Data Mining: Concepts and Techniques

DR. ALI ALSALEHY

Instructor & Course Welcome

- CS 466: Data Mining – Concepts and Techniques
- Instructor: Dr. Ali Suliman AlSalehy
- Office: M035
- Email: Salehyal@rcjy.edu.sa
- This is an online course
- Please read and follow the logistics carefully
- Clear rules = smooth semester for everyone

Office Hours (Online Only)

To ensure equal access for all students, **all office hours are online**

GROUP OFFICE HOURS

- Held every Wednesday
- Open to the entire section
- Time and link announced in advance
- Best for common questions and general clarification

INDIVIDUAL OFFICE HOURS

- One-on-one meetings
- By appointment only
- Email me to schedule a suitable time

Email Rules (Very Important)

To help me respond efficiently:

- Always include [CS466] in the email subject
- Example:

  [CS466] Question about Assignment 1
- Emails without the tag may take longer to receive a response
- This is not optional - it helps me manage course emails fairly

Attendance Policy

Attendance will be counted and recorded

- Attendance is a learning tool, not just a requirement
- You are responsible for attending regularly
- Since this is an online course:
  - Stable internet is your responsibility
  - Attend from a quiet and suitable place
- Attendance records are updated weekly
  - Check them regularly

In-Class Assignments

- There will be short in-class assignments during lectures
- Based on the previous lecture
- This means you must study before class
- **Important details:**
  - Assignments are random
  - May appear at the beginning, middle, or end of the lecture
  - You must be present to attempt them

If you miss one or two:

- Do not panic
- Focus on the remaining ones
- They are designed to be reasonable if you stay engaged

How to Use Office Hours Effectively

Office hours are for support and clarification

You may use them to:

- Ask about lectures or assignments
- Clarify concepts you already studied
- Discuss different solution approaches

Before attending, please:

- Attend the lecture
- Review the slides
- Try on your own first
- Prepare specific questions

Instead of:

- “I don’t understand Chapter 2”

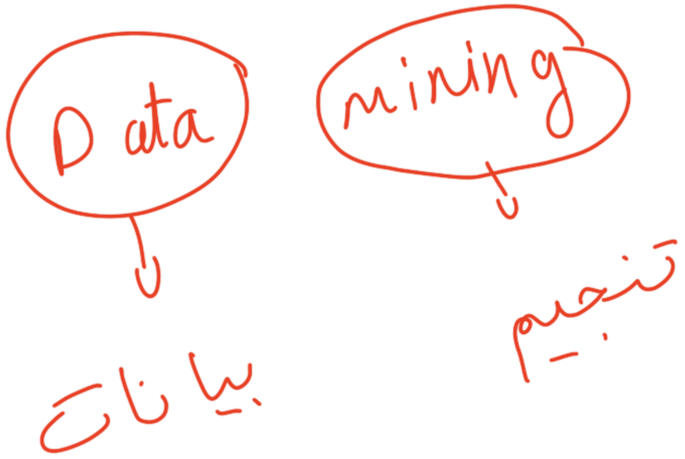
Try:

- “I’m confused about the difference between X and Y”

Additional Notes & Expectations

- Office hours are not for re-teaching full lectures
- Bring examples or problems you worked on
- Respect time limits
- Email early if you need extended discussion
- If many students ask the same question, it may be addressed in group office hours

The more prepared and engaged you are, the more you will benefit from this course.


Data
mining
بيانات
تنجيم

Introduction

- Data explosion in business, science, social media, sensors.
- Raw data alone has limited value.
- Challenge: Too much data, not enough insight.
- Data Mining = turning data into useful knowledge.
- Key applications:
  - Business decision support
  - Fraud detection
  - Healthcare & bioinformatics
  - Scientific discovery

Warm-up Question

Why do you think companies like Amazon, Netflix, or hospitals care about data?

شركات يقصد البيانات

Why Data Mining?

- Supports better decision-making.
- Finds hidden patterns not visible in raw data.
- Two main goals:
  - **Prediction** – forecast future outcomes.
  - **Description** – summarize key characteristics.

Another Question

If we already have databases and statistics, why do we need “data mining”?

Netflix

what happened? ← Data base → يخزن الأفلام

what is happening? ← statistic → متوسط المشاهدات

what will happen? ← Data mining → يحلل الفيلم المناسب

why?

What is Data Mining?

- Process of extracting **patterns and knowledge** from large datasets
- Part of the **KDD process (Knowledge Discovery in Databases)**.
- Uses tools from statistics, machine learning, AI, and databases.

V. Important

The KDD Process

تنظيف البيانات ازاله التشويش

1. Data cleaning → remove noise.
2. Data integration → combine sources. → دمج البيانات في اكثر من مصدر
3. Data selection → choose relevant data. اختيار البيانات
4. Data transformation → format properly. تحويل شكل مناسب
5. Data mining → discover patterns. اكتشاف الانماط
6. Pattern evaluation → test for usefulness. تقييم الانماط
7. Knowledge presentation → visualization & reporting.

كيفي المعرفه رسومات تقارير

Another Question

What kinds of data do you think we can mine?

What Kinds of Data can be Mined?

- Structured: databases, tables.
- Semi-structured: XML, JSON.
- Unstructured: text, images, video.
- Spatiotemporal: weather, traffic.
- Stream data: sensors, financial markets.
- Web/social media: clickstreams, networks.

Another Question

What kinds of patterns would be most useful to discover in data?

What Kinds of Patterns can be Mined?

- Association rules – find correlations (“milk ⟶ bread”).
- Classification – assign labels (spam vs. not spam).
- Clustering – group by similarity.
- Sequential patterns – discover order/sequence.
- Outliers – detect anomalies (fraud).
- Trends & evolution – observe changes over time.

Summary

- Data mining = extracting useful patterns from data.
- Important for prediction and description.
- Can handle many types of data.
- Finds patterns: association, classification, clustering, anomalies, trends.
- Foundation for big data analytics, AI, and decision-making.