

# Intelligent educational data management – The unsupervised learning approach

Prof. Cleo Sgouropoulou

Department of Informatics and Computer Engineering  
University of West Attica



HELLENIC MINISTRY OF  
Education, Religious Affairs and Sports



Co-funded by the  
Erasmus+ Programme  
of the European Union

# Cleo Sgouropoulou



Professor in the Department of Informatics and Computer Engineering,  
University of West Attica

Board of Directors, University of West Attica



Board of Directors, National Infrastructures for Research & Technology

National Delegate, European Committee for Standardization



CEN/TC 353 "ICT for Learning, Education, and Training"

CEN/TC428 "ICT Professionalism and Digital Competences"



ISO JTC1 SC36 "Information Technology for Learning, Education & Training"



Convener TC48-WG3, Hellenic Mirror Committee – ICT for LET



Director of the Educational Technology and eLearning Systems Lab



# EduTEL main research initiatives

eduTEL



Adaptive  
Educational  
Hypermedia  
Systems

Personalization &  
AI in Education

Chatbots & Large  
Language Models

Digital Skills

Virtual &  
Augmented Reality

Educational Data  
Mining & Learning  
Analytics

Online Learning  
Platforms &  
MOOCs

Learning Design

Gamification &  
Game-based  
Learning

Open Educational  
Resources

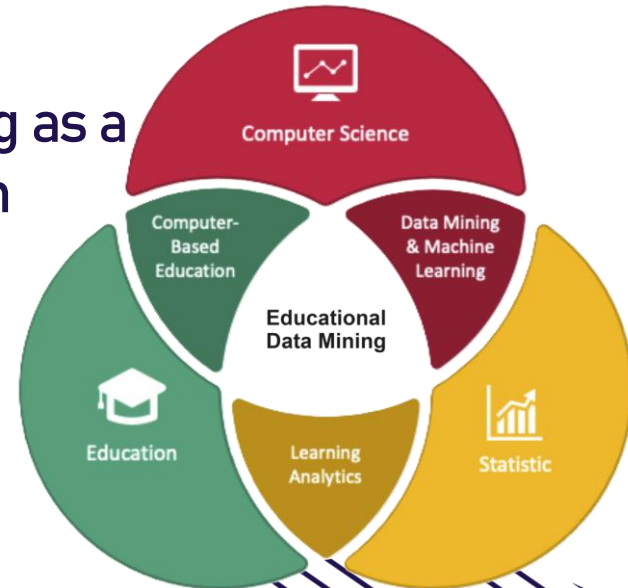
M-Learning,  
U-Learning &  
Wearable  
technologies

Educational  
Robotics



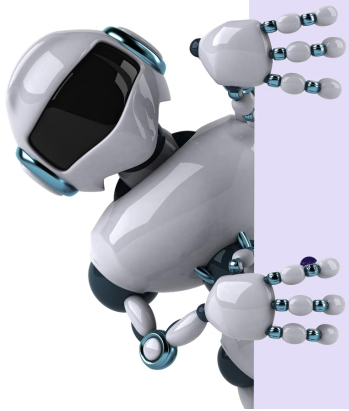
# Overview

- Overview of Artificial Intelligence and Machine Learning
- Principles and Practices of Intelligent Educational Data Management
- Examples of exploring Descriptive Clustering as a Guide to Unsupervised Learning in Education with ChatGPT



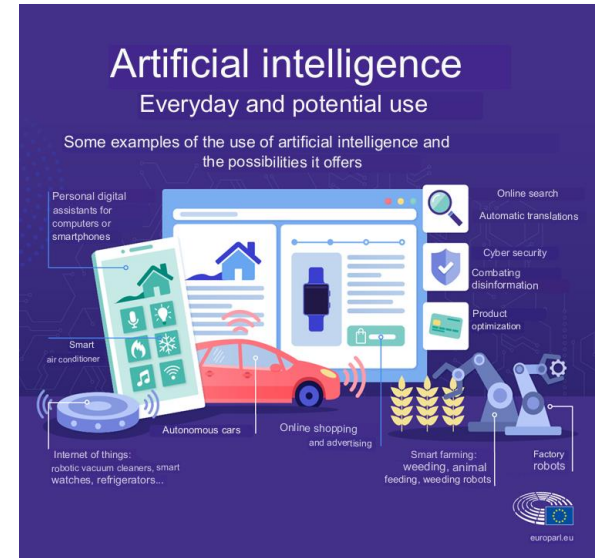
# Artificial Intelligence (AI)

- AI mimics human cognitive functions, such as learning and creativity
- AI enables machines to understand environments, solve problems, and pursue goals by processing data from sources like cameras
- AI systems adapt their behavior by analyzing outcomes of past actions and autonomously solving problems



# Evolution and Impact of AI

- Some AI Technologies have been around for more than 50 years
- Factors Contributing to AI Development:
  - Evolution of computers
  - Availability of countless data
  - Advancements in new algorithms
- Impact and Future Expectations:
  - Already affecting our daily lives
  - Expected to bring huge changes
  - Will play a central role in the digital transformation of our society



# AI-Enhanced Data Management

- **Diverse Data Access**  
Access to various data types: text, images, videos, sensor data
- **AI-Driven Insights**  
Leverage AI to analyze and extract insights from vast data collections
- **Enhanced Decision-Making**  
Utilize AI to process large volumes and make informed decisions
- **Pattern Recognition**  
Discover hidden patterns, relationships, and trends using AI techniques
- **Handling Data Complexity**  
AI's ability to manage both structured and unstructured data
- **Innovation and Improvement**  
Combine AI with data to drive innovation and enhance decision-making



# AI for Knowledge Mining





# Machine Learning

- Machine learning, a branch of computer science, evolved from AI's pattern recognition and computational learning.
- It involves designing algorithms that learn from data to make predictions or decisions.
- Closely related to computational statistics and mathematical optimization, influencing its methods and applications.

## What is ... ?



### Machine Learning

Machine learning is the field behind a great many of the artificial intelligence programs that we encounter in daily life right now.

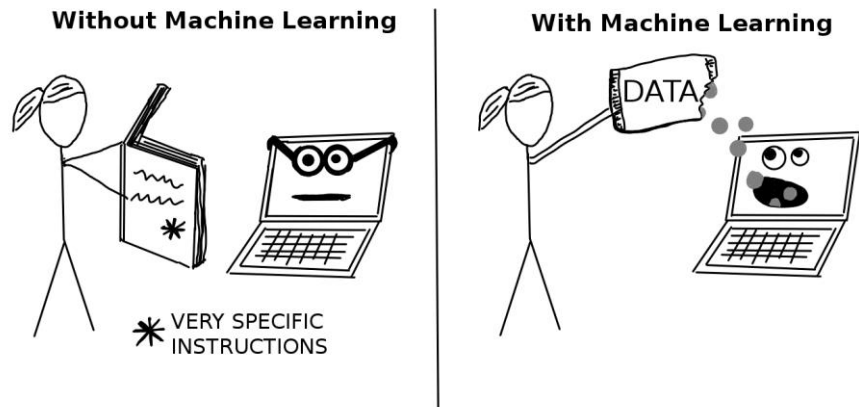
It's a method that AI tools use to acquire new information.

Machine learning gives AI tools the ability to learn without being explicitly taught or programmed with new information, which makes all kinds of other things possible.

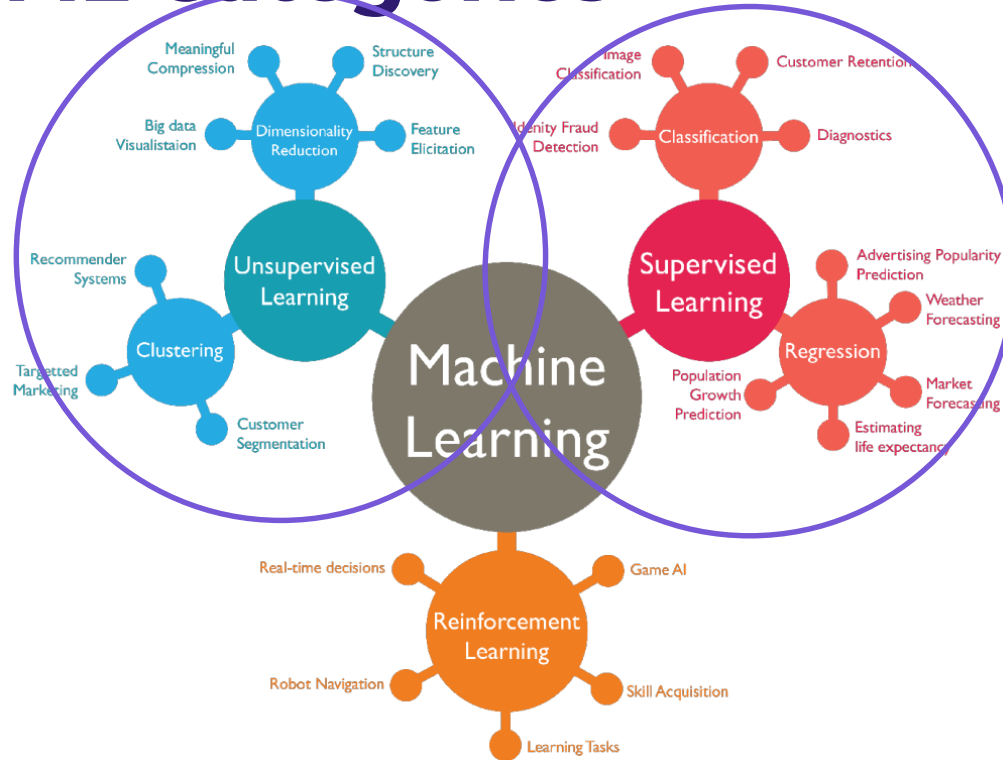


# Machine Learning

- Machine learning used for complex computational tasks.
- Applications include spam filtering, character recognition, search engines, and computer vision
- Develops predictive models and algorithms in data analysis
- Enables reliable decision-making and uncovers data trends and interrelationships



# Basic ML Categories



# Supervised vs Unsupervised

In **supervised learning**, we are provided with a dataset where each record is labeled with the corresponding class (group)

The goal is to create a model that can accurately classify new data into one of the pre-existing classes when presented with new instances

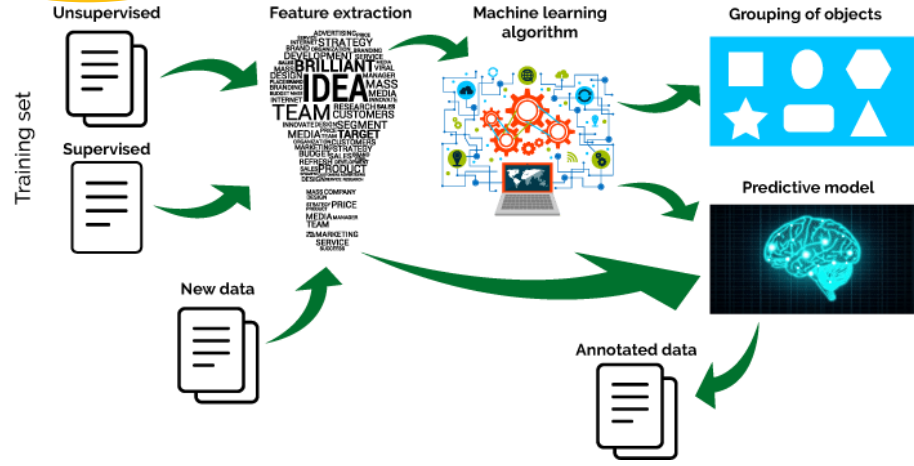
In **unsupervised learning**, we are given a dataset without any labels indicating the class of each record.

The goal is to analyze this data to discover potentially interesting patterns or facts, such as grouping similar records into clusters

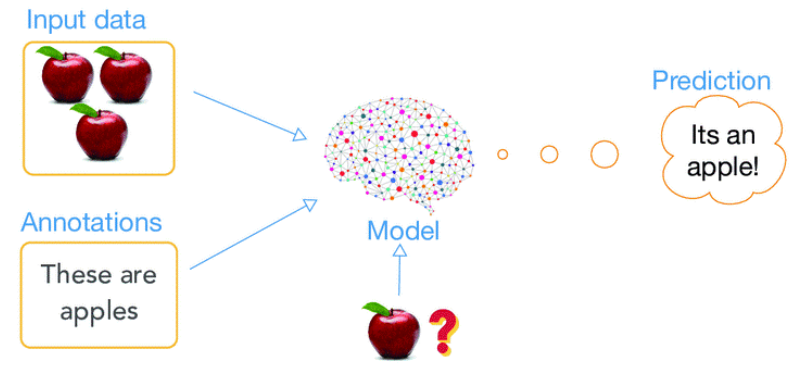
A paradigm of unsupervised learning is data mining



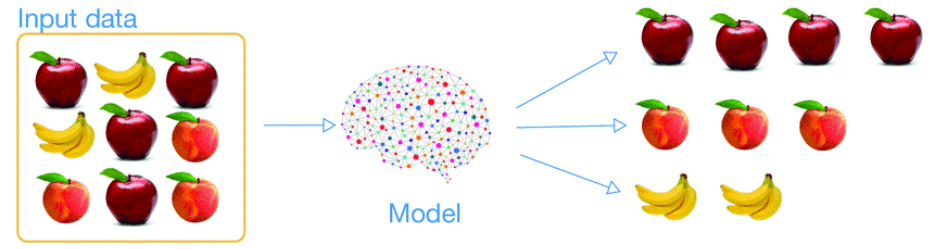
# Machine Learning



## supervised learning



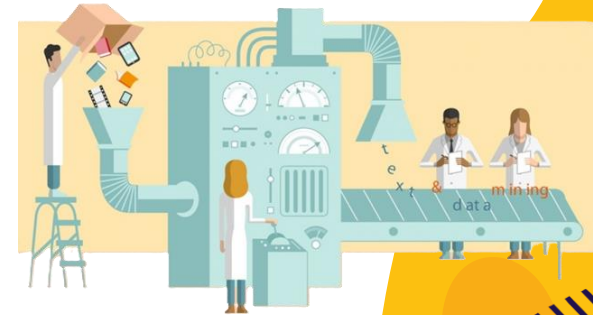
## unsupervised learning





# Why Data Mining in Education?

- Big data analysis enables educational institutions to explore complex datasets where traditional methods would be difficult or impossible
- Educational research teams utilize data analysis to study and accurately understand educational phenomena.
- Educational institutions can analyze student behavior and preferences to design more effective educational strategies, enhancing learning outcomes and potentially reducing operational costs



# Why Data Mining in Education?





- Analyze observed trends to enhance educational outcomes
- Identify extreme behaviors in student interactions that may indicate potential issues
- Apply predictive analytics to assess student performance and learning needs
- Conduct case checks to ensure accuracy in student records and assessments
- Identify potential learning groups based on student abilities and preferences
- Determine connections between various educational concepts and student learning patterns

*How*  
**EDUCATIONAL DATA MINING & LEARNING ANALYTICS**  
*can help:*





Educational data mining focuses on developing new tools and algorithms for discovering data patterns

Learning analytics focuses on applying tools and techniques at larger scales in instructional systems

**EDUCATIONAL DATA MINING CAN ANSWER QUESTIONS LIKE:**

-  What sequence of topics is most effective for a specific student?
-  Which student actions are associated with better learning and higher grades?
-  Which actions indicate satisfaction and engagement?
-  What features of an online learning environment lead to better learning?

**LEARNING ANALYTICS CAN ANSWER QUESTIONS LIKE:**

-  When are students ready to move on to the next topic?
-  When is a student at risk for not completing a course?
-  What grade is a student likely to receive without intervention?
-  Should a student be referred to a counselor for help?



# Knowledge Extraction Process

**Knowledge Extraction: A two-way, iterative process**

Educators and administrators often start without a precise idea of the needed information

**Initial Findings Prompt Further Inquiry**

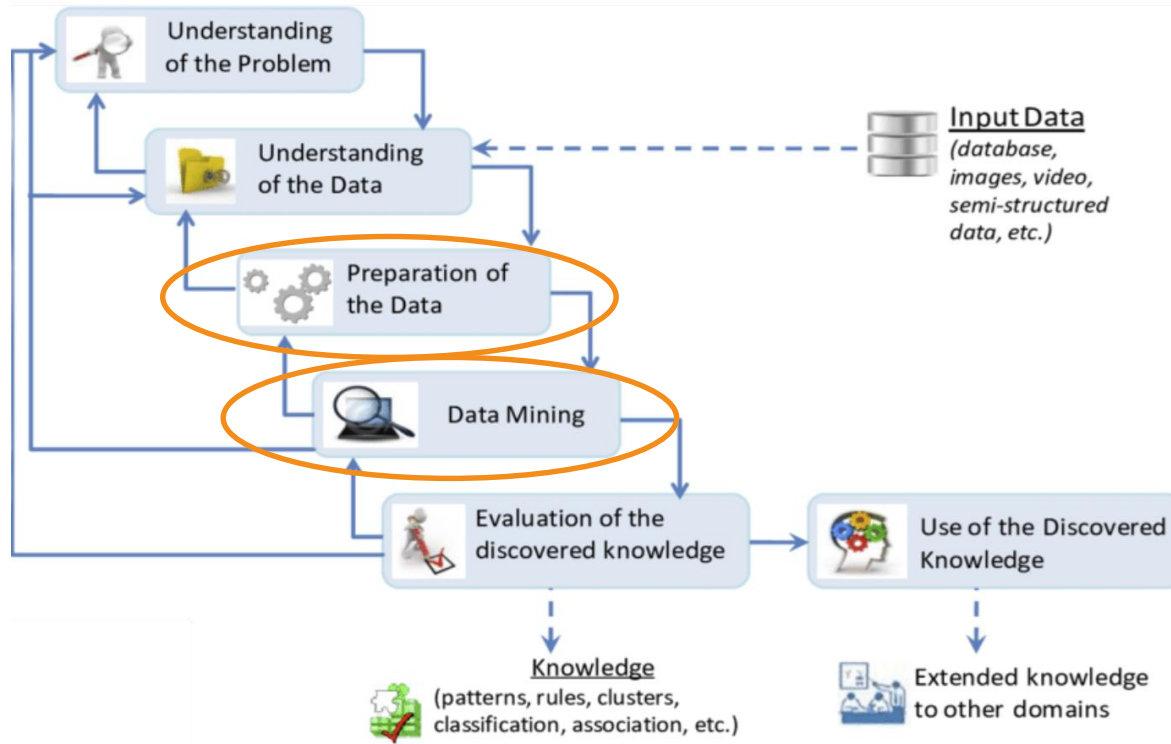
Conclusions from educational data analysis typically generate new questions and prompt additional investigations.

**Adapting Analysis Methods**

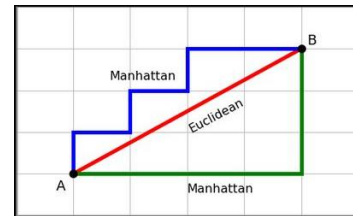
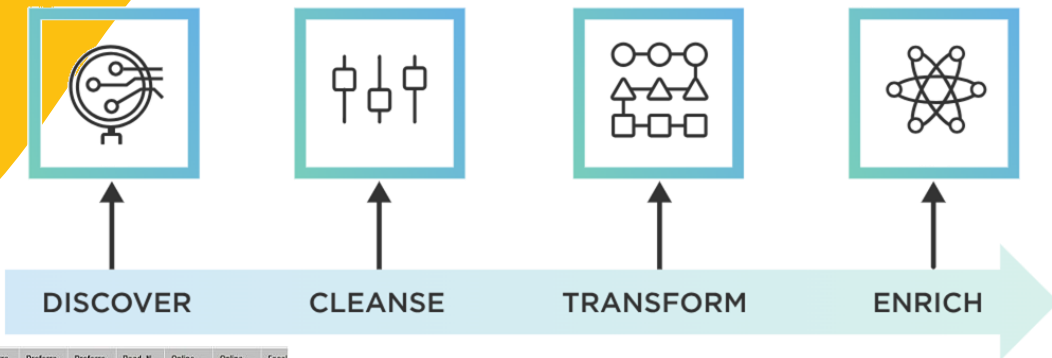
If data analysis results in educational settings are not effective, the process may need to be redesigned to align more closely with student and teacher needs



# Knowledge Extraction Process

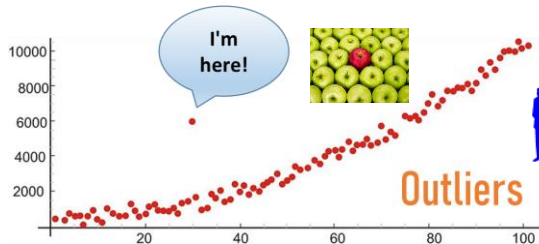


# Data Preparation and Transformation



1	Gender	Race	Birth_Year	Marital_	Years_o...	Hours_P...	Preferre...	Preferre...	Preferre...	Read_N...	Online_...	Online_...	Facel
2	M	White	1972	M	8	1	Firefox	Google	Yahoo	Y	N	N	Y
3	M	Hispanic	1981	S	14	2	Chrome	Google	Hotmail	Y	N	N	Y
4	F	African A...	1977	S	6	2	Firefox	Yahoo	Yahoo	Y	Y		Y
5	F	White	1961	D	8	6	Firefox	Google	Hotmail	N	Y	N	N
6	M	White	1954	M	2	3	Internet...	Bing	Hotmail	Y	Y	N	Y
7	M	African A...	1982	D	15	4	Internet...	Google	Yahoo	Y	N	Y	N
8	M	African A...	1981	D	11	2	Firefox	Google	Yahoo		Y		Y
9	M	White	1977	S	3	3	Internet...	Yahoo	Yahoo	Y			Y
10	F	African A...	1959	M	6	2	Firefox	Google	Gmail	N	Y	N	N
11	M	White	1987	S	12	1	Safari	Yahoo	Yahoo	Y		Y	Y
12	F	Hispanic	1959	D	12	5	Chrome	Google	Gmail	Y	N	N	Y

Missing values



Outliers

- Temperature
- Cold
- Medium
- Heat

- Excursion
- False
- True
- True

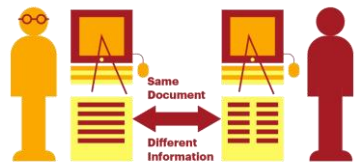


Population



Sample

## Inconsistent Data



# Achieving Excellence in Education with Advanced Strategies

## Innovative Assessment Techniques:

Employ advanced tools to assess student learning in real-time. Provide immediate feedback to help students adjust their learning strategies promptly.

## Scalable Educational Practices:

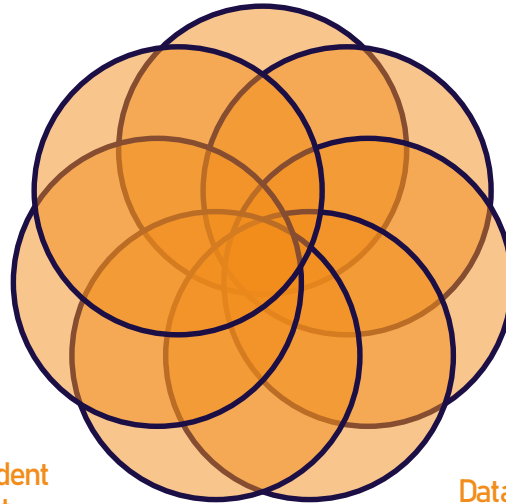
Apply strategies across various educational levels and settings, from K-12 to higher education. Scale solutions to accommodate different class sizes and learning environments efficiently.

## Increased Student Engagement:

Engage students with interactive and relevant learning experiences. Motivate learners through technology-enhanced methods that align with their interests and capabilities.

## Effective Grouping:

Create dynamically balanced teams based on data insights. Foster collaborative learning environments that leverage diverse student strengths.



## Tailored Learning Experiences:

Customize assignments and learning paths to individual performance levels or learning styles. Increase student engagement and improve academic outcomes through personalized education.

## Enhanced Learning Materials:

Deliver targeted educational content that addresses specific needs and preferences. Use AI to recommend resources that maximize learning efficiency and retention.

## Data-Driven Insights:

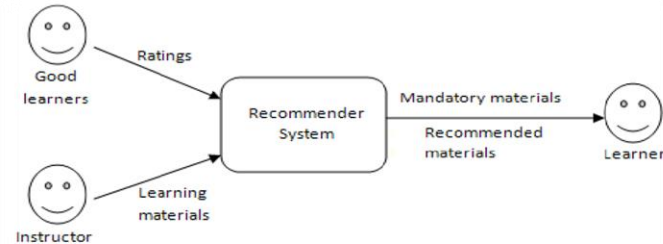
Utilize analytics to gain a deeper understanding of student behaviors and needs. Make informed decisions that enhance teaching strategies and student support.



# **Exploring Clustering: A Guide to Unsupervised Learning in Education with ChatGPT**

# Key Strategies for Personalized Education: Three Examples Using ChatGPT

<i>Example 1:</i> Team Formation Through Strategic Clustering	<i>Example 2:</i> Tailored Assignment Delivery	<i>Example 3:</i> Personalized Learning Material Recommendations
Create balanced teams by strategically mixing students	Deliver assignments that are suitably challenging for each student	Recommend specific learning materials for each student



# How to Perform Unsupervised Learning Tasks in ChatGPT

## 1. Prepare your Dataset

Ensure your dataset is correctly formatted

Include essential attributes for each student (e.g., Student ID, Learning Style, Performance Level, Test Score) in a simple table or list format

## 2. Paste the Dataset into ChatGPT

Start a new session in ChatGPT

Paste your dataset directly into the chat

## 3. Specify the Aim

Clearly state your analysis objective, such as

*"I want to use unsupervised learning to form balanced project teams based on learning style, performance level, and test scores."*



# How to Perform Unsupervised Learning Tasks in ChatGPT

## 4. Choose the Unsupervised Learning Method

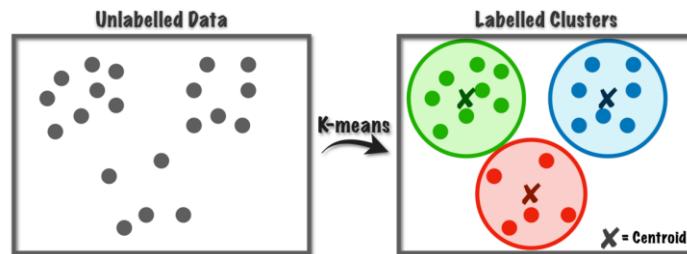
Specify the unsupervised learning technique to use, e.g.

*"Please apply k-means clustering to this dataset to identify suitable groups for team formation."*



If unsure, ask for recommendations, e.g.

*"Which unsupervised learning method would you recommend for forming teams from this dataset?"*





# How to Perform Unsupervised Learning Tasks in ChatGPT

## 5. Request Specific Analysis or Outputs



Request ChatGPT to simulate the analysis and provide specific outputs, e.g.  
*"Please simulate k-means clustering with 5 clusters and describe the characteristics of each cluster."*



Ask for suggestions on team formations based on clustering results, e.g.  
*"Suggest balanced teams based on the clustering results."*



# How to Perform Unsupervised Learning Tasks in ChatGPT

## 6. Discuss and Refine the Results

Review the suggestions provided by ChatGPT

Discuss the results to refine or adjust the groupings, e.g.

*"Can you redistribute the teams to increase diversity in learning styles?"* or  
*"How would the teams change if we used a different number of clusters?"*



## 7. Ask for Implementation Advice

Inquire about best practices for real-world implementation, e.g.

*"What are the best practices for implementing these team assignments in a classroom setting?"*



Seek advice on monitoring effectiveness, e.g.

*"How should I monitor the effectiveness of these teams?"*



# Example 1: Balanced Student Teams



## Goal

- To create teams that are formed by strategically mixing students from different clusters to create balanced teams that complement each other's skills and learning preferences



## Technique

- k-Means Unsupervised Algorithm



## Tool

- ChatGPT

# Step 1. Data Simulation

We will use a sample dataset for 30 students, each characterized by several attributes:

**Learning Style:** Coded as 0 for Visual and 1 for Auditory

**Performance Level:** Numerically encoded from 1 (Poor) to 4 (Excellent).

**Test Score:** A continuous variable scaled from 0 to 10.

Conceptual representation of the dataset →

Student	Learning Style	Performance Level	Test Score
1	0	2	5.4
2	1	3	7.5
3	0	1	3.2
4	1	4	8.9
5	0	2	5.0
...	...	...	...
30	0	1	4.5

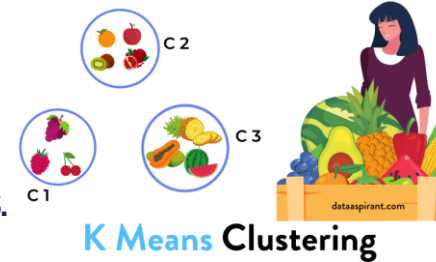
# Step 2: k-Means Clustering Explanation

## Key Concepts:

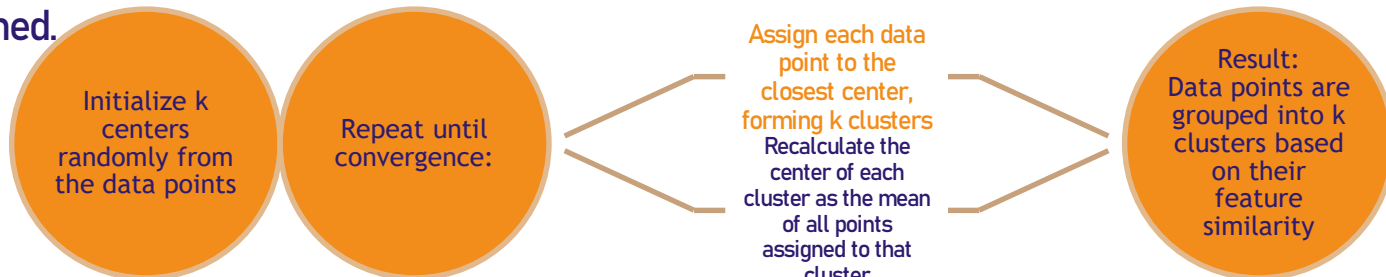
**Centers (Centroids):** The mean position of all the points in a cluster.

**Assignment:** Determining the nearest centroid for each data point.

**Update:** Continuously adjusting centroids to the center of their respective clusters.



The objective of k-means is to minimize the total variance within each cluster, aiming to form clusters where members are as similar to each other as possible, thereby making the clusters compact and well-defined.



# Step 3: Applying k-Means to the Dataset

We decide on  $k=5$  for clustering our dataset, using the attributes (Learning Style, Performance Level, Test Score) to define clusters. After clustering, we can identify which students belong to each cluster based on their similarities:

Cluster 0:  
Students 1, 7, 11,  
19, 23, 29

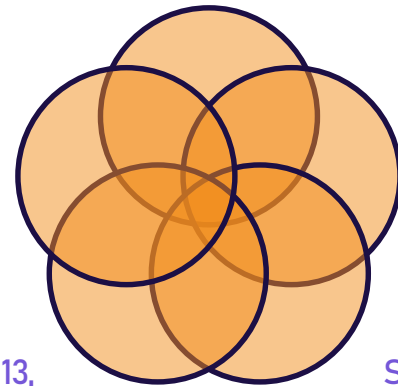
Hypothetical Cluster Assignments:

Cluster 4:  
Students 5, 15,  
27

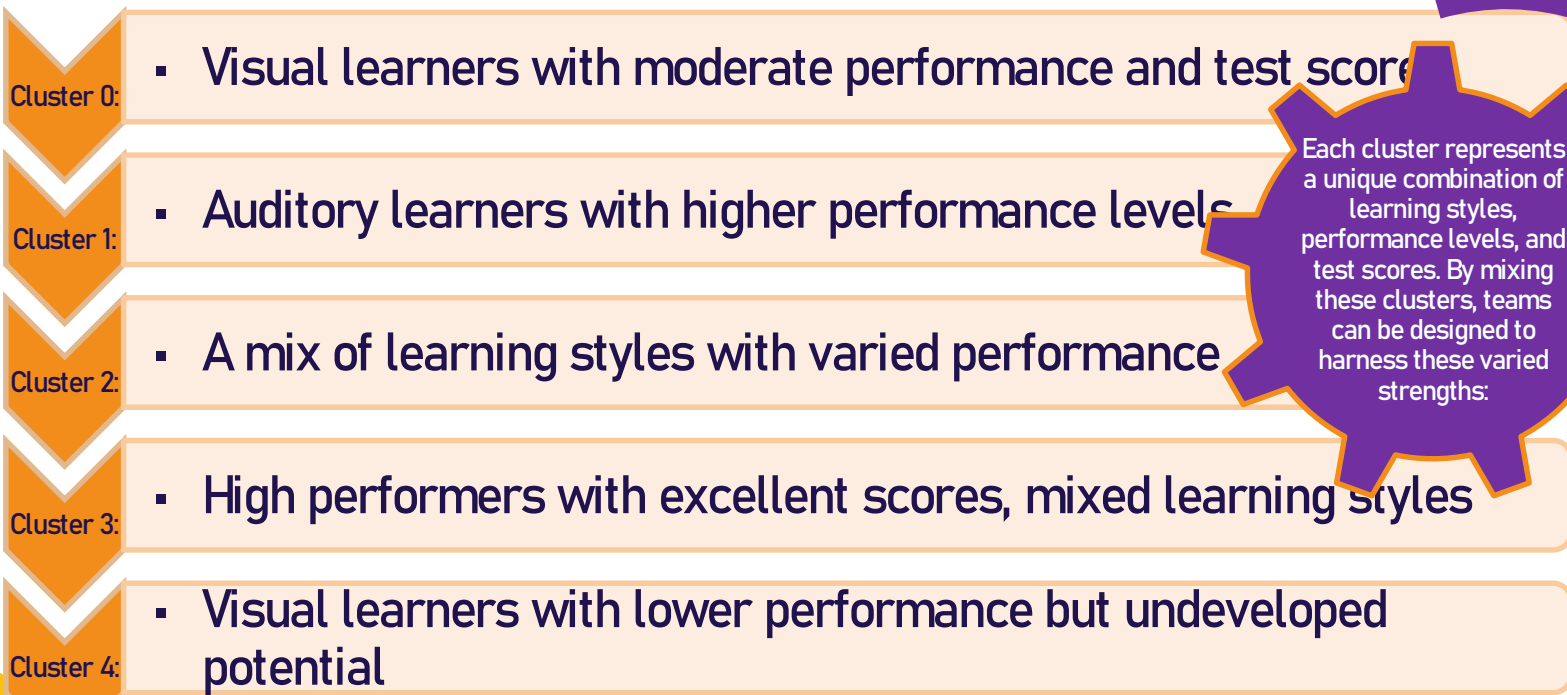
Cluster 1:  
Students 2, 6, 10,  
14, 18, 22, 26, 30

Cluster 3:  
Students 4, 9, 13,  
17, 21, 25

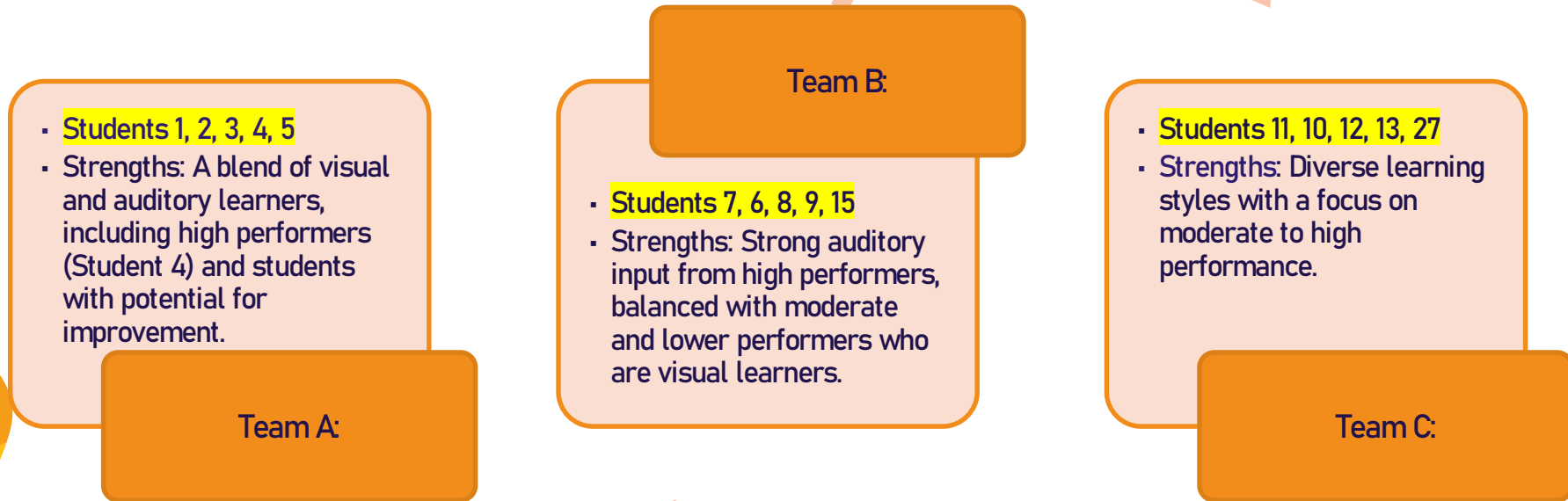
Cluster 2:  
Students 3, 8, 12,  
16, 20, 24, 28



# Step 4: Clusters for Team Strengths

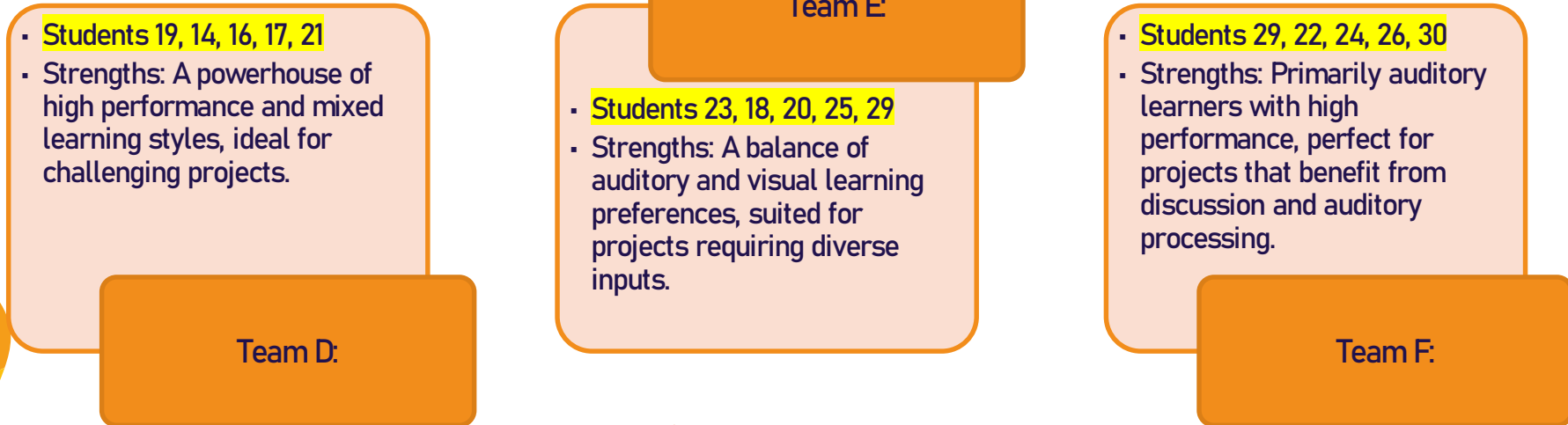


# Step 5: Forming the Teams





# Step 5: Forming the Teams



# Example 2: Tailored Assignments



## Goal

- To understand the capabilities of each student based on their learning style, performance level, and test scores to tailor assignments that are suitably challenging



## Technique

- k-Means Unsupervised Algorithm



## Tool

- ChatGPT

# Step 1. Data Recap

We will use the sample dataset for 30 students, each characterized by several attributes:

Student	Learning Style	Performance Level	Test Score
1	0	2	5.4
2	1	3	7.5
3	0	1	3.2
4	1	4	8.9
5	0	2	5.0
...	...	...	...
30	0	1	4.5

# Step 2: Clustering the Students

- Applying k-Means Clustering

Using k-means clustering, we will group students into clusters based on similarities in their learning styles, performance levels, and test scores. This will help identify groups of students who might benefit from similar types of assignment difficulties.

- k-Means Clustering Steps

**Standardize the Features:** Ensure each feature contributes equally.

**Determine the Number of Clusters:** Use an approach like the Elbow Method.

**Execute k-Means:** Cluster the students using the standardized features.

**Analyze Clusters:** Examine the characteristics of each cluster to understand the academic capabilities and needs of the grouped students.



# Step 3: Assignment Recommendation

- Understanding Cluster Characteristics

Once clusters are formed, analyze each cluster to determine the **common characteristics of the students** within. This might show, for instance, that:

**Cluster 1** consists of high performers with high test scores, predominantly auditory learners.

**Cluster 2** includes students with moderate performance levels and average test scores, mostly visual learners.

**Cluster 3** captures students who are performing poorly with low test scores.



# Step 3: Assignment Recommendation

- Tailoring Assignment Difficulties

Based on the cluster analysis, assign **difficulty levels to assignments** that match the capabilities of the students in each cluster:

**Cluster 1 (High Performers):** Recommend more challenging assignments to push these students' boundaries and enhance their learning potential.

**Cluster 2 (Moderate Performers):** Assign moderately challenging assignments that encourage learning and growth without overwhelming these students.

**Cluster 3 (Low Performers):** Recommend basic level assignments that focus on foundational skills and concepts to build confidence and improve understanding.



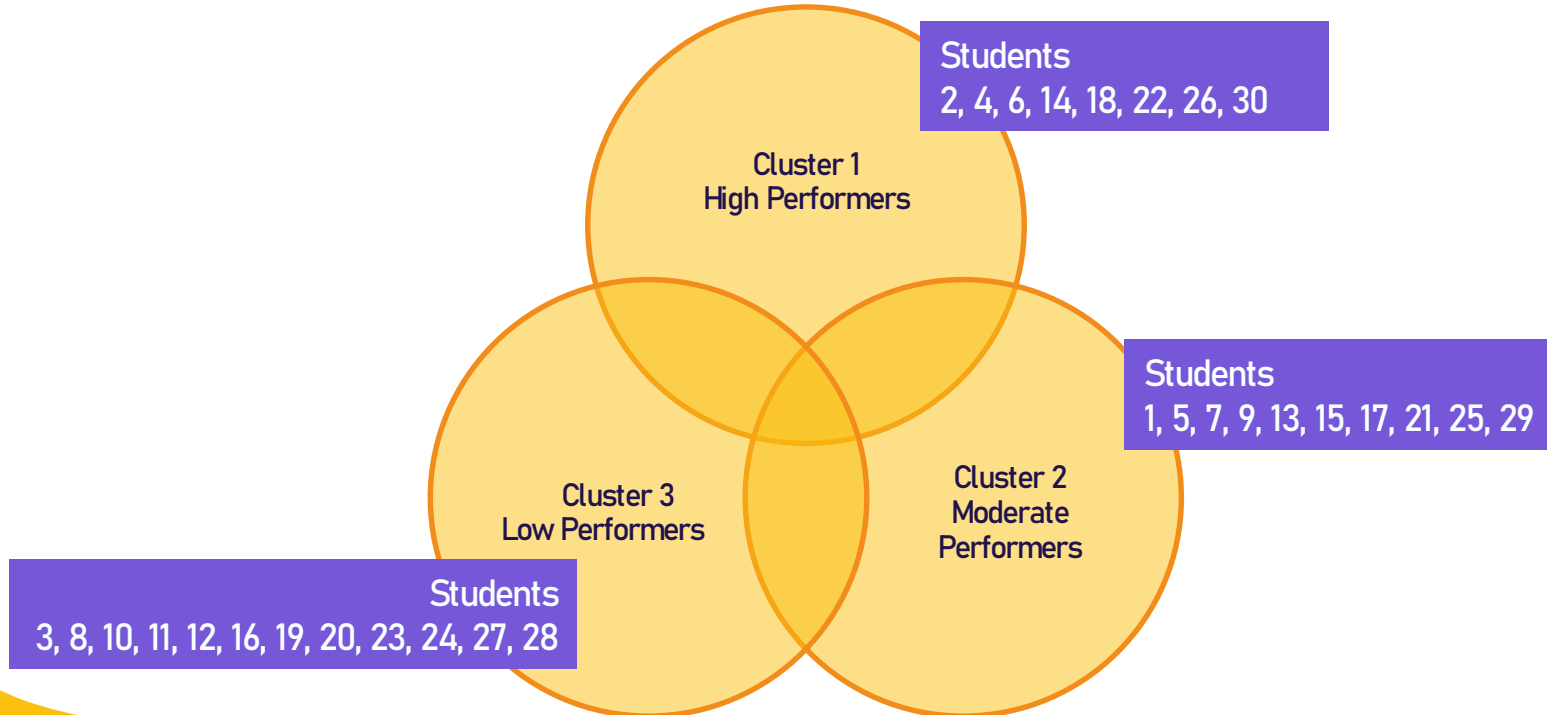
# Step 4: Implementation Strategy

- Pseudo-implementation

Given the clusters and their characteristics, the assignment difficulty can be programmatically assigned as follows:

```
for each student in dataset:  
  if student in Cluster 1:  
    recommend "High Difficulty"  
  elif student in Cluster 2:  
    recommend "Medium Difficulty"  
  else:  
    recommend "Low Difficulty"
```

# Step 5: Student Distribution





# Step 5: Team Formation

- **Students: 2 (C1), 1 (C2), 3 (C3)**
- Strengths: High-performing auditory learner, moderate visual learner, and a student who may benefit from additional support.

Team A:

Team B:

- **Students 4 (C1), 5 (C2), 8 (C3)**
- Strengths: High performer with strong auditory skills, a solid moderate performer with visual learning style, and extra support needs.

Team C:

- **Students 6 (C1), 7 (C2), 10 (C3)**
- Strengths: : Combination of top-tier auditory skills, stable visual learning, and foundational support.

# Step 5: Team Formation

- Students 14 (C1), 9 (C2), 11 (C3)
- Strengths: Strong auditory learning presence, mixed with moderate visual skills and support to improve.

Team D:

- Students 18 (C1), 13 (C2), 12 (C3)
- Strengths: High academic achievers with auditory preference, balanced by moderate visual and supportive needs.

Team E:

- Students 22 (C1), 15 (C2), 16 (C3)
- Strengths: Mix of top academic skills, average performance, and low scores needing boost.

Team F:



# Example 3: Material Recommendation



## Goal

- To recommend learning materials for each student



## Technique

- Hierarchical Clustering



## Tool

- ChatGPT

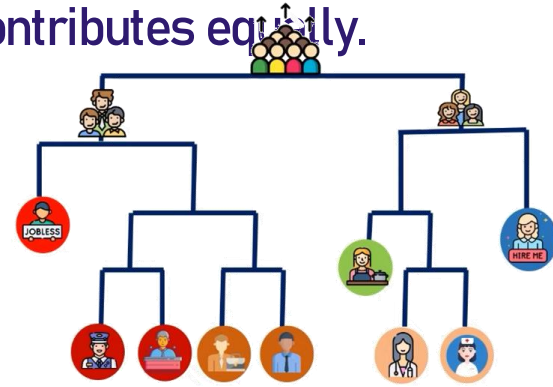
# Step 1. Data Recap

We will use the sample dataset for 30 students, each characterized by several attributes:

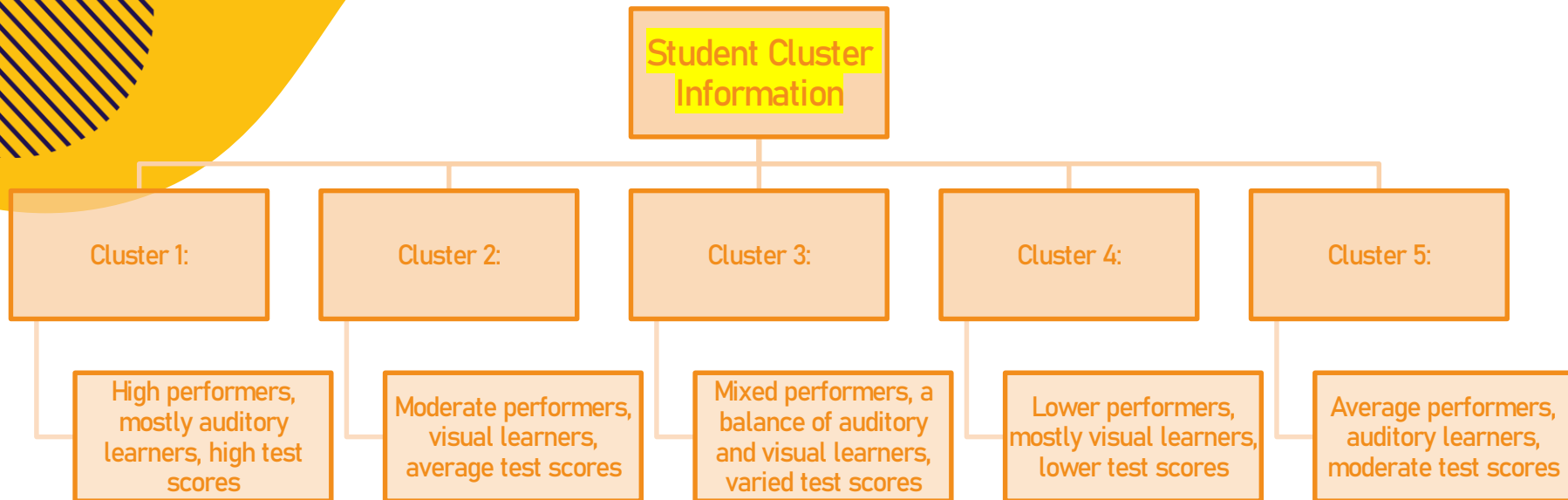
Student	Learning Style	Performance Level	Test Score
1	0	2	5.4
2	1	3	7.5
3	0	1	3.2
4	1	4	8.9
5	0	2	5.0
...	...	...	...
30	0	1	4.5

# Step 2: Clustering the Students

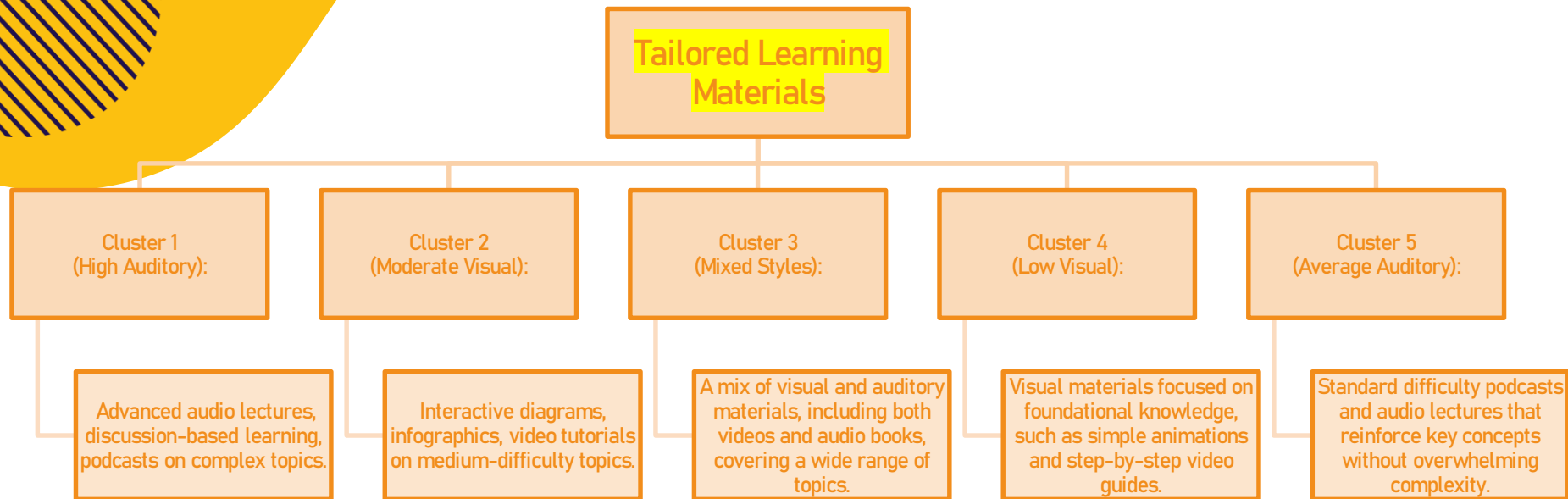
- Applying Hierarchical Clustering
- Standardize the Data: Scale the data so each feature contributes equally.
- Perform Clustering:  
Use an agglomerative clustering approach.  
Apply the Ward linkage method to minimize the variance within each cluster.
- Create a Dendrogram:  
Use visual tool to help determine where to cut the tree to define clusters.
- Decide on the Number of Clusters: Analyze the dendrogram to select an appropriate cutoff, creating a practical number of clusters, such as 5.



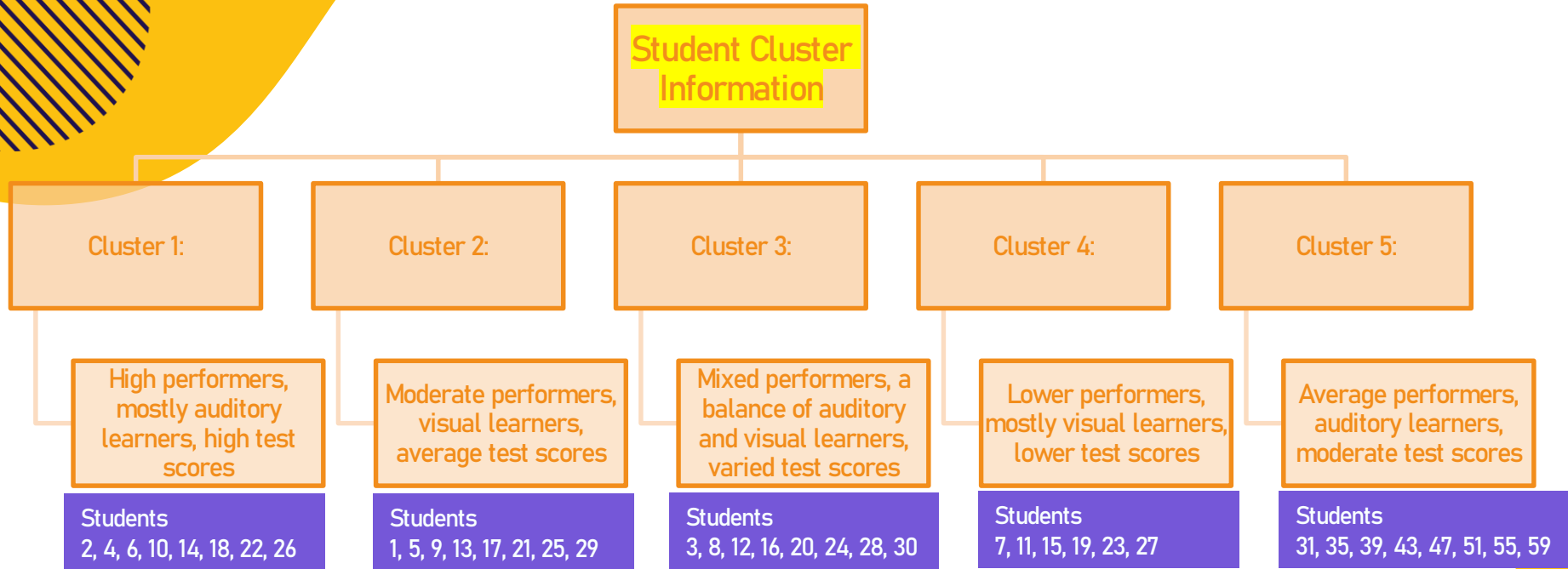
# Step 3: Recommendation Clusters



# Step 3: Recommendation Clusters



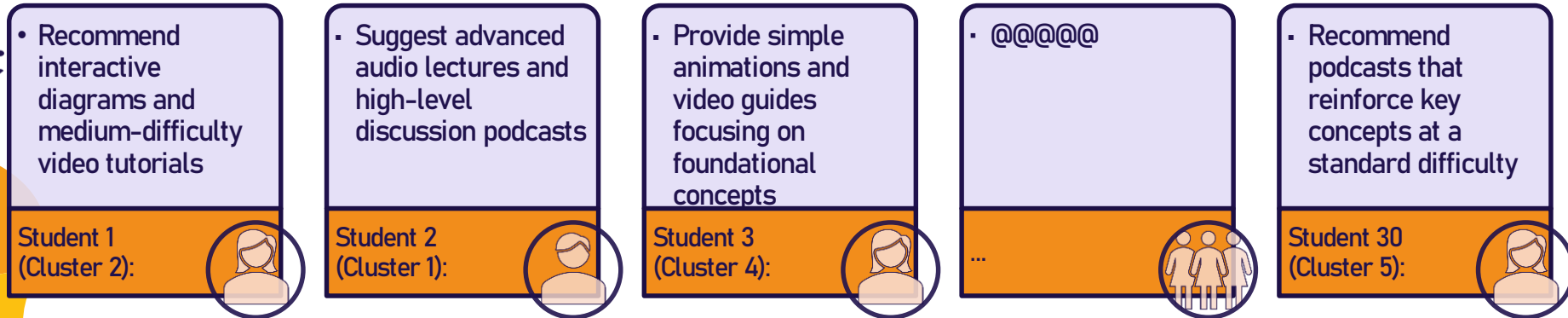
# Step 3: Recommendation Clusters





# Step 4: Personalized Recommendations

- Using the cluster assignments, we can now recommend specific learning materials to each student. Here's how this might look:



# Step 5: Implementation and Monitoring

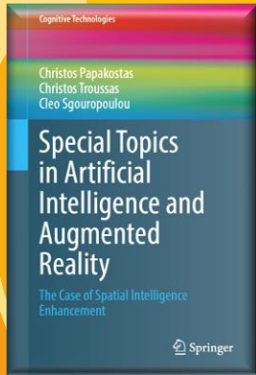
- Implementation
  - Assign the recommended materials to each student based on their cluster
  - Integrate these materials into the students' learning plans
- Monitoring
  - Track student engagement and progress with the assigned materials
  - Gather feedback on the effectiveness of the materials and make adjustments as necessary



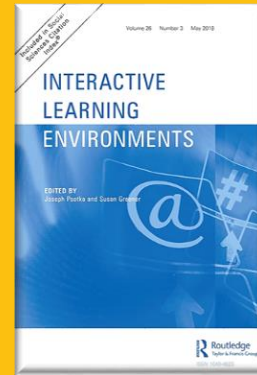
# Conclusions



# Further Reading: Selected Publications



Papakostas, C., Troussas, C., Sgouropoulou, C. (2023). *Special Topics in Artificial Intelligence and Augmented Reality: The Case of Spatial Intelligence Enhancement*, Book, Springer.



Troussas, C., Giannakas, F., Sgouropoulou, C., Voyiatzis, I. (2023). Collaborative activities recommendation based on students' collaborative learning styles using ANN and WSM. *Interactive Learning Environments*, 31(1), 54–67.



Krouska, A., Troussas, C. & Sgouropoulou, C. (2023). A novel group recommender system for domain-independent decision support customizing a grouping genetic algorithm. *User Model User-Adap Inter*, 33, 1113–1140.



Krouska, A., Troussas, C., Voulodimos, A., Sgouropoulou, C. (2022). A 2-tier fuzzy control system for grade adjustment based on students' social interactions, *Expert Systems with Applications*, 203, 117503.



**Thank you for your kind  
attention!!**

**Prof. Cleo Sgouropoulou**  
**[csgouro@uniwa.gr](mailto:csgouro@uniwa.gr)**

