CSCI 347: Data Mining

Course orientation

Instructor details

- Adiesha Liyanage

- Email: a.liyanaralalage@montana.edu

- Office hours:
  - Monday 12:00 PM- 02:00 PM
  - Tuesday 10:00AM - 11:00 AM

- Office: Barnard 349 (right next to conference room)

TA details

- Name: TBD
- Email:
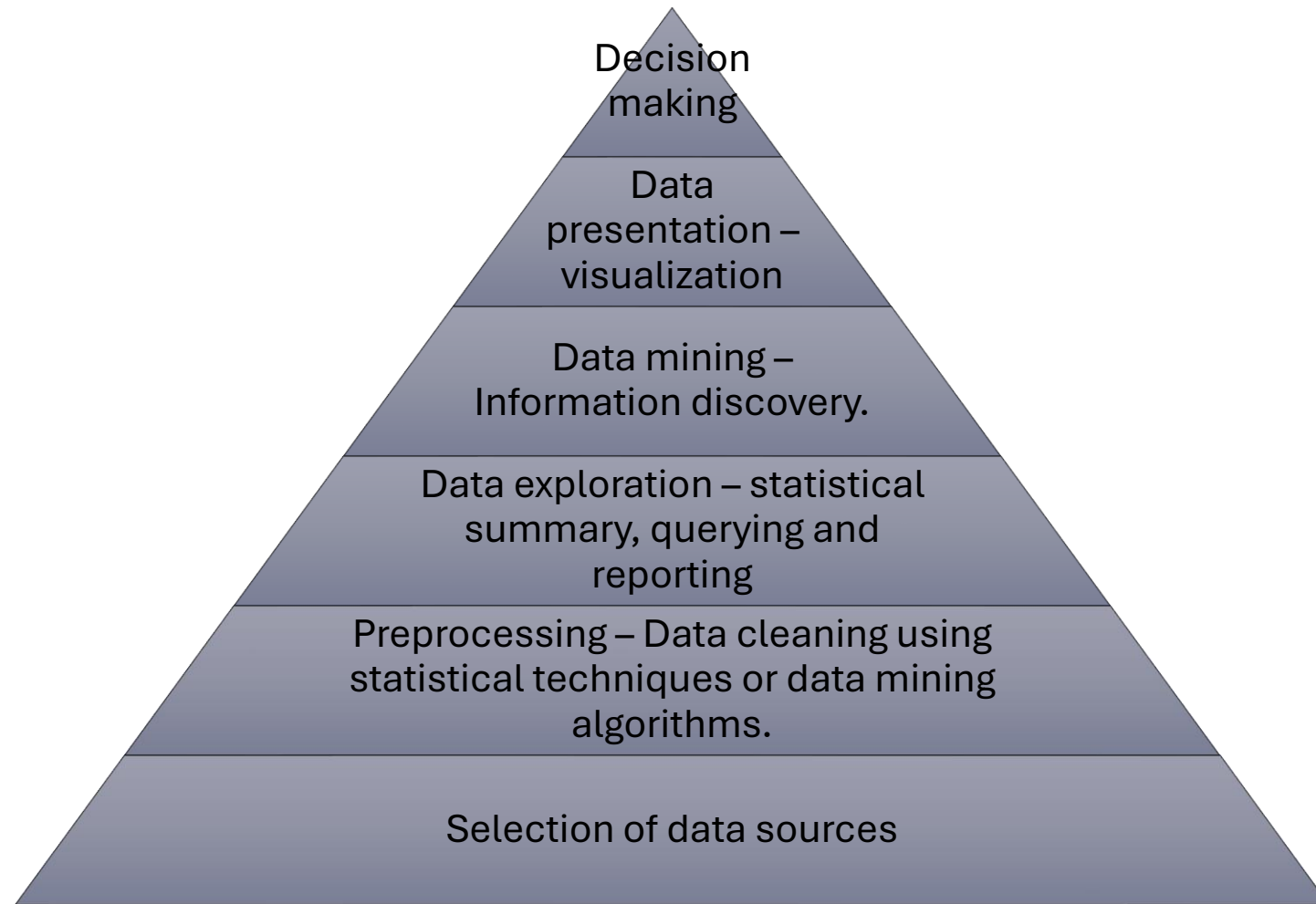- Office hours:
  - TBD

Activity

- Let's go around the class—please tell us your name, where you're from (City/State), and your major and something fun you did during the winter break.
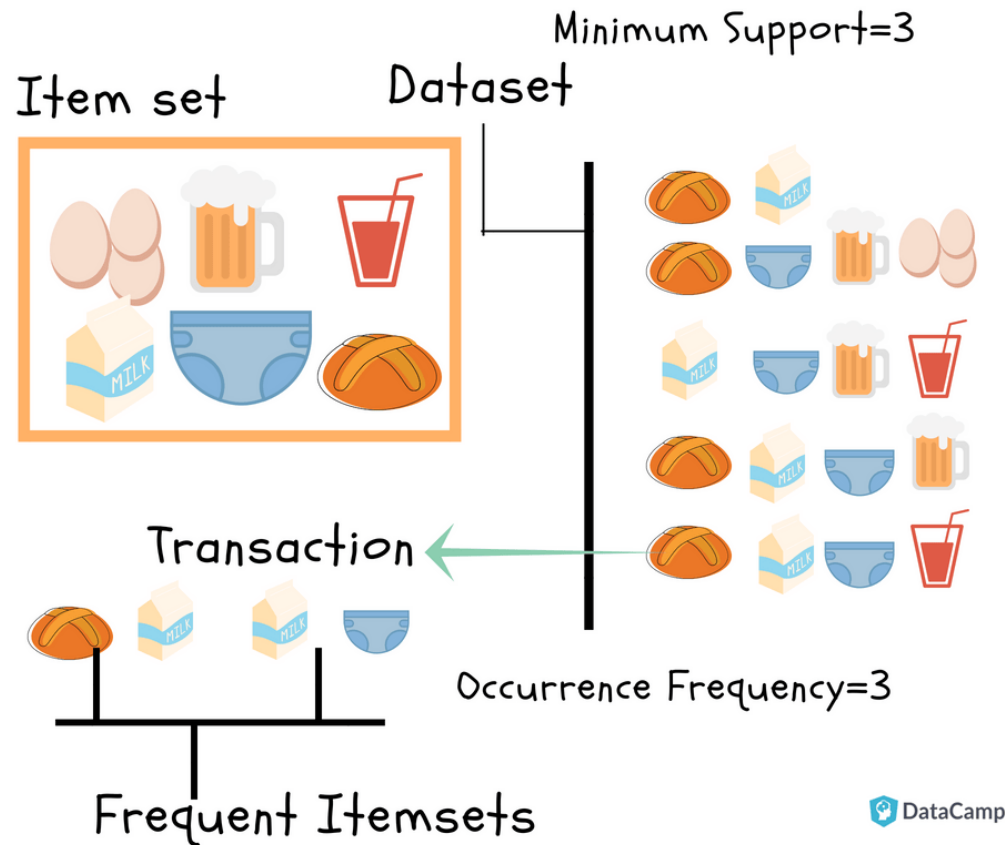
What is Data Mining?

- "the process of finding valid, novel, useful, and potentially understandable patterns in data" — G. Piatetsky-Shapiro, KDnuggets

- "algorithms for learning, analysis, data management and visualization of large datasets" — C. Faloutsos, CMU

- "Discovery of useful, possibly unexpected, patterns in data" — J. Ullman, Stanford

- "providing tools to discover knowledge from data" - J. Han in his textbook, Data Mining: Concepts and Techniques

- "the process of discovering insightful, interesting, and novel patterns, as well as descriptive, understandable, and predictive models from large-scale data" — M. J. Zaki and W. Meira Jr. in Data Mining and Analysis: Fundamental Concepts and Algorithms

- Like how mining gold is digging through earth and rock for the valuable bits, Data mining is sorting though large data sets to find valuable information.

- Marketing

- Risk Management

- Fraud detection

- Cybersecurity

- Medical diagnosis

Process of Data Mining

Decision making

Data presentation – visualization

Data mining – Information discovery.

Data exploration – statistical summary, querying and reporting

Preprocessing – Data cleaning using statistical techniques or data mining algorithms.

Selection of data sources

Data Mining Application: Market basket Analysis

https://www.datacamp.com/tutorial/market-basket-analysis-r

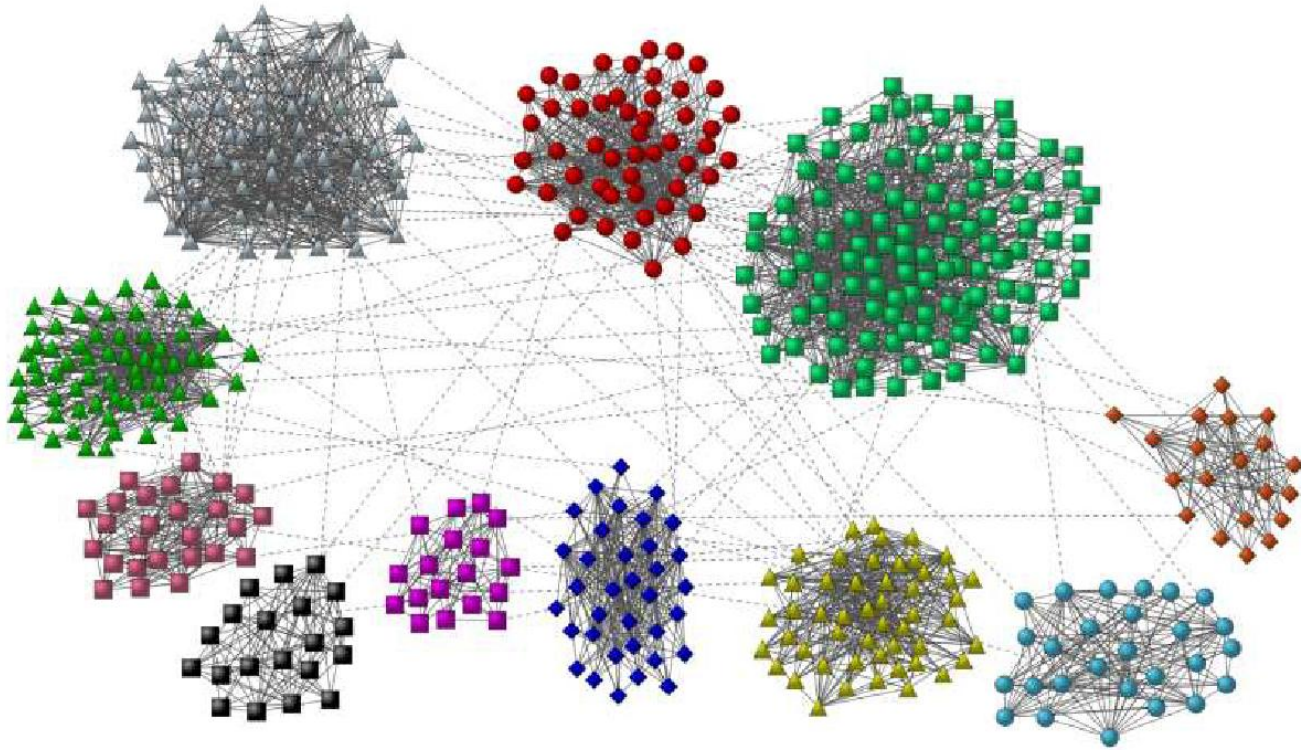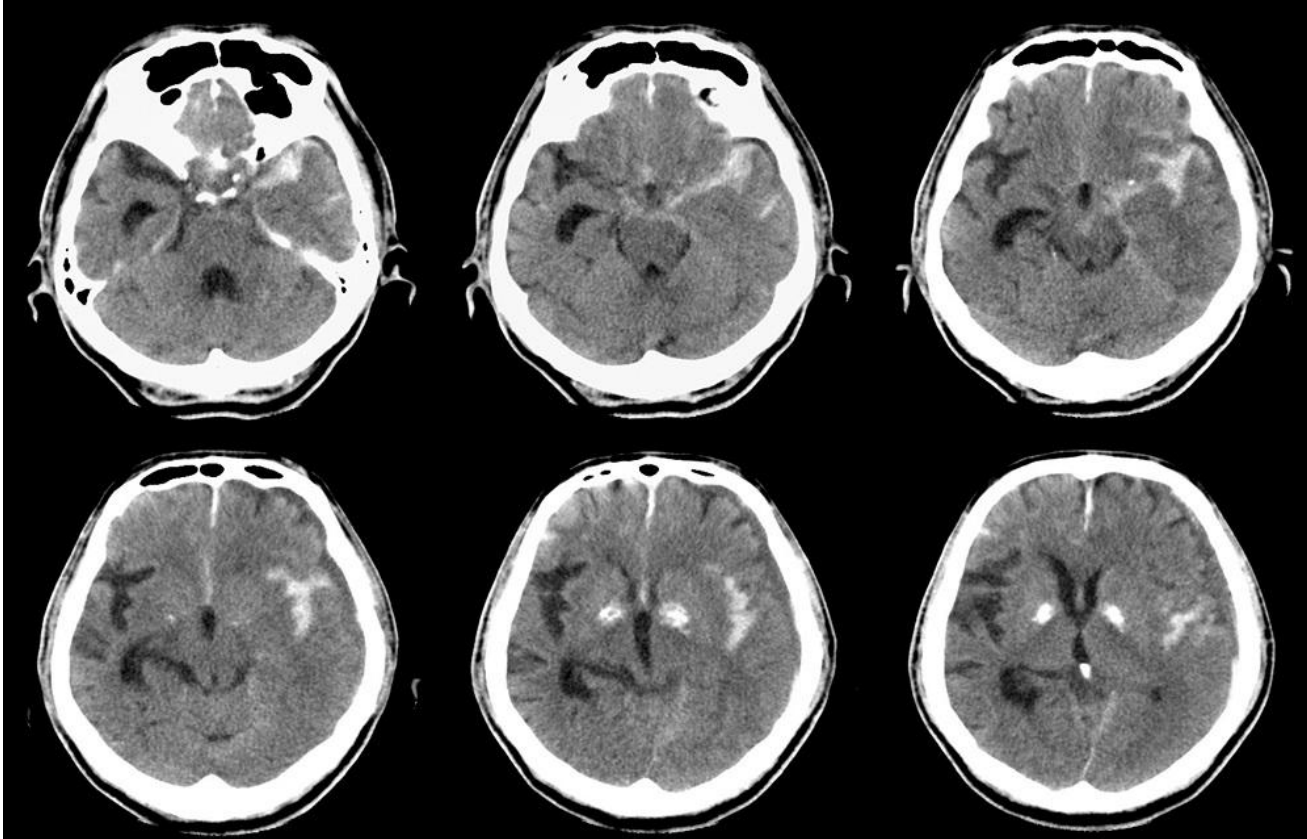# Data Mining Application: Finding communities in social network

Image taken from: Community detection in graphs, Fortunato, Santo

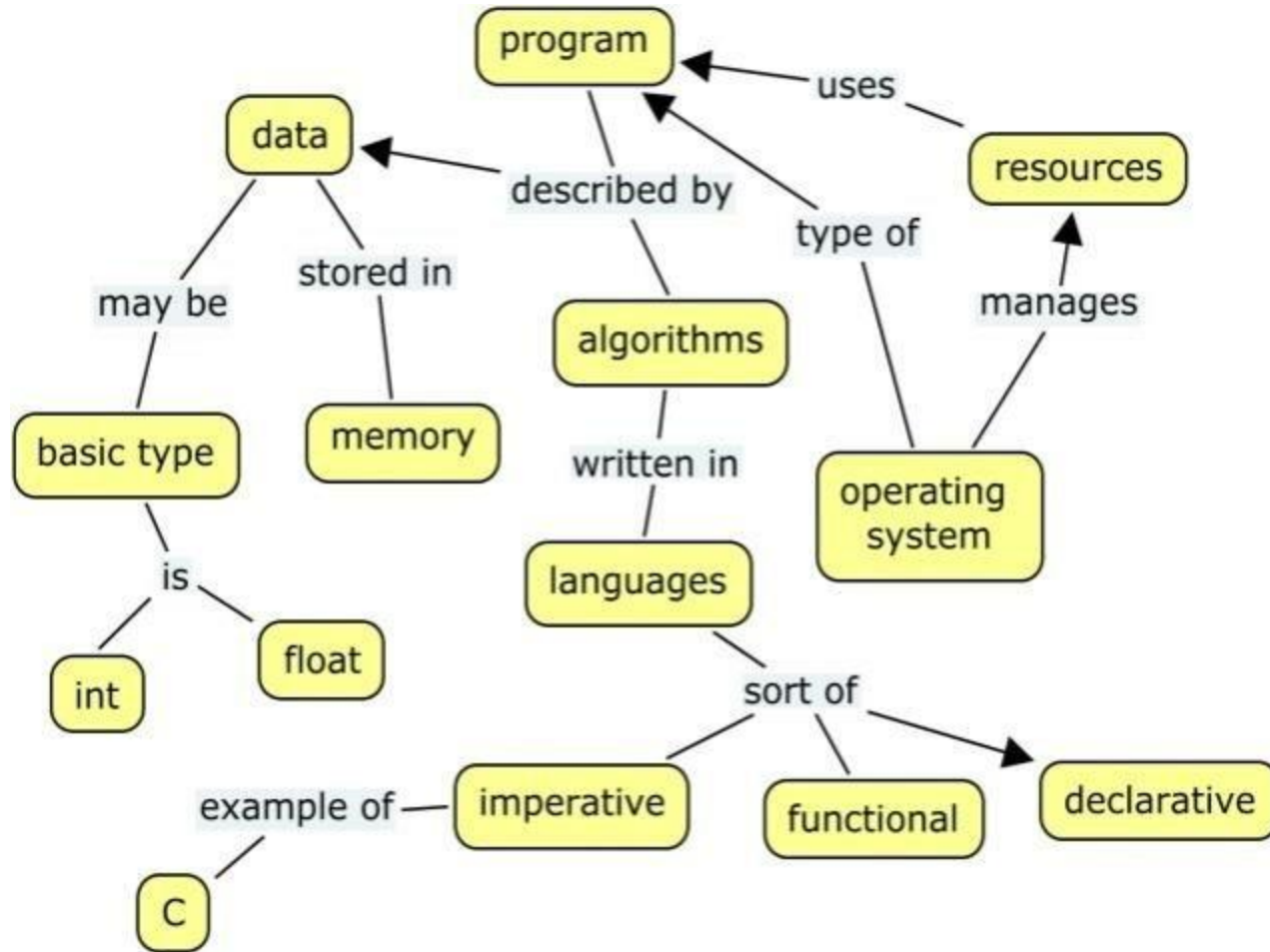# Data Mining Application: Anomalies in healthcare data

Deep learning algorithms can identify abnormalities on head computed tomography (CT) scans in patients who present
with head trauma or stroke symptoms, according to study results published in the Lancet*.

https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(18)31645-3/fulltext

Concept Maps

- "Concept map is a graphical tool for organizing and structuring knowledge by depicting concepts as nodes, and relationships between concepts as edges" – Wei et al. in Concept mapping in computer science education.
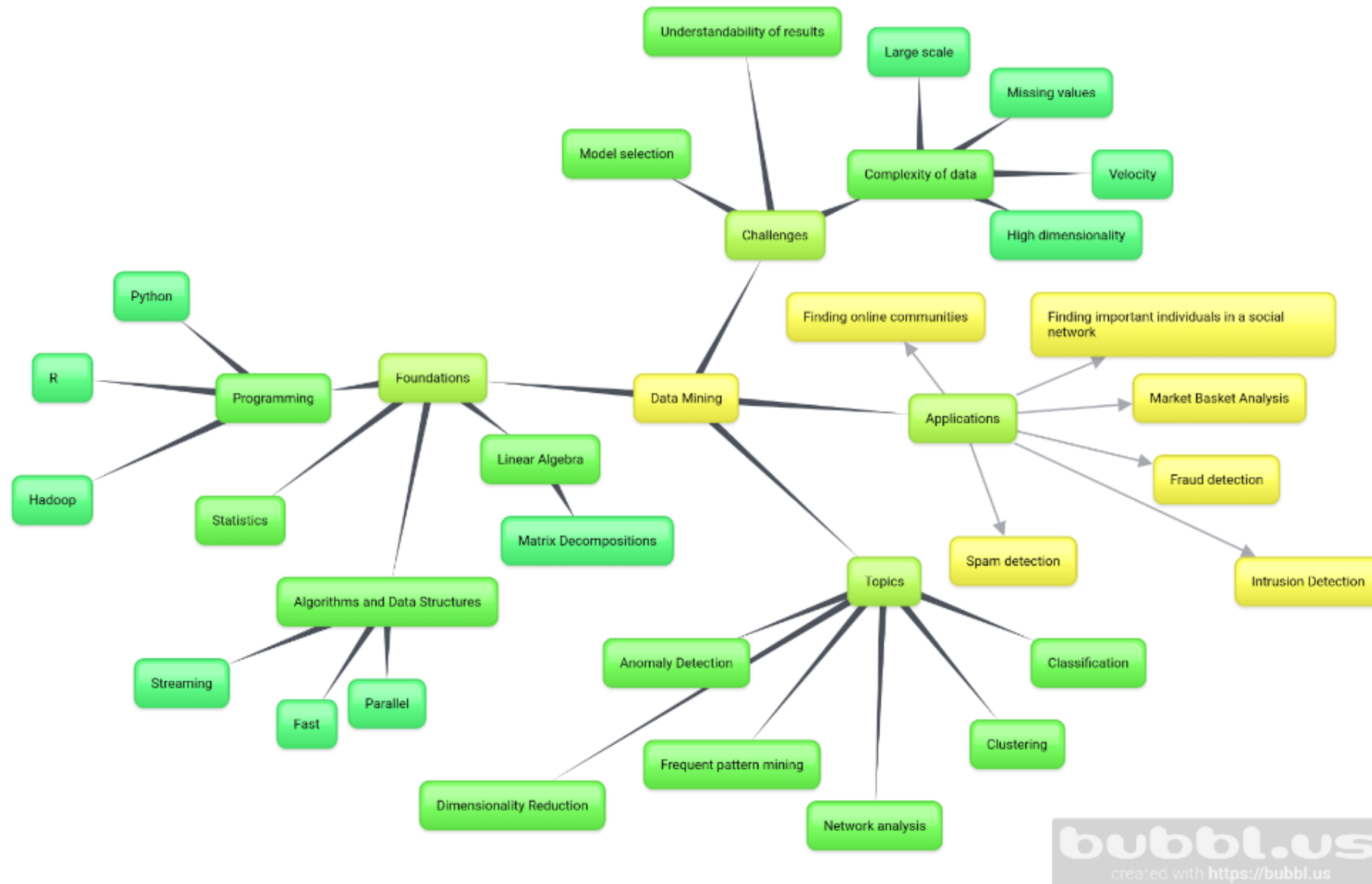
Example concept maps

CM for operating systems - An Evaluation Methodology for Concept Maps Mined from Lecture Notes: An Educational Perspective

Activity: Let's create a data mining concept map

- CM helps to visualize important concepts related to a theme.
  - Work with people around you.
  - Think about following questions:
    - What are applications in data mining ?
    - What are the foundational skills required to be an expert in this area ?
      - Programming languages
      - Mathematical skills etc.
    - What are the topics related to Data mining that comes to your mind?
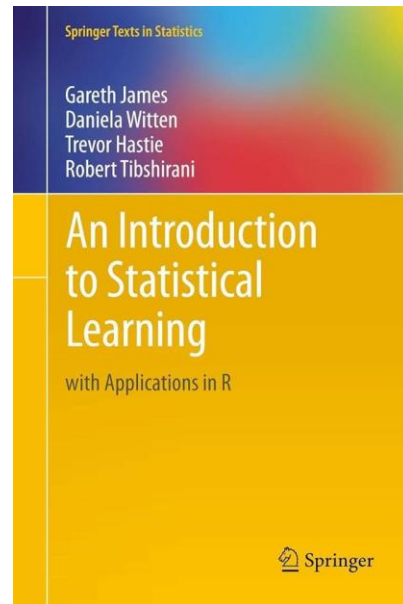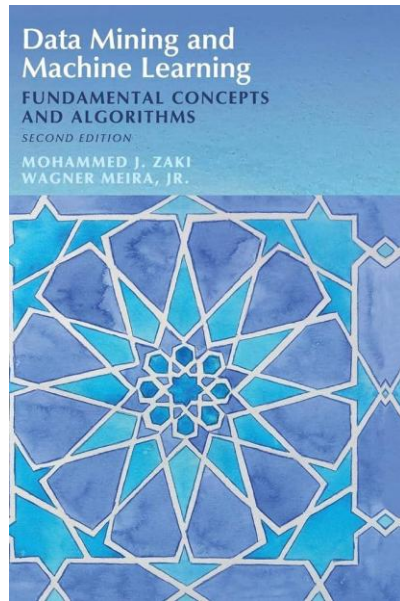   - What are challenges in Data mining ?

Summary

- Topics: Common data formats, exploratory data analysis, data preparation, graph analysis, dimensionality reduction, clustering, classification and few advanced topics.

Textbooks

- Data Mining and Analysis: Fundamental Concepts and Algorithms by Mohammed J. Zaki and Wagner Meira, Jr.

- Optional: An introduction to Statistical Learning (ISLR)

Programming Language and Software

- We will use Python as the programming language

- We use Jupyter Lab as the IDE for the course, please make sure that you have installed this before the next session.

- https://www.datacamp.com/tutorial/installing-jupyter-notebook

- Have your laptop with you for in-class activities.

Communication

- Post your questions on MSU CSCI Discord channel.
  - CSCI-347-DATA-MINING (CATEGORY)
- I will post the material on Canvas.
- I will use GitHub page for updating the course schedule.
  - https://github.com/adiesha/CSCI347Spring2026

Pre-requisites

- Students are expected to have taken statistics/probability course and have basic working experience in programming.

Evaluation and grading

Grading	Weight of each category (from 100)
Quizzes	10
Attendance	6
Assignments	30
Mini projects (3 projects)	24
Mid-term exam	10
Final Project	20
Bonus (depending on the situation and this is entirely up to me)	Up to 2
Total	100 + (up to 2)

Quizzes

- Canvas weekly or per topic quizzes (you can expect lot of quizzes)
- But these quizzes will be relatively short.
- No makeup quizzes, I will drop lowest quiz from your overall grade.

Attendance

- 6% of your grades would be attendance.
- If you have 80% of the attendance, then you will get full grade for your attendance score.
- We will use iClicker to record attendance.
- This is a free service.
- Use the following link to register for the course.
- https://join.iclicker.com/ZABU
- Register one device for iClicker attendance.
- At start of each class, you must use your device to record attendance.
- If you encounter problems with registration talk with UIT.

Attendance

- Please do not come to me at the end of the semester and tell me that you could not get the iClicker working.

- Resolve these issues in the beginning of the semester.

- The attendance server is run automatically during the class time, this would include holidays, I will remove these days manually from the server so do not worry about them.

- During the tests you do not have to record attendance.

Assignments

- There will be written assignments about every other week or after completing each chapter.

- The assignments should be turned in on Canvas.

- For coding parts of the assignments, you can submit the JupyterLab files along with other written materials.

Mini Projects

- There will be 3 mini projects.

- They should be submitted as a well-commented report in PDF format along with your code.

- The mini-projects should be completed in teams of 3-4 members.
  - This is a must. Since we have lot of students, individual project grading is too much for the TA for timely grades.
  - If you create groups of 1 or 2 people, then there will be a penalty.
  - Only one submission is required. (Multiple submissions by different people will result penalties)

Mid term test

- There will be one in-class mid-term exam in this course.

- I will announce the exact exam dates later.

- If you need accommodations, please let me know before the exam.

- Roughly, speaking exam would be in April. (I will announce the dates later)

Final project

- Students should propose a data mining problem and dataset and offer a solution using the methods we learn in class.

- Groups of 3-4 students.

- More information will be shared during the semester.

Special needs and accommodations

- If you have a documented disability and need or might need any accommodation, please reach out to the Office of Disability Services as soon as possible.

- They will instruct you on what you need to do to get accommodation for this class.

Course survey and evaluation

- If 70% of the students complete the course survey and evaluation, all the students will get some bonus points. (I will decide the exact value of the bonus points; you can expect it to be between 0.25— 0.75).

Classroom etiquette

- Please note that, disruptive behavior during the class time will not be tolerated and will result in such disruptive students being asked to leave the lectures.

- Please be kind and respectful to other students in the class.

- I will add more requirements here if it deems appropriate.

Course schedule

- https://github.com/adiesha/CSCI347Spring2026

Assignment 0 – About You

- Please update your canvas photo.

- Please create a single slide with the following information.
  - Your name (with pronunciation hints).
  - Hometown/Country (optional).
  - Your major.
  - One thing you would like to learn from this course.
  - One interesting fact about you.

- Convert it to a pdf and submit it.