



CS280 Elements of Data Processing (Online)

| | | | |
|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|---------|
| Instructor Information | Professor An Email: dranteaching@hotmail.com | | |
| Term | June 27, 2022 - July 22, 2022 | Credits | 4 units |
| Course Delivery | The class will be delivered in the format of online. Other than recorded lecture videos, the instructor will arrange 2 hours' real-time interactions with students per week (via discussion forum, zoom meeting, and WeChat). The workload students are expected to complete to properly pass this course is about 15 hours per week. | | |
| Required Texts (with ISBN) | Recommended texts: J. Han, M. Kamber and J. Pei, Data Mining: Concepts and Techniques, 3 rd ed., Morgan Kaufmann, 2012. ISBN: 978-0-12-381479-1. Bing Liu, Web Data Mining, Springer, 2011. ISBN: 978-3-642-26891-5. | | |
| Prerequisite | Students are expected to have completed one of computer programming courses such as Python, C++, Java, C#, etc. or have good knowledge of one of such programming languages. | | |



Course Overview

This course covers both theoretical foundations and practical techniques and tools for data processing. Topics include data representation, cleaning, transformation and analysis, visualization, privacy, clustering and classification methods, information retrieval, data and web mining, model evaluation.

Learning Outcomes

The students will be able to:

1. Have a fundamental understanding on data, data representation and storage, processing, visualization, and management.
2. Identify and use current data processing techniques, skills, and tools to perform effective data processing and analysis.
3. Have a basic knowledge of information retrieval, data mining, recommender systems, and model evaluation.

Program Outcomes

This course addresses the following program outcomes:

- An ability to apply knowledge of computing and mathematics appropriate to the discipline
- An ability to analyze a problem, and identify and define the computing requirements appropriate to its solution
- An ability to design, implement, and evaluate a computer-based system, process, component, or program to meet desired needs
- An ability to use current techniques, skills, and tools necessary for computing practice
- The capability for critical and independent thinking and skills for lifelong learning
- Respect for academic integrity and the ethics of scholarship



Grading Policy

| | |
|--------------|-----|
| Quizzes | 10% |
| Presentation | 10% |
| Assignments | 30% |
| Midterm | 20% |
| Final Exam | 30% |

Grading Scale is as follows

| Number grade | Letter grade | GPA |
|--------------|--------------|-----|
| 90-100 | A | 4 |
| 85-89 | A- | 3.7 |
| 80-84 | B+ | 3.3 |
| 75-79 | B | 3 |
| 70-74 | B- | 2.7 |
| 67-69 | C+ | 2.3 |
| 65-66 | C | 2 |
| 62-64 | C- | 1.7 |
| 60-61 | D | 1 |
| ≤59 | F (Failure) | 0 |



Class Schedule

| Date | Lecture | Readings | Online Teaching Arrangement |
|--------|-----------------------------------------------------------------------------------------------------|-------------------------------|--------------------------------------------------------------------------------------------------|
| Day 1 | Why Processing Data, Data Representation, Type of Attributes, Basic Statistical Description of Data | HKP: 3.1, 2.1-2.2 | approximately 50 minutes pre-recorded video lectures plus 50 minutes online interaction via Zoom |
| Day 2 | Data Integration and Cleaning: Missing Values and Outlier Detection and Removal | HKP: 3.2, 12.1-12.2 | approximately 50 minutes pre-recorded video lectures plus 50 minutes online interaction via Zoom |
| Day 3 | Transformation by Normalization, Discretization by Binning | HKP: 3.5.1-3.5.3 | approximately 50 minutes pre-recorded video lectures plus 50 minutes online interaction via Zoom |
| Day 4 | Data Dimension Reduction | HKP: 3.4 | approximately 50 minutes pre-recorded video lectures plus 50 minutes online interaction via Zoom |
| Day 5 | Text Preprocessing and Information Retrieval Query languages and processing | L: 6.1-6.3, 6.5-6.6 | approximately 50 minutes pre-recorded video lectures plus 50 minutes online interaction via Zoom |
| Day 6 | Entropy and Information Gain | HKP: 8.2.2 | approximately 50 minutes pre-recorded video lectures plus 50 minutes online interaction via Zoom |
| Day 7 | Association Rules | L: 2.1-2.2 | approximately 50 minutes pre-recorded video lectures plus 50 minutes online interaction via Zoom |
| Day 8 | Data Visualization, Clustering and Clustering Visualization | HKP: 2.3, 10.1-10.2 L: 4.2 | approximately 50 minutes pre-recorded video lectures plus 50 minutes online interaction via Zoom |
| Day 9 | Presentation | | |
| Day 10 | Midterm | | |
| Day 11 | Classification Methods: Decision Trees, K-Nearest Neighbor | HKP: 8.2, 9.5.1 L: 3.9 | approximately 50 minutes pre-recorded video lectures plus 50 minutes online interaction via Zoom |
| Day 12 | Classification Methods: Naïve Bayes, Combining Classifiers | HKP: 8.3 | approximately 50 minutes pre-recorded video lectures plus 50 minutes online interaction via Zoom |



| | | | |
|--------|-------------------------------------------------------------------------------------------------------------------------------|----------------------------|-----------------------------------------------------------------------------------------------------------|
| Day 13 | Experimental Design and Evaluations | HKP: 8.5.1-8.5.5 L: 6.4 | approximately 50 minutes pre-recorded video lectures plus 50 minutes online interaction via Zoom |
| Day 14 | Link Analysis and Social Network Analysis | L: 7.1 | approximately 50 minutes pre-recorded video lectures plus 50 minutes online interaction via Zoom |
| Day 15 | Link Analysis and Social Network Analysis | L: 7.1 | approximately 50 minutes pre-recorded video lectures plus 50 minutes online interaction via Zoom |
| Day 16 | PageRank | L: 7.3 | approximately 50 minutes pre-recorded video lectures plus 50 minutes online interaction via Zoom |
| Day 17 | Assessing Correlations and Recommender Systems | HKP: 2.4.7 L:12.4 | approximately 50 minutes pre-recorded video lectures plus 50 minutes online interaction via Zoom |
| Day 18 | Data Preprocessing and Web Usage Mining | L: 12.1-12.3 | approximately 50 minutes pre-recorded video lectures plus 50 minutes online interaction via Zoom |
| Day 19 | Data Linkage, Privacy and Bloom Filters, Social and Ethical Implications of Big Data Analytics, Cloud Computing Project | HKP: 13.4 | approximately 50 minutes pre-recorded video lectures plus 50 minutes online interaction via Zoom |
| Day 20 | Final Exam | | |