



SECP2753 Data Mining

Group Task 2

Prepared by group 9

Ahmad Ziyaad A23CS0206

Goe Jie Ying A23CS0224

Teh Ru Qian A23CS0191



Q1 - What is Knowledge Discovery in Databases (KDD)?

Definition 1:

Knowledge Discovery in Databases, commonly referred to as KDD, is a systematic approach to uncovering patterns, relationships, and actionable insights from vast datasets. [4]

Definition 2:

Knowledge discovery in databases (KDD) is one proper methodology to analyze and understand such huge amounts of data. [5]

Definition 3:

Knowledge discovery in databases (KDD) is entire process of extracting knowledge and using interesting patterns in databases. [9]

Q2 - What are the steps in the KDD process?

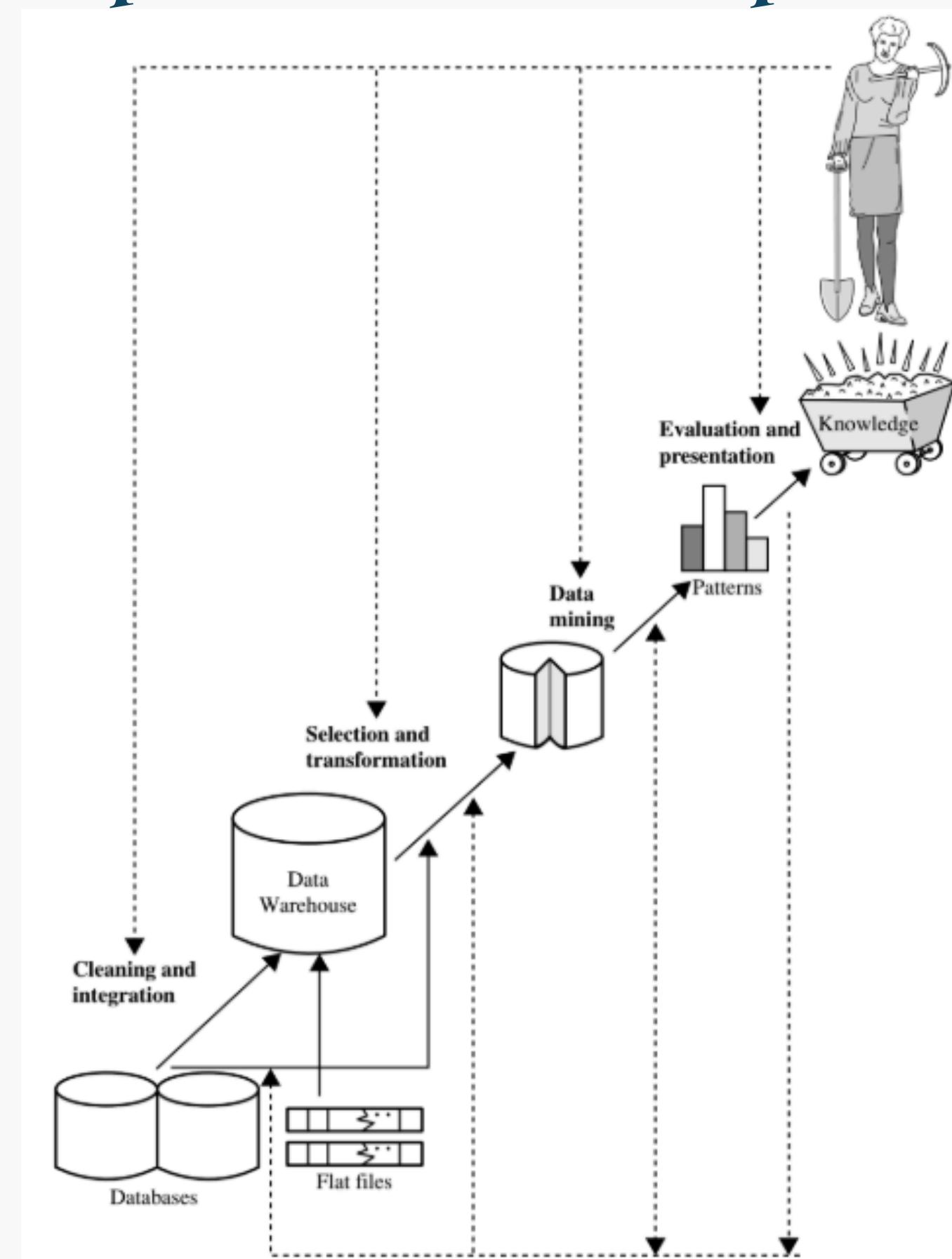


Figure 1: Data mining as a step in the process of knowledge discovery [6]

Q2 - What are the steps in the KDD process? [9]

- Data cleaning - **remove noise or irrelevant data**
- Data integration - combine **multiple** data sources
- Data selection - **extract** relevant data needed for analysis
- Data transformation - **convert, aggregate, or summarize** data to fit analysis needs
- Data mining - **extract** data **patterns**
- Pattern evaluation - **identify** the discovered patterns
- Knowledge presentation - **visualize and interpret** extracted **knowledge** for users

Q3 - How does KDD differ from Data Mining? [9] ••••

Aspect	KDD	Data Mining
Definition	Entire process of extracting knowledge	The steps in applying mining algorithms during the process of knowledge discovery in databases
Scope	Include multidiscipline, such as statistics, artificial intelligence, machine learning, database science, and information retrieval	Applying algorithms to extract patterns from data
Analogy	Making a burger	Following the specific steps (e.g., following McDonald's or Burger Ramli's recipe)



Q4 - What is Data and Insights?

Definition 1:

Data refers to unanalyzed user observations, while insights are the actionable opportunities based on research and business goals. [10]

Definition 2:

Data is a collection of facts, representing measurements or descriptions of a specific situation. Insight refers to an analyst or business user discovering a pattern in data or a relationship between variables that they didn't previously know existed. [11]



Q5 - What is the difference between data and insights?

.....

Aspect	Data	Insights
Definition	A collection of facts [11]	Discovering patterns in data [11]
Characteristics	An unanalyzed collection of observations about users that may include transcripts, notes, metrics, or survey output [10]	Focused explanations of opportunities, based on other user research and business context [10]
Usage	Needs to be put into context and processed into information before it can become useful [12]	Reveal new patterns, relationships, and possibilities, and allow us to predict the future [12]



Q6- What is CRISP-DM?

Business Understanding

- Understand the project objectives and requirements from a business perspective.

Data Understanding

- Collect data and get familiar with it, identify data quality issues, and discover initial insights.

Data Preparation

- Clean and transform data into a suitable format for modeling.

Modeling

- Select and apply modeling techniques, and fine-tune model parameters.

Evaluation

- Assess the model's performance and ensure it meets business objectives.

Deployment

- Deliver the results in a usable form, such as reports or predictive systems.



Q7- How does Information Retrieval differ from Data Mining?

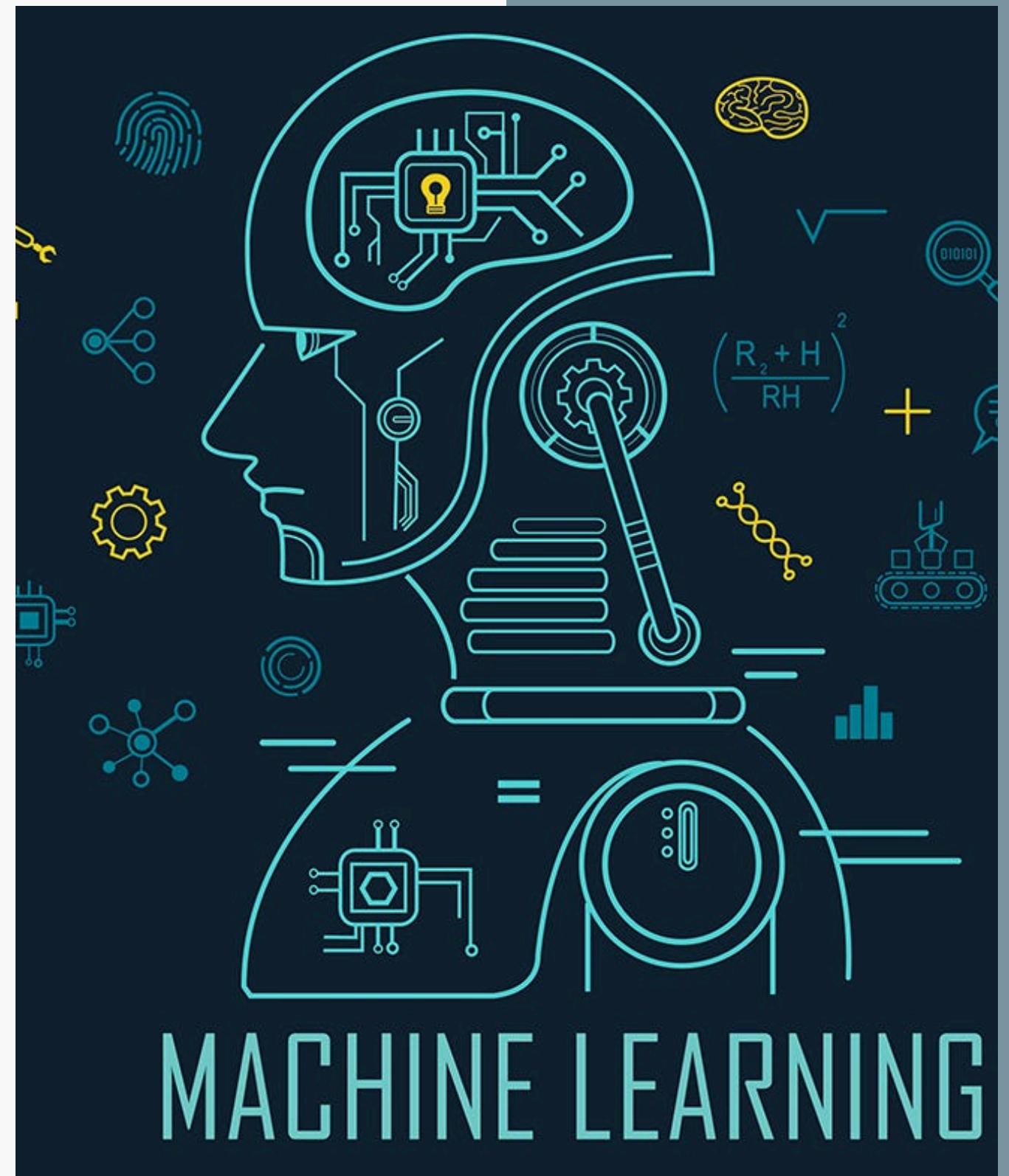
Information Retrieval	Data Mining
Focuses on finding and retrieving relevant information from large datasets (like documents or websites) based on user queries.	Focuses on discovering hidden patterns, correlations, and trends in large datasets using statistical and machine learning techniques.
Deals mainly with unstructured or semi-structured data such as text, images, or multimedia.	Works primarily with structured data like databases or spreadsheets.
Goal is to find specific information that matches a query.	Goal is to extract useful knowledge from data, often without a specific query.

Q8- What is Machine Learning?

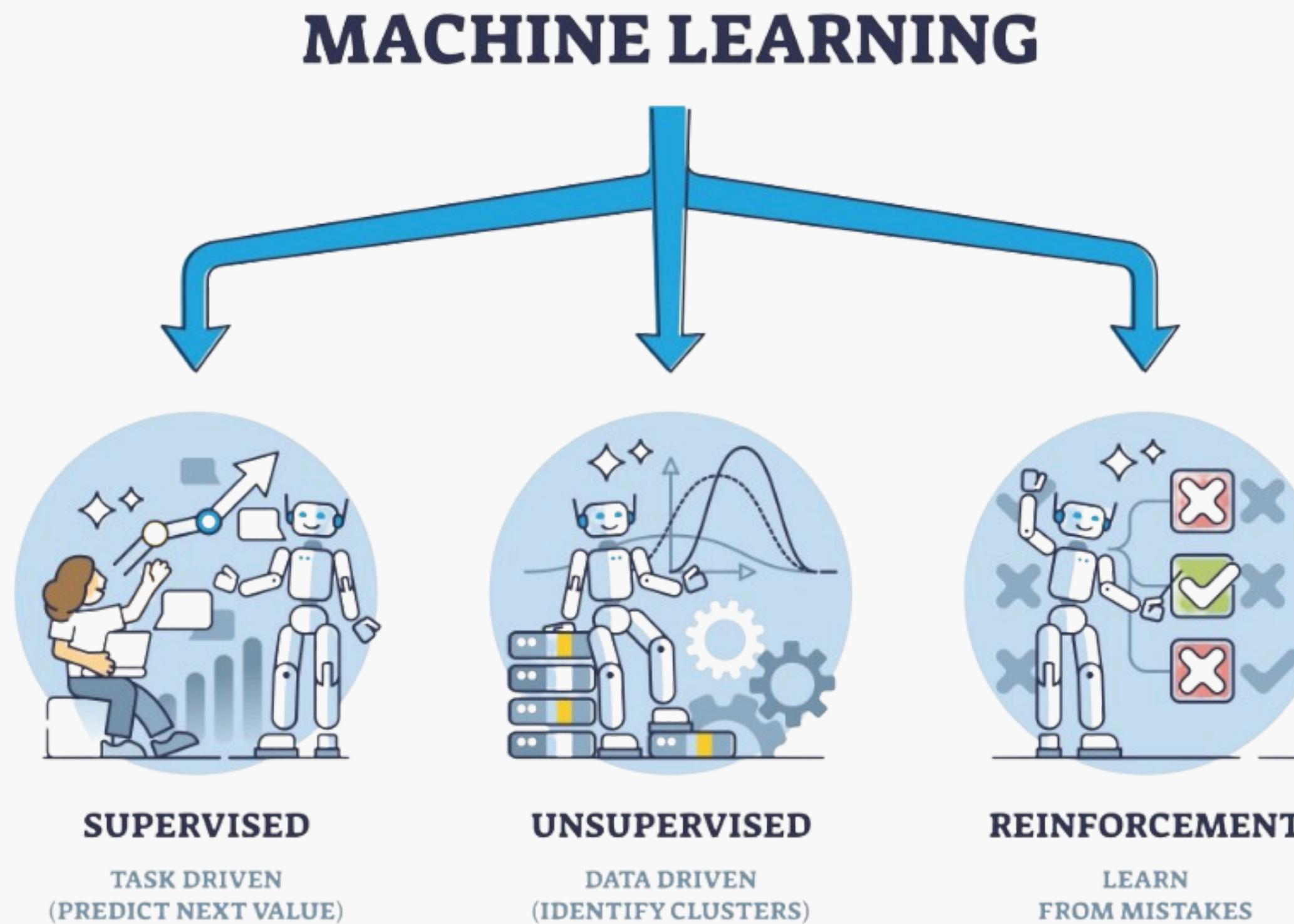
Machine Learning:

Machine Learning is a branch of artificial intelligence (AI) that enables computers to learn from data and improve their performance on a task without being explicitly programmed.

It involves developing algorithms that can identify patterns in data and make decisions or predictions based on it.



Q9- What are the main types of Machine Learning?



Q1o- How is DM different from ML?

Data Mining	Machine Learning
The process of discovering patterns, correlations, and insights from large datasets.	A method of teaching computers to learn from data and make predictions or decisions.
Focuses on extracting useful information from data.	Focuses on learning algorithms that can improve from experience.
Often uses machine learning techniques as tools.	Machine learning can be used independently to build predictive models.

Q11 - What is Artificial Intelligence (AI)?

Artificial Intelligence (AI) is the field of computer science focused on creating systems capable of performing tasks that typically require human intelligence, such as learning, reasoning, and problem-solving. [1]

AI can be classified into 2 categories [1] :

- narrow AI - designed for specific tasks
- general AI - perform any intellectual task that human can.



Q12 - What is the difference between ML and AI? ••••

Aspect	AI [1] [8]	ML [1] [8]
Definition	AI creates machines that perform tasks needing human intelligence.	ML enables machines to learn from data to make decisions.
Scope	AI includes machine learning and other technologies.	ML is a subset of AI focused on data-driven learning.
Goal	AI simulates human-like intelligence.	ML aims to improve performance through data.
Techniques Used	AI uses learning, logic, planning, and more.	ML uses algorithms like supervised and unsupervised learning.



Q12 - What is the difference between ML and AI? ••••

Continue.....

Aspect	AI	ML
Example	Chatbots, voice assistants (Siri).	Spam filters, recommendations (Youtube).
Data Dependency	AI can use rules or data.	ML depends on large amounts of data to learn.
Human Involvement	AI can work autonomously once set up.	ML requires human input for data and evaluation.



*Q13 - What are some **real-world applications** of AI and potential careers in this field?*



Healthcare [2]

- AI-driven radiology systems have reduced diagnostic errors by 30%, aiding early disease detection.

Finance [2]

- AI-powered fraud detection systems have decreased fraud losses by 45%, helping financial institutions identify and prevent fraudulent activities.



Retail [2]

- AI-driven systems can boost revenue by 20% by analyzing purchasing behavior to suggest products tailored to individual preferences.

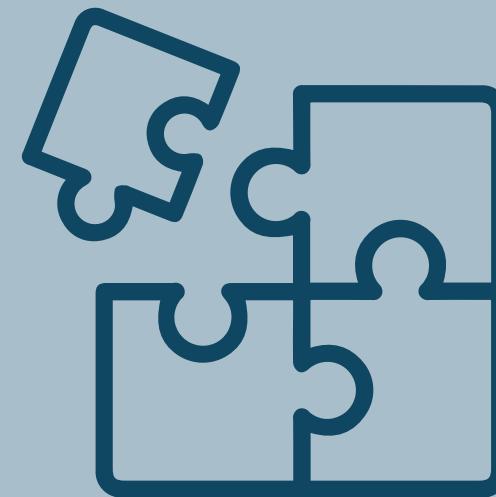
Q13 - What are some real-world applications of AI and potential careers in this field?

Continue.....



AI Research Scientist [3]

- Average salary of \$143,184 with 17,940 open positions.



Machine Learning Engineer [3]

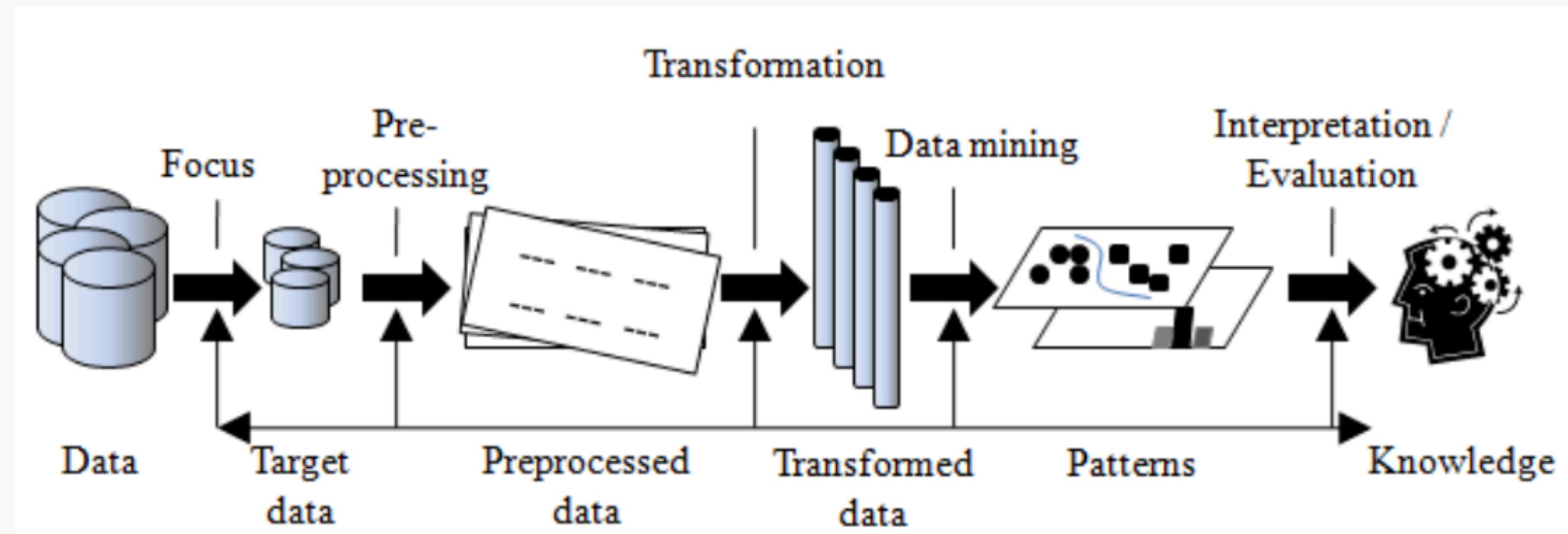
- Average salary of \$118,859 with 6,817 open positions.



AI Engineer [3]

- Average salary of \$105,000 with 4,352 open positions.

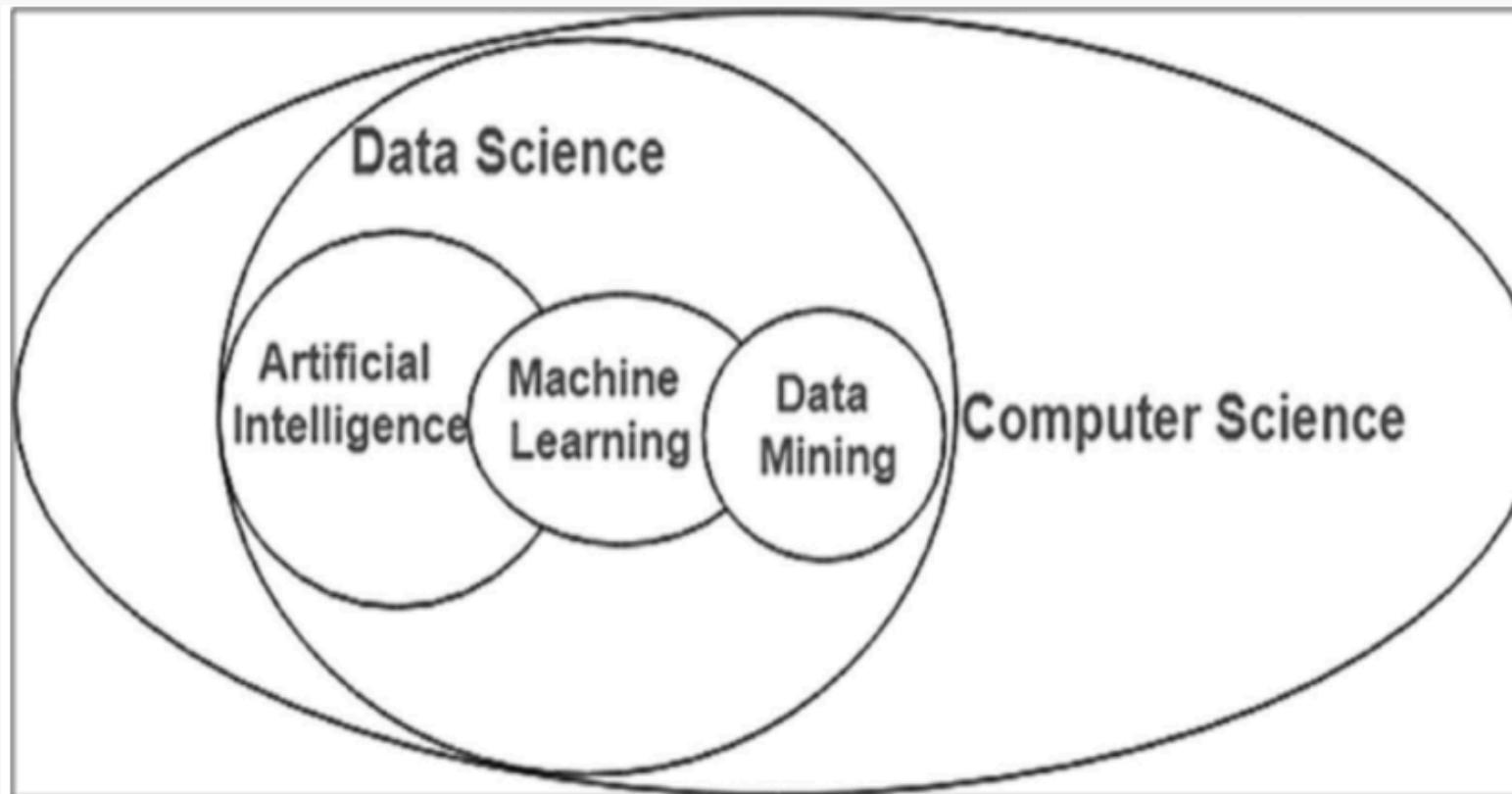
Q14 - Provide a figure that illustrates the relationships among KDD, DM, ML and AI.



- KDD - **process** that discovering useful knowledge from data. [7]
- DM - one of the **key steps** within the KDD process. Process of analyzing large datasets to identify patterns and it plays a central role in turning raw data into useful knowledge.

Q14 - Provide a figure that illustrates the relationships among KDD, DM, ML and AI.

Continue.....



- AI includes **Machine Learning** (ML) and **Data Mining** (DM) as subfields. [8]
- **Machine Learning** (ML) is often used within **Data Mining** (DM) techniques.
- **Data Mining** (DM) is a key step in the **KDD** process, which includes multiple stages for preparing, analyzing, and interpreting data.

Reference

- [1] Russell, S., & Norvig, P. (2016). Artificial Intelligence: A Modern Approach (3rd ed.). Pearson Education.
https://api.pageplace.de/preview/DT0400.9781292153971_A27091185/preview-9781292153971_A27091185.pdf
- [2] Samradni, “Top 5 Applications of Artificial Intelligence,” Analytics Insight, Nov. 07, 2024.
<https://www.analyticsinsight.net/artificial-intelligence/top-5-applications-of-artificial-intelligence> (accessed Apr. 05, 2025).
- [3] P. Peralta, “10 most in demand AI jobs for 2024,” Digital Insurance, Jan. 02, 2024. <https://www.digin.com/list/these-10-ai-careers-are-the-most-in-demand> (accessed Apr. 05, 2025).
- [4] S. Chumbar, “Knowledge Discovery in Databases (KDD): A Practical Approach,” Medium, Sep. 24, 2023.
<https://medium.com/@shawn.chumbar/knowledge-discovery-in-databases-kdd-a-practical-approach-f28247493be4>
- [5] “knowledge discovery in database - an overview | ScienceDirect Topics,” Sciencedirect.com, 2016.
<https://www.sciencedirect.com/topics/computer-science/knowledge-discovery-in-database>
- [6] jiawei han, micheline kamber, and jian pei, “Data Mining Third Edition,” 2011. Available:
<https://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>

Reference

- [7] Cifci, Mehmet Akif & Hussain, Sadiq. (2018). Data Mining Usage and Applications in Health Services. JOIV : International Journal on Informatics Visualization. 2. 10.30630/joiv.2.4.148.
https://www.researchgate.net/figure/Knowledge-discovery-in-databases_fig1_326022435
- [8] “The Relationship between Data Mining, Machine Learning, and Artificial Intelligence,” Ebrary, 2018.
https://ebrary.net/190208/health/relationship_data_mining_machine_learning_artificial_intelligence (accessed Apr. 06, 2025).
- [9] SECP2753 Data Mining Lecture Slide Module 1b_2024 Introduction to Data Mining_Part2
- [10] W. L. in R.-B. U. Experience, “Data vs. Findings vs. Insights: The Differences Explained,” Nielsen Norman Group, Apr. 23, 2023. <https://www.nngroup.com/articles/data-findings-insights-differences/>
- [11] “What are Data Insights: Definition & Best Practices.” Qlik. <https://www.qlik.com/us/data-analytics/data-insights>
- [12] “The Difference Between Data, Information, and Insight - A Data Pro,” adata.pro, Jul. 07, 2022.
<https://adata.pro/blog/the-difference-between-data-information-and-insight/>