as in using data analysis and guided advanced analytics functions to reveal insights
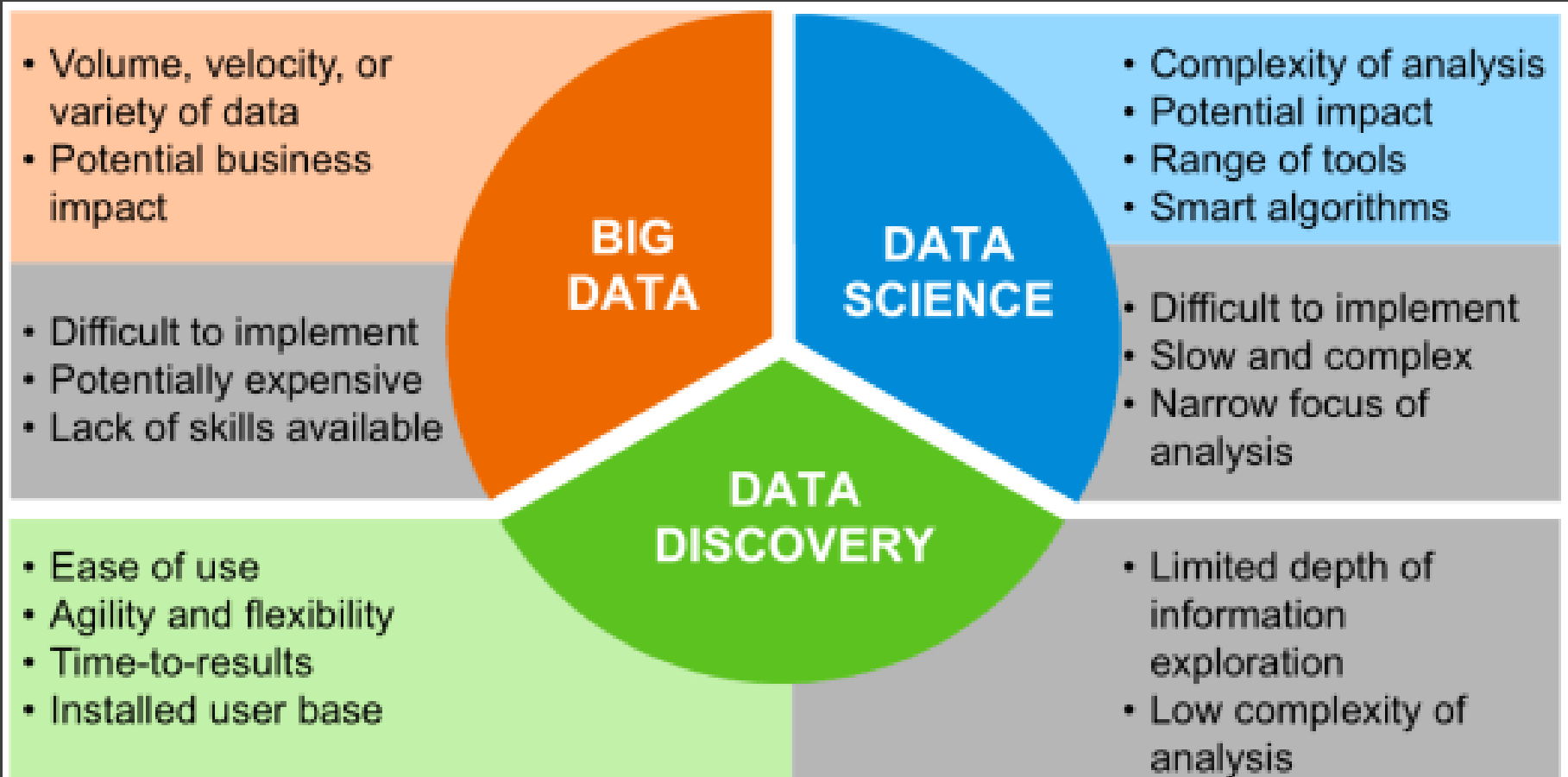
Data Discovery

Sherman, R. (2015). Data Integration Process. Business Intelligence Guidebook , Elsevier Inc.

Big Data, Data Discovery, and Data Science

- Source: Gartner. Big Data Discovery is the combination of Big Data, Data Science, and Data Discovery.

- Volume, velocity, or variety of data
- Potential business impact

- Difficult to implement
- Potentially expensive
- Lack of skills available

BIG DATA

DATA SCIENCE

- Complexity of analysis
- Potential impact
- Range of tools
- Smart algorithms

- Difficult to implement
- Slow and complex
- Narrow focus of analysis

DATA DISCOVERY

- Ease of use
- Agility and flexibility
- Time-to-results
- Installed user base

- Limited depth of information exploration
- Low complexity of analysis

Introduction to Deep Learning

- Deep learning is the subfield of artificial intelligence that focuses on creating large neural network models that are capable of making accurate *data-driven decisions*.

- Deep learning is particularly suited to contexts where the data is complex and where there are large datasets available.

Today most online companies and high-end consumer technologies use deep learning.

- **Facebook** uses deep learning to analyse text in online conversations.

- **Google, Baidu, and Microsoft** all use deep learning for image search, and also for machine translation.

- All modern **smart phones** have deep learning systems running on them; for example, deep learning is now the standard technology for **speech recognition, and also for face detection** on digital cameras.

- In the healthcare sector, deep learning is used **to process medical images** (X-rays, CT, and MRI scans) and **diagnose health** conditions.

- **self-driving cars**, where it is used for localization and mapping, motion planning and steering, and environment perception, as well as tracking driver state.

The relationship between artificial intelligence, machine learning, and deep learning

- Machine learning involves the development and evaluation of algorithms that **enable a computer to extract (or learn) functions from a dataset** (sets of examples)

- Deep learning **enables *data-driven decisions*** by <u>identifying and extracting patterns from large datasets</u> that <u>accurately map from sets of complex inputs to good decision</u> outcomes

Artificial intelligence

Machine learning

Deep learning
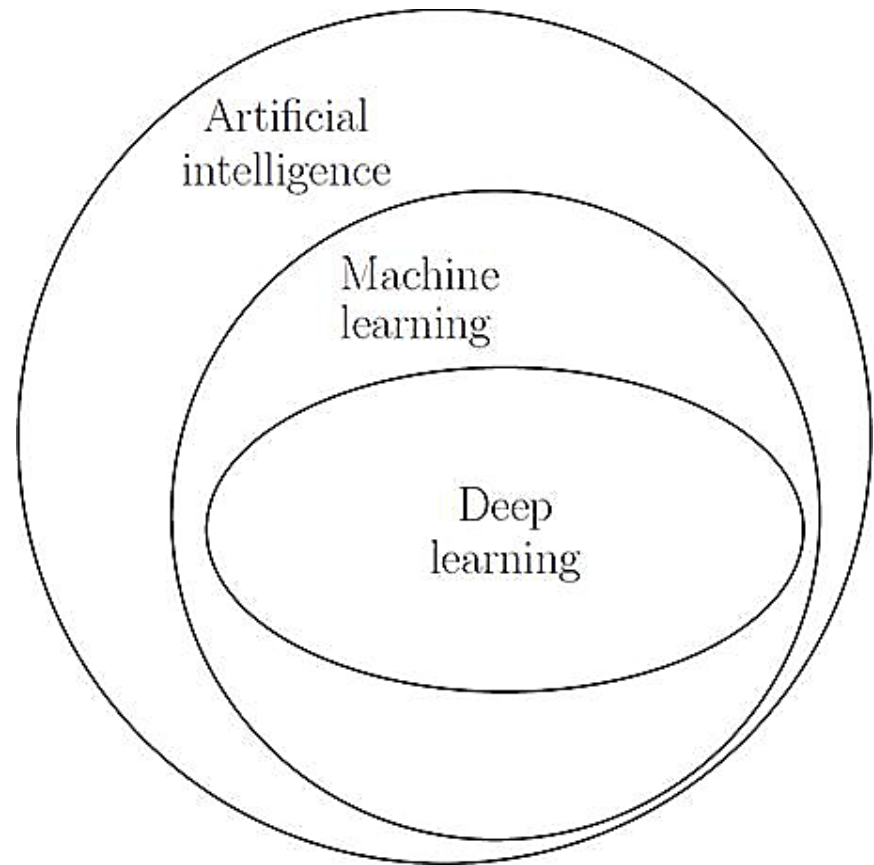
Machine Learning

- A machine learning algorithm is a **search process designed to choose the best function**, from **a set of possible functions**, to **explain the relationships between features in a dataset.**
- To get an intuitive understanding of what is involved in extracting, or learning, a function from data, examine the following set of sample inputs to an unknown function and the outputs it returns.

$$\text{function}(\text{Inputs}) = \text{Output}$$

$$\text{function}(5,5) = 25$$

$$\text{function}(2,6) = 12$$

$$\text{function}(4,4) = 16$$

$$\text{function}(2,2) = 04$$

Three key ingredients in machine learning:

Data (a set of historical examples).

A set of functions that the algorithm will search through to find the best match with the data.

Some **measure of fitness** that can be used to evaluate how well each candidate function matches the data.
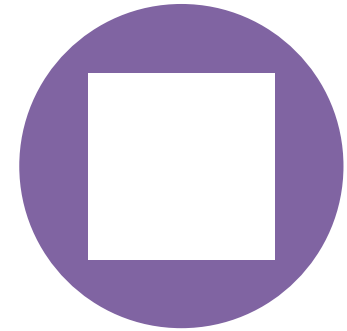
Types of machine learning

SUPERVISED
MACHINE LEARNING

UNSUPERVISED
MACHINE LEARNING

REINFORCEMENT
LEARNING

Supervised Machine Learning

- Supervised ML, each example in the dataset is **labelled with the expected output** (or target) value
  - the benefit of having these target values in the dataset is that the machine learning algorithm can use these values **to help the learning process**.
  - It does this by **comparing the outputs a function** produces with the target outputs specified in the dataset, and **using the difference (or error) to evaluate the fitness of the candidate function**, and use the fitness evaluation to guide the search for the best function.

Data Input (X)	Data Output (y)	Real-World Application
History of customers' purchases	A list of products that customers have never bought	Recommender system
Images	A list of boxes labelled with an object name	Image detection and recognition
English text in the form of questions	English text in the form of answers	Chatbot, a software application that can converse
English text	German text	Machine language translation
Audio	Text transcript	Speech recognition
Image, sensor data	Steering, braking, or accelerating	Behavioral planning for autonomous driving

Examples of Supervised Learning

Unsupervised Machine Learning
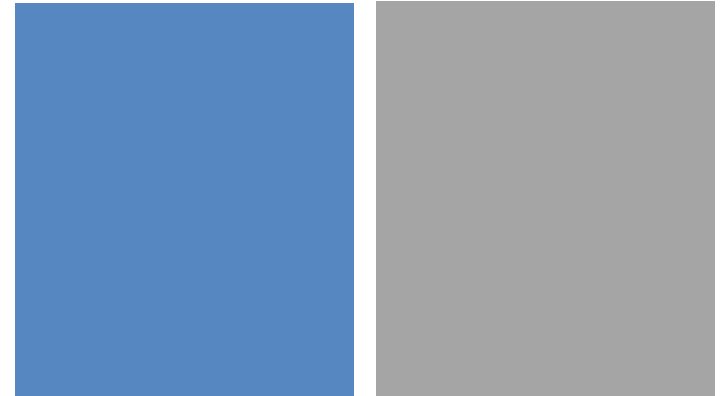
- Unsupervised ML is generally **used for clustering data**.

- For example, this type of data analysis is **useful for customer segmentation**, where a company wishes to segment its customer base into coherent groups so that it can target marketing campaigns and/or product designs to each group.

- Instead, the machine learning algorithm tries to identify functions that map similar examples into clusters, such that the examples in a cluster are more similar to the other examples in the same cluster than they are to examples in other clusters.

- The fitness functions used in unsupervised machine learning generally reward candidate functions that result in higher similarity within individual clusters and, also, high diversity between clusters.

Reinforcement Machine Learning

- Reinforcement learning is most relevant **for online control tasks**, such as robot control and game playing.
- In these scenarios, an agent needs to learn a policy for **how it should act in an environment in order to be rewarded**.
- The goal of the agent is to learn a mapping from its current observation of the environment and its own internal state (its memory) to what action it should take: for instance, *should the robot move forward or backward* or *should the computer program move the pawn or take the queen*.
- Deep learning can be applied to all three machine learning scenarios: supervised, unsupervised, and reinforcement.

Benefits from Business Perspective for using Machine Learning

- **Simplify product marketing:** One of the issues that any organization faces is determining what to sell and when, based on customer preferences.
  - Sales campaigns are expensive, so having one fail usually isn't an option.
  - Knowing what the customer wants is incredibly difficult unless you can **analyse huge amounts of buying data, which is something that machine learning does well.**
- **Predict future sales accurately:** Being in business can seem a little like gambling because you can't be quite sure that your bets will pay off.
  - A machine learning solution can **follow sales minute by minute and track trends** before they become obvious.
  - The capability to perform this kind of **tracking** means that you can **more accurately tune sales channels to deliver optimal results and ensure that stores have enough of the right products to sell**.

Benefits from business perspective for using machine learning

- **Forecast medical and other employee downtime:** Oddly enough, some organizations end up having problems because employees choose the worst possible times to be absent from work.
  - In some cases, these absences seem unpredictable, such as medical needs, while in others you could possibly predict them, such as a sudden need for personal time.
  - By tracking various trends from easily available data sources, you can **track both medical-type and personal-type absences** for your industry as a whole, location as a whole, and your organization in particular to ensure that you have **enough people to get the job done at any given time**.
- **Reduce data entry errors:** Some kinds of data entry errors are relatively easy to avoid by **using form features correctly or incorporating a spell checker** into your application.
  - In addition, adding certain kinds of pattern matching can help reduce capitalization errors or incorrect phone numbers.

Benefits from business perspective for using machine learning

- **Improve financial rule and modelling precision:** Machine learning enables you to perform tasks such as portfolio management, algorithmic trading, loan underwriting, and fraud detection with greater precision.
  - You can't eliminate human participating in such cases, but the human and machine working together can become an incredibly efficient combination that won't allow many errors to pass unnoticed.
- **Foresee maintenance needs:** machine learning can help predict when a system will need cleaning based on past performance and environmental monitoring.
  - You can also do things like plan for replacement or repair of certain equipment based on past repairs and equipment statistics.
  - A machine learning solution can even enable you to determine whether replacement or repair is the better option.
- **Augment customer interaction and improve satisfaction:** You can find a wealth of information about customers through online sources, including everything from recent purchases to consistent buying habits.
  - By combining all this data with a good machine learning solution and customer support personnel who have discerning eyes, you can appear to have personally created a special solution for each customer, even though the time required to do so is minimal.

Machine Learning Limitations

- **Massive amounts of training data are needed:** Unlike programmed solutions of the past, a machine learning solution **relies on massive amounts of data to train it**.
  - As **problem complexity increases, the number of data points required to model a particular problem increases**, making even more data necessary.
  - Although humans generate increasingly larger amounts of data in specific problem domains and the computing power needed to process this data also increases daily, some problem domains simply lack enough data or enough processing power to make machine learning effective.
- **Labelling data is tedious and error prone:** When using the supervised learning technique someone must **label the data to provide the output value**.
  - The labelling process for huge amounts of data is both **tedious and time consuming**, making machine learning difficult at times.
  - The problem is that a human can look at any number of examples of something like a stop sign and know that they're all stop signs, but a computer must have every stop sign individually labelled.

Machine Learning Limitations

- **REMEMBER Machines can't explain themselves:** As machine learning solutions become more flexible and capable; the amount of hidden functionality becomes greater as well.
  - In fact, when dealing with deep learning solutions, you find that the solution contains one or usually more hidden layers that the solution creates but that humans haven't taken the time to explore.
- **Bias makes the results less usable:** An algorithm can't tell when **data contains various mistruths** in it.
  - Consequently, it regards all data as being unbiased and completely truthful.
  - As a result, any analysis performed by an algorithm trained using this data is **suspect**
- **Machine learning solutions can't cooperate:** One of the most important advantages of being human is the ability to collaborate with others.
  - Knowledge potential increases exponentially as each party to a potential solution submits its piece of knowledge to create a whole that is much greater than the sum of its parts.
  - A single machine learning solution remains a single machine learning solution because of it can't generalize knowledge and thereby contribute to a comprehensive solution with multiple cooperative parties.

Gartner Top 10 Data and Analytics Trends, 2021

Accelerating Change

1 Smarter, Responsible, Scalable AI

2 Composable Data and Analytics

3 Data Fabric Is the Foundation

4 From Big to Small and Wide Data

Operationalizing Business Value

5 XOps

6 Engineering Decision Intelligence

7 D&A as a Core Business Function

Distributed Everything

8 Graph Relates Everything

9 The Rise of the Augmented Consumer

10 D&A at the Edge

gartner.com/SmarterWithGartner

Gartner.

Top big data trends in 2021

Edge computing

Explosive growth in data generated from cloud systems, sensors, smart devices and video streaming is driving adoption of edge computing. Data processing is done on the periphery of the network as close to the originating source as possible.

Cloud and hybrid cloud computing

Cloud computing enables organizations to process nearly limitless amounts of data. Hybrid cloud approaches are being developed to enable companies in regulated industries to take advantage of cloud's economic and technical advantages.

Data lakes

These large repositories store structured and unstructured data in its native format. Data scientists often extract just what's needed for a project, eliminating costly ETL processes required of centralized data warehouses.

Machine learning and AI technologies

Machine learning and other AI technologies are revolutionizing big data analytics. AI's ability to ingest and analyze massive amounts of structured and unstructured data is being used by companies to optimize and improve business operations.

Big Data, Machine Learning and Deep Learning Trends

- https://towardsdatascience.com/these-are-the-big-data-trends-2020-49c4db330ba1

- https://www.gartner.com/smarterwithgartner/gartner-top-10-trends-in-data-and-analytics-for-2020/

- https://www.crn.com/news/cloud/5-emerging-ai-and-machine-learning-trends-to-watch-in-2021