جامـــــعة المـــــستـقبـل
AL MUSTAQBAL UNIVERSITY

كلية العلوم

قســـم الانظمة الطبية الذكية

Lecture: (2)

Data-Information-Knowledge (DIK Hierarchy)

Subject: Artificial Intelligence Principles
Level: Third
Lecturer: Prof. Dr. Mehdi Ebady Manaa

1. Data-Information-Knowledge (DIK Hierarchy)

The Data-Information-Knowledge-Wisdom (DIKW) pyramid illustrates the relationship between data, information, knowledge, and wisdom. Its primary goal is to provide valuable insights derived from data.

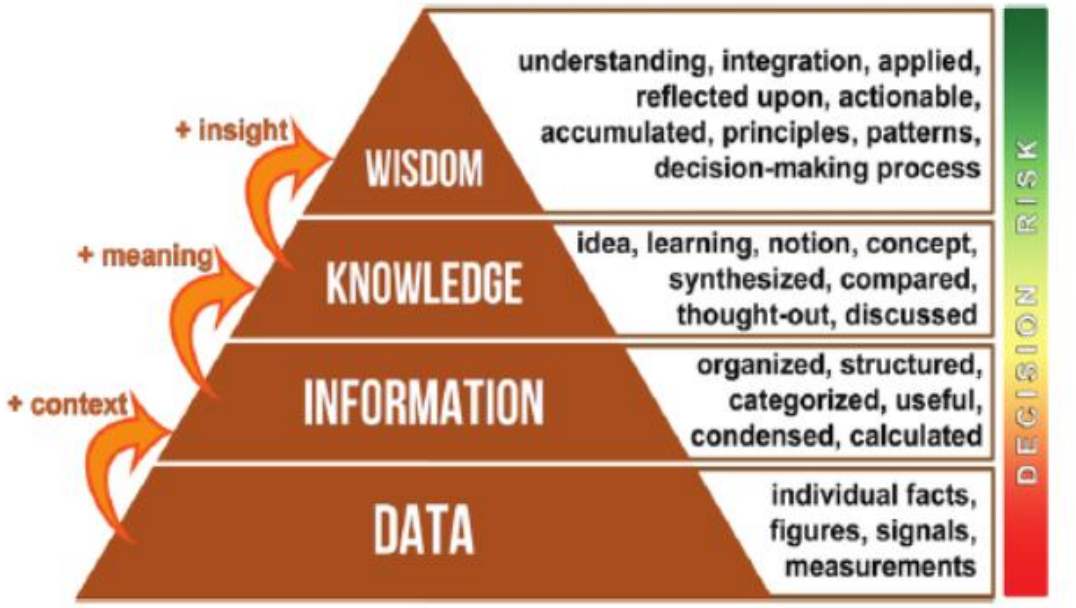

Figure 1: The Data-Information-Knowledge Pyramid

The DIKW model is widely used in data science and big data analytics to understand how data is transformed into meaningful context. By analyzing and converting data into actionable insights, it helps decision-makers make accurate and informed decisions.

The DIKW model consists of four levels: data, information, knowledge, and wisdom. Each level is built upon the next level where the flow starts from top level to bottom to up. The more details about each level can be seen below .

A. Data: Data refers to unprocessed facts and figures that lack context. In its raw form, data holds limited value. Tasks at this level include data collection, processing, and standardization. Data can be classified into three types. Figure (2) shows the main classification of data types.

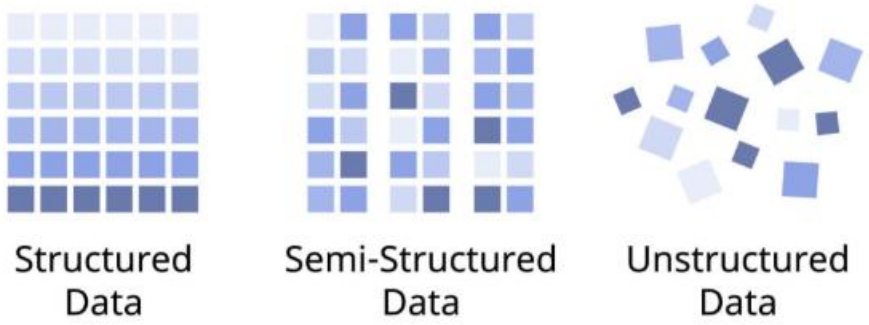

Figure 2: Types of Big Data

B. Structure _Data

The structure data is organized in a meaningful way to make it readable by both humans and machines. It is represented as discrete fields in a tabular format, allowing access either individually or aggregated with other fields.

C. Unstructured Data

Unstructured data refers to unorganized information that lacks a predefined data model. It includes various types, such as text, images, videos, sensor

data, web pages, and social media content. A key characteristic of unstructured data is its rapid growth rate, which largest size than the structure data. Unstructured data is usually stored in a non-relational database also known as NoSQL Database

D. Semi Structure Data

The semi structure data is partially organized in structure format, but it doesn't have a fixed scheme. XML, Emails, Zip file and JSON are common examples of semi structure data. They contain the metadata that provides additional information about the scheme of data.

The main distinctive features of semi structure is the ability to process data faster rather than structure data. This is achieved by indexing data in a more dynamic way. Information: The Information is raw data that has been organized, processed, and interpreted. This process involves data analysis, collection, extraction, and visualization. Information provides context, meaning, and valuable insights that guide decision-makers in taking the correct actions. It answers the questions of "who," "what," "where," and "when".

E. Knowledge: It analyses the information to explore and interpreting the trends, and relationship.it require expertise to provide justification of "why" and how the phenomena happen.

F. **Wisdom:** The ability to apply knowledge effectively in order to understand customer preferences and market trends. This wisdom helps decision-makers make accurate decisions about the target audience and allocate resources to increase user engagement.

2. Life Cycle Process for Converting the Data to Information

Figure (3) shows the data lifecycle scheme

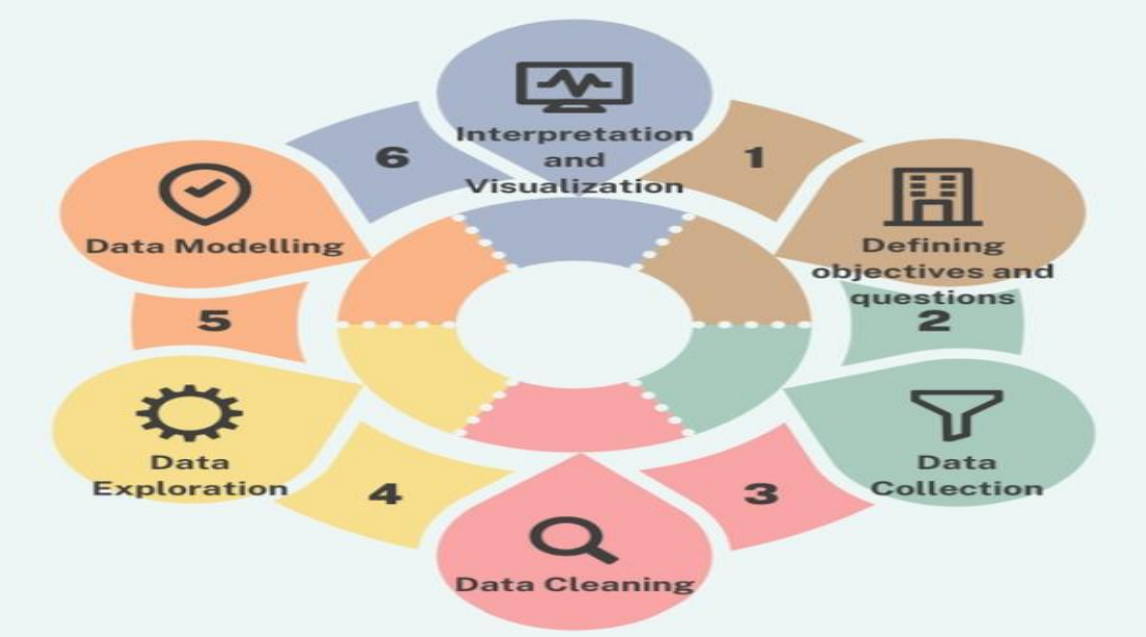

Figure 2: Data Lifecycle Scheme

A) Data _Collection

- The data can be collected from various sources, such as satellite sensors, cameras, and social media posts, and provide valuable information
- The Internet of Things (IoT) sensors is the main resource that has been used
- to gather data for everyday objects, such as water pumps.

- This data helps predict failures before they disrupt water supplies, ensuring that communities continue to have access to essential resources.

B) Data Processing

- Data Processing can be divided into part data cleaning and data transformation.
- Data preprocessing is an important step in machine learning as it converts data into a readable format.
- Organizations can gain actionable insights from data to improve business strategy, enhance operational planning, and forecast market trends.
- Proper data processing has a significant impact on the reliability and accuracy of AI applications. Poorly processed data can lead to bias and unreliable results.

A.Data Cleaning

Data cleaning is the process of removing the incorrect, corrupted, duplicated, missing data, and outline. The process of cleansing data will vary from database to other.

The incorrect data leads to unreliable machine learning results even though the result looks correct. Data cleansing gives the decision maker deep insight into comprehensive information, leading to enhance the productivity of organization.

B. Data Transformation

In Python, data transformations are implemented using two techniques. The first technique is normalizing the data to follow a normal distribution. This is done by utilizing Min-Max scaling which transfers the values to a range between 0,1.

The second technique is standardization where data is transferred to have zero mean ($\mu = 0$) and standard deviation of 1.

C.Data Analysis

- The analysis can offer plenty of advantages to business such as better products, more quickly and improve companies overall efficiency. The insights gained from such analysis can significantly influence a company's success.

- There are various types of analysis that businesses can employ to drive innovation, improve operational efficiency, and reduce risks. The choice of analysis depends on the specific problem to be addressed or the questions that need to be answered. For instance, trend analysis is a valuable tool for examining performance indicators, such as Key Performance Indicators (KPI) over time.

- Descriptive Statistics summarize and analyses the main features of databases without making inference such as measure the center tenancy (mean, median, mode), measure dispersion (standard deviation, variance).Data Visualizations

- Visualization techniques are a process to represent data in visual elements, which is very important step to help the decision maker draw the relevant and actionable information.

- The most common ways to represent data visually are charts and graphs. There are many types of charts used to present information graphically. It is crucial to ensure that the type of visualization used best illustrates the pattern, problem, or trend that needs highlighting; not all types of visualizations are appropriate for all kinds of data.

- There are various types of visualization tools (Line Charts, Column Charts, Bar Charts, Pie Charts)

- There are many factors considered when choosing a visualization
  - The number of variables, which are the characteristics measured, that need is shown
  - The number of data points, or units of information, in each variable
  - Whether the data illustrates changes over time (hourly, daily, weekly)
  - The need to make a comparison or correlation between different data
  - points.

3. DIK in Real-Use case

EXAMPLE#1

Cosmetic shop has 500 customers who need to make sense of their data and to use it to identify trends, make decisions, and predict future conditions. The market manager wants to increase the percentage rate. The market manager is employed the DIK framework.

Data:

- ❖ Costumer purchase history
- ❖ Costumer Demographics features (Address, age, height, weight, educational _level)
- ❖ Social media post review
  - Seasonal sales trend

Information

- ❖ Top-purchase products for location
- ❖ Customer engagement level in social media

- Customer segments based on group age
- Purchase fluctuations rate based on educational _level

Knowledge

- -The center of the city is the occupy the highest purchase rate
- Focused on high liked post type
- Create age-specific promotions(discount)
- -Seasonal sales trends

Wisdom

Understanding these patterns helps the cosmetic shop make informed decisions regarding customer segments and engagement levels, as consequence, it would enhance the shop's net revenue.

Example 2:

An AI startup has launched a solution to detect deep-fake posts, sounds, and videos. However, the sales rate has been recorded as very low. The

product manager has decided to take action to improve sales. To assist in boosting sales, the product manager can utilize the DIK (Data, Information, Knowledge) framework.

Data:

- The sale data
- Customer feedback
- Market trend
- Social media review

Information:

- Group most popular AI solution for deep fake detection according to region
- Average features usage frequently in market trend
- Computing the SD for positive and negative social media review

Knowledge

- Interpreting AI platform usage through exploring the user feedback

- Discover the relationship between the features frequently and the sale percentage.

- Explore uncovered pattern of sale marketing rate according to region

Wisdom

-Understanding these patterns helps the CEO make informed decisions about methods to increase purchase rates, which, in turn, can enhance company profit.