CS 504 Principles of Data Management and Mining
Syllabus
Fall 2023

Contact Information
Name: Binqian Yin, Ph.D.
Office Number: Buchanan Hall D215J
Email: byin2@gmu.edu
Office Hour: Mon/Wed 3:00-4:00pm

Time and Place of Class Meeting

<table>
<thead>
<tr>
<th>Section</th>
<th>Day</th>
<th>Time</th>
<th>Building</th>
<th>Room</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>Monday</td>
<td>4:30-7:10pm</td>
<td>Innovation Hall</td>
<td>132</td>
</tr>
<tr>
<td>002</td>
<td>Wednesday</td>
<td>4:30-7:10pm</td>
<td>Art and Design Building</td>
<td>2003</td>
</tr>
</tbody>
</table>

Course Description and Goals
CS 504 is an introductory, 3-credit course that explores a wide variety of topics related to data science. While it covers some aspects of databases and machine learning, the specific focus of this course is on data itself: how to store, query, and extract meaningful structure from it. Topics covered will include database management, data models, SQL, and popular models in machine learning. Students will have the opportunity to apply their knowledge through project, reports, quizzes, and exams.

The goal of this course is helping students develop problem-solving and analytical thinking skills within the context of data science. Provide an overview of modern approaches for managing and analyzing data. Enhance students' design and programming skills in order to work effectively with large data sets.

Text Book
- *Introduction to Data Mining* (Second edition) Pang-Ning Tan, Michael Steinbach, and Vipin Kumar
Grading Weights

The course has a maximum of 100 percent points (without the Extra Credits) distributed as follows:

• Quiz: 20%
• Midterm Exam: 25%
• Project: 30%
• Final Exam: 25%

You can earn Extra Credits through the following:

• Homework % (5% * 2, optional)

Both your TA and I are committed to offering reasonably detailed and prompt feedback on all assessments. If you have any questions or concerns about your grades, don't hesitate to contact us. You can reach us during scheduled office hours or through your "@gmu.edu" email address. Grades will be changed only when a grading error has been made. If there is an error in your grade, please contact us within a week of receiving the feedback. After this period, we may be unable to make any change.

Grading Scale

This course is graded according to the standard graduate grading policies, with the following cutoffs:

A+ : ≥ 95
A  : ≥ 90, < 95
A- : ≥ 88, < 90
B+ : ≥ 85, < 88
B  : ≥ 80, < 85
B- : ≥ 78, < 80
C  : ≥ 70, < 78
F  : < 70

Due Dates & Grace Period

Throughout the semester, you are allotted three "free" late days (grace period) to use for assignment and project submissions. These late days can be used all at once or distributed across project and different assignments —whatever best suits your needs. Keep in mind that each late day is triggered automatically following an assignment's due date: the first one minute after, the second 24 hours and one minute after, and the third 48 hours and one minute after. Any late submissions beyond these three days will result in a 20% penalty per day.

Penalties of late submission will be calculated based on the time of the most recent submission.
Absence of the Quiz and Exam
The student must take the exams and quizzes in the scheduled time of the course enrolled.
For the absence of quiz:
- The make-up quiz will not be offered.
- Zero will be recoded if student absence of the quiz.
For the absence of Mid-term and Final exam:
- Must get the permission from instructor before the exam at least 5 business day. The proof of absence excuse should be provided.
- The make-up exam must be taken within the specified time period.
- Zero will be recoded for midterm exam if without prior permission from instructor.
- Unexcused absence from the final exam will result in grade F for the course.

Office Hour
No appointment is needed for regular office hours. If additional meeting time is required, it can be arranged by making an appointment in advance via email.

Email
If you have any questions or concerns, please send email to byin2@gmu.edu.
When sending an email, please use the subject format 'CS504-section (choose 001 or 002) + (Your Name) + (Title of the Inquiry)'.

Course Update
Any updates related to this course will be announced on Blackboard, where you can also find all the resources for this course. Please check Blackboard regularly to ensure you do not miss any updates.

Honor Code
Please take the time to acquaint yourself with the University's overarching honor code, as well as the specific honor code for the Department of Computer Science. Any violations of academic integrity will be reported to the Office of Academic Integrity.
# CS 504 Course Schedule for Fall 2023
## Section 002

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Reading Chapter*</th>
</tr>
</thead>
<tbody>
<tr>
<td>08/23/23</td>
<td>Lecture 1: Introduction of the course</td>
<td>a. Chapter 1 &amp; 2</td>
</tr>
<tr>
<td>08/30/23</td>
<td>Lecture 2: ER &amp; EER Modeling</td>
<td>a. Chapter 3 &amp; 4</td>
</tr>
<tr>
<td>09/06/23</td>
<td>Lecture 3: Relational Database Design/ Lecture 4: Basic SQL</td>
<td>a. Chapter 5, 9 &amp; 6</td>
</tr>
<tr>
<td>09/13/23</td>
<td>Lecture 5: More SQL</td>
<td>a. Chapter 7 &amp; 30</td>
</tr>
<tr>
<td></td>
<td>Lecture 6: Database Security and Authorization</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Homework 1 (Due on 9/19/2023 11:59 PM)</td>
<td></td>
</tr>
<tr>
<td>09/20/23</td>
<td>Lecture 7: Relational Algebra</td>
<td>a. Chapter 8</td>
</tr>
<tr>
<td>09/27/23</td>
<td>Lecture 8: Basic of Functional Dependencies and Normalization</td>
<td>a. Chapter 14</td>
</tr>
<tr>
<td>10/04/23</td>
<td>Midterm Exam</td>
<td>Lecture 1~8</td>
</tr>
<tr>
<td>10/11/23</td>
<td>Lecture 9: Data and Data Preprocessing</td>
<td>b. Chapter 1 &amp; 2</td>
</tr>
<tr>
<td>10/18/23</td>
<td>Lecture 10: Probability Theory, Machine Learning Introduction</td>
<td>b. Chapter 4 &amp; 7.2</td>
</tr>
<tr>
<td>10/25/23</td>
<td>Lecture 11: Linear Regression Model</td>
<td>b. Chapter 4.6</td>
</tr>
<tr>
<td>11/01/23</td>
<td>Lecture 12: Supervised Learning Method 1: KNN &amp; Naïve Bayes Classifier</td>
<td>b. Chapter 4.3 &amp; 4.4</td>
</tr>
<tr>
<td></td>
<td>Homework 2 (Due on 11/14/2023 11:59 PM)</td>
<td></td>
</tr>
<tr>
<td>11/15/23</td>
<td>Lecture 14: Supervised Learning Method 3: ANN</td>
<td>b. Chapter 4.7</td>
</tr>
<tr>
<td>11/22/23</td>
<td>No Class — Thanks Giving Holiday</td>
<td></td>
</tr>
<tr>
<td>11/29/23</td>
<td>Lecture 15: Model Evaluation</td>
<td>b. Chapter 3.6</td>
</tr>
<tr>
<td>12/06/23</td>
<td>Final Exam</td>
<td>Lecture 9~15</td>
</tr>
</tbody>
</table>

*The “reading chapter” is based on the text book listed below:
  b. *Introduction to Data Mining* (Second edition) Pang-Ning Tan, Michael Steinbach, and Vipin Kumar