

**Instructor:** Dr. Ching-yu (Austin) Huang

**Email:** [chuang@kean.edu](mailto:chuang@kean.edu)

**Website:** <http://www.kean.edu/~chuang>

**Office Hours:** <https://huang.youcanbook.me/> Tuesday - Thursday: 9:30 – 10:30 am; 3:30 – 4:30 pm

**Class:** Monday, Wednesday (11:00 - 12:15), **synchronized ONLINE** using Blackboard Live Session

**Prerequisites:** **CPS 5740 or CPS 3740. Students without prerequisite(s) must withdraw from the class.**

| Week | Date        | Topic   | Reading | Assignment             |    |
|------|-------------|---|---------|------------------------|----|
|      | 1/18/21 Mon | <b>Martin Luther King Holiday</b>                           |         | <b>School closed</b>   |    |
| 1    | 1/20/21 Wed | Overview  |         | <b>HW 1 start</b>      | 1  |
|      | 1/25/21 Mon | Introduction, projects                                      | Ch 1    |                        | 2  |
| 2    | 1/27/21 Wed | Correlation, visualization                                  | Ch 1    | <b>Research start</b>  | 3  |
|      | 2/1/21 Mon  | Getting to know your data, SQL                              | Ch 2    |                        | 4  |
| 3    | 2/3/21 Wed  | SQL, aggregation and report, similarity                     | Ch 2    |                        | 5  |
|      | 2/8/21 Mon  | Data Processing - text and image                            | Ch 3    |                        | 6  |
| 4    | 2/10/21 Wed | Continue, review for midterm #1                             | Ch 3    |                        | 7  |
|      | 2/15/21 Mon | <b>President Day Holiday</b>                                |         | <b>School closed</b>   |    |
|      | 2/17/21 Wed | <b>Midterm #1</b>   |         | <b>HW 1 due</b>        | 8  |
| 5    | 2/22/21 Mon | Database vs Data Warehouse                                  | Ch 4    | <b>HW 2 start</b>      | 9  |
|      | 2/24/21 Wed | Data Warehouse and OLAP                                     | Ch 4    |                        | 10 |
| 6    | 3/1/21 Mon  | Associations and Correlations                               |         |                        | 11 |
|      | 3/3/21 Wed  | Mining Frequent Pattern                                     | Ch 6    |                        | 12 |
| 7    | 3/8/21 Mon  | Basic concepts of Image classification                      |         |                        | 13 |
|      | 3/10/21 Wed | Image classification - Google TensorFlow                    |         | <b>HW 2 due</b>        | 14 |
| 8    | 3/15/21 Mon | Text mining and text similarity                             | Ch 8    | <b>HW 3 start</b>      | 15 |
|      | 3/17/21 Wed | Search engine and Web crawling                              | Ch 8    |                        | 16 |
| 9    | 3/22/21 Mon | Classification: impurity, gain, decision tree               | Ch 9    |                        | 17 |
|      | 3/24/21 Wed | Classification: Linear regression and SVM, review midterm 2 | Ch 9    |                        | 18 |
| 10   | 3/29/21 Mon | <b>Midterm 2</b>  |         | <b>Midterm 2</b>       | 19 |
|      | 3/31/21 Wed | Review midterm 2 answers, SNP clustering                    |         |                        | 20 |
|      | 4/1/21 Thur | <b>Last day to withdraw with 'W' grade</b>                  |         | <b>0% refund</b>       |    |
| 11   | 4/5/21 Mon  | Cluster analysis and K-Means                                | Ch 10   |                        | 21 |
|      | 4/7/21 Wed  | Hierarchical cluster analysis                               | Ch 10   |                        | 22 |
| 12   | 4/12/21 Mon | Outlier detection   | Ch 12   |                        | 23 |
|      | 4/14/21 Wed | Continue outlier Detection                                  | Ch 12   | <b>HW 3 due</b>        | 24 |
| 13   | 4/19/21 Mon | Data mining trends and impacts                              | Ch 13   |                        | 25 |
|      | 4/21/21 Wed | <b>Research presentation</b>                                |         |                        | 26 |
| 14   | 4/26/21 Mon | <b>Research presentation</b>                                |         |                        | 27 |
|      | 4/28/21 Wed | <b>Kean Research Days – case study</b><br>Review for Final  |         | <b>Monday schedule</b> | 28 |
| 15   | 5/3/21 Mon  | <b>Final exam</b>   |         |                        | 29 |
|      | 5/5/21 Wed  | Review final exam answers                                   |         |                        | 30 |

**Late revision: 1/5/2021**

- **Course Description:** This course covers fundamentals of knowledge discovery and data mining concepts, techniques, algorithms and languages; architectures, designs and technology; and includes applications in business and the sciences.
- **Textbook:** Data Mining: Concepts and Techniques, Third Edition by Jiawei Han, Micheline Kamber and Jian Pei, 2012. ISBN: 978-0-12-381479-1.
- **Objectives:** Upon completion of this course, students will be able to:
  - A. Summarize knowledge discovery and data mining (KDDM).
  - B. Summarize the historical perspective and the social impact of KDDM.
  - C. Compare and summarize the differences between KDDM and information access/retrieval.
  - D. Analyze KDDM processes, concepts, techniques, algorithms and languages.
  - E. Demonstrate KDDM applications in business and the sciences.
- **Course Grading:** 2 Midterm (2x160 pts), 1 Final (180 pts), 3 Homework (3x100 pts), 1 Research and presentation (200 pts)

**Class Policies:**

- **This is a hands-on course. Students are required to bring personal laptop every class.**
- Instructional Methods: lecture with slides, student presentations, class discussions and exercises, and project creations.
- **A:**  $\geq 940$ , **A-:** 939-890, **B+:** 889-840, **B:** 839-800, **B-:** 799-760, **C+:** 759-720, **C:** 719-680, **D:** 679-640, **F:** below 640 (C or better is needed for CS and IT majors. B- or better is needed for CIS Graduate Program.)
- It is students' responsibilities to verify all the scores posted on **the class website** that is announced on the first day of the class.
- **Assignments should be used MySQL connecting through a Linux server that is announced on the first day of the class.**
- Mid-Term Progress information will be provided to students by the instructor via KeanWise.
- No late work will be accepted without permission in advance from the instructor. Late work is penalized.
- No make-up exams are provided without documentation of medical or family emergency. Please arrange your schedule for the next 16 weeks to arrive in class on time and prepared.
- Attendance and participation are required.
- All work turned in under your name must be your own. **No credit will be given for an assignment or homework that is copied – in part or in total – from another person.**
- **Exam policy:** During class examinations, if a student leaves the classroom for any reason during the exam time, **the exam must be turned in before the student departs, concluding the exam for that student.** The student cannot resume the exam on return later. You may not make or take a phone call, text or access the internet during class or a test.
- **Classes may not be audio or digitally captured**, unless permission has been granted by the instructor for that specific class.
- **Public posting of course materials from this class is prohibited.** Materials presented in the classroom are for the personal use of registered students during the semester only. Homework, programs, and other classwork should not be shared, distributed or publicized in any way, for the protection of other students and in respect of the intellectual property rights of the faculty and copyright owners.

**Important University Policies and Information:**

- The Kean Academic **Calendar** can be found online at <https://www.kean.edu/offices/registrars-office/academic-calendar>
- **Tutoring and learning Support services:** <http://www.kean.edu/content/tutoring-services>
- Students should review the **Student Code of Conduct**, as it discusses expectations of appropriate conduct in the classroom: <https://www.kean.edu/KU/Code-of-Conduct>
- Students are responsible to review and understand the University Academic Integrity Policy available via the following link: <https://www.kean.edu/academic-integrity>
- Read Kean **Students Rights and Responsibilities** at <https://www.kean.edu/offices/disability-services/rights-and-responsibilities>
- Register for the university's **emergency notification system** at <https://www.kean.edu/offices/university-police-0/campus-alert>
- Students must have a **valid Kean email account**. You can request one at <https://www.kean.edu/~ocisweb/forms.htm>
- **Title IX:** Title IX of the Education Amendments of 1972 (Title IX) prohibit discrimination on the basis of sex in education programs or activities. Sexual harassment in any form will not be tolerated at Kean University. Sexual harassment by students should be reported to the Office of Affirmative Action Programs, Office of the Vice President for Student Affairs or the Office of Community Standards and Student Conduct immediately. Information about the University's Sexual Misconduct Policy may be found at the following: <https://www.kean.edu/offices/policies/sexual-misconduct-policy>
- **Americans with Disabilities Statement & Non-Discrimination Statement:**  
Kean University is an affirmative action, equal opportunity institution. Students with documented disabilities who may need special instructional accommodations or who may need special arrangements in the event of an evacuation should notify the instructor as soon as possible, no later than the second week of the term. Students may contact the Office of Disability Services in Downs Hall Room 122 or call 908-737-4910 to discuss special needs.
- **Kean University Non-Discrimination Policy Statement:**  
Kean University is an affirmative action, equal opportunity institution.

**Syllabus Updates:** **Instructor reserve the right to change the syllabus at any time throughout the semester.** All changes will be announced in class and a new syllabus will be electronically available to all students. If you are enrolled in this class past the add/drop date you are subjected to all rules in this syllabus.