



- Clustering: Automatic grouping of similar objects into sets--including kMeans/ kMedoids, and more
- Applications in Data Mining: Search Engines and Text Retrieval with Python
- Applications in Data Mining: Social Network Mining
- Applications in Data Mining: Advanced Neural Networks
- Applications in Data Mining: Ethics and Electronic Profiling

### Course Format

---

The format for this course is MIXED HYBRID. This means that you will be working asynchronously on Mondays, we will be meeting live in person on Wednesdays, and we will have a live Zoom meeting on Friday during class time. Please make every effort to attend the Zoom meeting.

Each Sunday evening, a communication from the instructor on Blackboard outlines the goals and tasks for the week. Generally, the work rhythm in this course is as follows:

- Monday: First weekly module--readings and videos with content-based quiz, Perusal annotations, Flipgrid, or other activity, due on Blackboard by 11:59 pm CST. Estimated time: 2 hours
- Wednesday: Second weekly module--readings and videos with application-based quiz and/ or collaborative coding assignment or discussion post, due on Blackboard by 11:59 pm CST. Estimated time: 2-3 hours
  - Courses with a physical attendance requirement meet in-person in the assigned classroom during the assigned class time. If student enrollment exceeds the capacity of the classroom, you will be assigned a day for your rotation.
- Friday by 11:59 pm CST: Third weekly module--weekly lab assignment due on Blackboard by 11:59 pm CST. Estimated time: 3 hours.
  - Courses specified as Mixed Hybrid meet for a Live Discussion on Zoom.

**ATTENDANCE POLICY:** Due to COVID-19, physical attendance in any course designated as in-person on-site is **optional**.

- If you decide to attend class in person, in keeping with the Valparaiso University Code of Student Conduct, you must self-monitor and submit a daily health questionnaire before coming to campus and follow the Campus COVID Classroom Conduct outlined below once you are inside the classroom.
- If you suspect you have been exposed to COVID-19, or if you exhibit symptoms of COVID-19, please do not come to class. Instead, call the Student Health Center at 219-464-5060 and follow their directions.
- If you are ill or in one of the vulnerable populations as defined in Valparaiso University's COVID-19 Campus Safety Plan (p.17), please know that all class meetings will be livestreamed and/ or recorded and available on Blackboard. The course can be taken fully online. Do not risk your health by trying to attend class in person if it is not safe for you or others.
- If you anticipate that COVID will disrupt your ability to participate in the course and complete all assignments on time, please contact me, and we will work out a plan to help you succeed.

IF YOU ARE AN INTERNATIONAL STUDENT, IT IS YOUR RESPONSIBILITY TO VERIFY WITH YOUR ADVISOR AND WITH OIP THAT YOU ARE IN COMPLIANCE WITH APPLICABLE VISA REGULATIONS BEFORE DECIDING NOT TO ATTEND CLASS IN PERSON.

## Textbooks & Materials

---

With the exception of an available computer, the required materials for this course are free:

- Tools (for more detail, see Software and Tools):
  - A computer (Windows, Mac, or Linux) with at least 4GB RAM
  - Any kind of spreadsheet software (MS Office 365 is free for Valparaiso University Students; you can also use Google Sheets or LibreOffice)
  - An account on Rstudio Cloud at <https://rstudio.cloud/>
  - Access to Google Colab at <https://colab.research.google.com/notebooks/welcome.ipynb>
- Books and other Materials (all free through Valparaiso University):
  - Yanchang Zhao, R and Data Mining: Examples and Case Studies. [https://cran.r-project.org/doc/contrib/Zhao\\_R\\_and\\_data\\_mining.pdf](https://cran.r-project.org/doc/contrib/Zhao_R_and_data_mining.pdf)
  - Joel Grus, Data Science from Scratch: First Principles with Python; available for free at <https://www.intra.valpo.edu/librarydocs/DataSciencefromScratch>
  - Ron Zacharski, A Programmer's Guide to Data Mining. <http://guidetodatamining.com/>
  - Ledolter, Johannes. Data Mining and Business Analytics with R. Somerset, NJ, USA: John Wiley & Sons, 2013. <https://ebookcentral.proquest.com/lib/valpo-ebooks/detail.action?docID=1204741>
  - LinkedIn Learning at Valparaiso University. Access at <https://www.linkedin.com/learning/login-ent?redirect=https%3A%2F%2Fwww.linkedin.com%2Flearning%2F>

## Workload and Grading

---

This 3-credit course requires SIGNIFICANT individual and teamwork. Plan on spending 8-9 hours a week on studying and applying the material You will be completing the following tasks every week:

- **Monday Work (10 points each = 140 points).** After studying the first module of the week, you will complete a short assessment that allows you to demonstrate your understanding of what you have just studied. That can be commenting on assigned reading or video viewing in Perusall or Flipgrid, a quiz in Blackboard, researching a topic and posting it on the Discussion Board, or even posting a video of you performing an exercise--or a combination of these. Monday Work varies from week to week, but always focuses on the assigned materials. It CANNOT be made up and is due by 11:59 pm CST on the Monday of the week in which it is given.
- **Wednesday Work (10 points each = 140 points).** Typically, Wednesday Work consists of one or two short coding problems to help you explore and practice your new skills for any of the end-of-week labs. This is collaborative and shared work, often in a Discussion thread with practice code (due Wednesdays) and responses or peer reviews (due Thursdays), or in Flipgrid format. Wednesday Work CANNOT be made up and is due by 11:59 pm CST on the Wednesday (and Thursday as needed) of the week in which it is given.

- **Friday Labs (10\*10 points each=100 points):** Some weeks contains a lab assignment about a theoretical, practical, or programming problem. Students will be assigned a team with which they may collaborate, but **all students will submit their own solution**. Solutions must posted to Blackboard by 11:59 pm CST on FRIDAY. NO EMAIL SUBMISSIONS ACCEPTED. There are 11 labs on the syllabus; the lowest lab grade will be dropped.
- **Midterm Exam (20 points):** The midterm exam requires that you solve set of problems on a database you know, based on the first 5 lab assignments. Please consult the workplan below for the midterm date.
- **Final Exam (100 points):** The final exam will be an individual set of tasks performed on a database that you will receive 24 hours before the exam opens. Please consult the workplan below for the course final date and time. You will have 110 minutes to do the following:
  - Solve the course final tasks you are given and practice typing the commands
  - In a max. 15-minute video, record your solution path with Screen-cast-o-matic (and webcam, in which you explain what you are doing and typing and why) and post your video to YouTube, then paste the YouTube link into the assignment on Blackboard. YOU MUST PRACTICE AND TIME THIS A FEW TIMES. UPLOADING A 15-MINUTE VIDEO TAKES TIME!

NOTE: You are responsible for the quality of your WiFi. Due to FERPA, no submissions by email are accepted. NO EXCEPTIONS.

You can earn up to 500 points in this course. Extra credit assignments can help you get there faster.

#### Letter Grade Conversion:

A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F
> 93%	90-93%	87-89.9%	83-86.9%	80-82.9%	77-79.9%	73-76.9%	70-72.9%	67-69.9%	63-66.9%	60-62.9%	< 60%

#### Assignment Submission, Late Work, and Academic Honesty

---

- **Assignment Submission:** All Assignments must be submitted on Blackboard. **BECAUSE OF FERPA LEGISLATION, I cannot accept any emailed assignments.**
- **Late Work:** Work is considered late if not posted to Blackboard by 11:59 pm CST of the day on which it is due.
  - ONLY Friday labs may be submitted late and will lose 50% of the grade. Any Friday labs that you did not complete, but now want to submit for grading MUST be submitted by the WEDNESDAY before the last day of instruction, 11:59 pm CST.
  - Since quizzes, tests, practice code/ peer review exercises, or discussion assignments are timebound, they can NOT be made up at a later date.
- **Academic Honesty:** This course upholds the Valparaiso University Honor Code, which permits students to do their academic work in an atmosphere of responsible freedom. For you, this means that **all work you submit for this course must be your own**.
  - **For labs and Discussions (Practice Code/ Peer Reviews):** If you decide to include anyone else's words or code (from blogs, webpages, coding forums like GitHub or Stackoverflow, purchased solutions, etc.), you must:

1. Give a clear **source citation** (including the exact location from which you copied these words or lines of code)
  2. Include an **explanation in your own words** of what the cited passage means or what the copied code does, why it works, and why it is better than your own.
- **For Midterm and Final:** You are allowed to use the course book, the instructor slides, and the work you have submitted in this course as authorized aid. You are NOT allowed to use any information from the internet or your friend/ current or former colleague/ pet rock or any source other than what is defined as authorized aid. Cell phones and smart watches must be left on a designated table in the room.

When the definition of unauthorized aid is in question, it is your responsibility to clarify your understanding of it with the instructor. Ignorance is not a valid excuse for violations of the Honor Code. Students should report suspected violations to the Honor Council.

In addition, you must write and sign with your name the following statement on all course work:

**I have neither given nor received, nor have I tolerated others' use of unauthorized aid.**

For more information about Valparaiso University's Academic Honor Code, case review cycles, and potential penalties, please refer to <http://www.valpo.edu/student/honorcouncil/index.php>

## Schedule

Week	Start Date	Topic	Applicable Materials and Tools (includes LinkedIn Learning courses where applicable)	Graded Work Due
1	08/17	Big Data and Basic Statistics Review	<ul style="list-style-type: none"> <li>• Grus 1</li> <li>• LinkedIn Learning course on Introduction to Data Science</li> <li>• Other materials as posted in Blackboard</li> </ul>	Monday Work (10 points) Wednesday Work (10 points)
2	08/23	Basic Statistics in R and Python: Measures of Central Tendency & Dispersion, Distance	<ul style="list-style-type: none"> <li>• Grus 5, 22</li> <li>• Zacharski 2</li> <li>• Other materials as posted in Blackboard</li> </ul>	Monday Work (10 points) Wednesday Work (10 points) Friday Lab_01 (10 points)
3	08/30	Wrangling Data: Preprocessing and Visualizing	<ul style="list-style-type: none"> <li>• Grus 3, 10</li> <li>• Other materials as posted in Blackboard</li> </ul>	Monday Work (10 points) Wednesday Work (10 points) Friday Lab_02 (10 points)
4	09/06	Advanced Statistics in R and Python: Measures of Association:	<ul style="list-style-type: none"> <li>• Grus 7, 22</li> <li>• Zacharski 2</li> <li>• Other materials as posted in Blackboard</li> </ul>	Monday Work (10 points) Wednesday Work (10 points)

		Covariance, Correlation, Cosine Similarity, Chi Square		Friday Lab_03 (10 points)
5	09/13	Linear Modeling with Regressions (Simple Linear, Multiple, Logistic)	<ul style="list-style-type: none"> <li>• Zhao 5</li> <li>• Ledolter 3, 4, 7</li> <li>• Grus 7, 14, 15, 16</li> <li>• Other materials as posted in Blackboard</li> </ul>	Monday Work (10 points) Wednesday Work (10 points) Friday Lab_04 (10 points)
6	09/20	Classification & Recommendation Systems: kNearest Neighbor; Naïve Bayes and Probability	<ul style="list-style-type: none"> <li>• Grus 12, 6 &amp; 13</li> </ul>	Monday Work (10 points) Wednesday Work (10 points) Friday Lab_05 (10 points)
7	09/27  <b>10/02</b>	Classification & Recommendation Systems: Decision Trees and Random Forest  <b>MIDTERM</b>	<ul style="list-style-type: none"> <li>• Zhao 4</li> <li>• Grus 17</li> <li>• Ledolter 13</li> <li>• Other materials as posted in Blackboard</li> </ul>	Monday Work (10 points) Wednesday Work (10 points) <b>MIDTERM</b>
8	10/04	Classification with Gradient Descent and Neural Networks	<ul style="list-style-type: none"> <li>• Grus 8, 18</li> <li>• Other materials as posted in Blackboard</li> </ul>	Monday Work (10 points) Wednesday Work (10 points) Friday Lab_06 (10 points)
9	10/11	Clustering Models: kMeans, Hierarchical Clustering, DBScan	<ul style="list-style-type: none"> <li>• Zhao 6</li> <li>• Zacharski 8</li> <li>• Ledolter 15</li> <li>• Grus 19</li> <li>• Other materials as posted in Blackboard</li> </ul>	Monday Work (10 points) Wednesday Work (10 points) Friday Lab_07 (10 points)
10	10/18	Applications in Data Mining: Deep Learning in TensorFlow	<ul style="list-style-type: none"> <li>•</li> </ul>	Monday Work (10 points) Wednesday Work (10 points) Friday Lab_08 (10 points)

11	10/25	Applications in Data Mining: Search Engines and Social Network Mining in R and Python	<ul style="list-style-type: none"> <li>• Zhao 10, 11</li> <li>• Zacharski 7</li> <li>• Ledolter 19, 20</li> <li>• Grus 9, 20, 21</li> <li>• Other materials as posted in Blackboard</li> </ul>	Monday Work (10 points) Wednesday Work (10 points)
12	11/01	Applications in Data Mining: Webscraping and Sentiment Analysis in R and Python	<ul style="list-style-type: none"> <li>• Zhao 10</li> <li>• Zacharski 7</li> <li>• Ledolter 19</li> <li>• Grus 20</li> <li>• Other materials as posted in Blackboard</li> </ul>	Monday Work (10 points) Wednesday Work (10 points) Friday Lab_09 (10 points)
13	11/08	Ethics and Data Mining	<ul style="list-style-type: none"> <li>• Cathy O'Neil, <a href="#">Weapons of Math Destruction</a> (2016). Ch. 4, 5, 6, 10</li> <li>• Materials about Facebook and Cambridge Analytica</li> </ul>	Monday Work (10 points) Wednesday Work (10 points) Friday Lab_10 (10 points)
14	11/15	Course Summary, Final Exam Preparation		Monday Work (10 points) Wednesday Work (10 points) Friday Lab_11 (10 points)
<b>Final</b>	11/23	Course Final Exam due 11:30-1:10 pm CST	<p><b>USE:</b> Books&amp;slides&amp; assignments</p> <p><b>DO NOT USE:</b> Friends, the internet, etc.</p>	<b>Course Final Exam</b>

## APPENDIX A: University Policies

### Diversity and Inclusion

---

Valparaiso University aspires to create and maintain a welcoming environment built on participation, mutual respect, freedom, faith, competency, positive regard, and inclusion. This course will not tolerate language or behavior that demeans members of our learning community based on age, ethnicity, race, color, religion, sexual orientation, gender identity, biological sex, disabilities (visible and invisible), socio-economic status, or national origin. The success of this class relies on all students' contribution to an anti-discriminatory environment where everyone feels safe, welcome, and encouraged to engage, to explore, and ultimately, "to embark on a rewarding personal and professional journey" (Pres. Heckler).

### Title IX

---

Valparaiso University strives to provide an environment free of discrimination, harassment, and sexual misconduct (sexual harassment, sexual violence, dating violence, domestic violence, and stalking). If you have been the victim of sexual misconduct, we encourage you to report the incident. If you report the incident to a University faculty member or instructor, she or he must notify the University's Title IX Coordinator about the basic facts of the incident. Disclosures to University faculty or instructors of sexual misconduct incidents are not confidential under Title IX. Confidential support services available on campus include: Sexual Assault Awareness & Facilitative Education Office "SAAFE" (219-464-6789), Counseling Center (219-464-5002), University Pastors (219-464-5093), and Student Health Center (219-464-5060). For more information, visit <http://www.valpo.edu/titleix/>.

### Access and Accommodation Services

---

The Access & Accommodations Resource Center (AARC) is the campus office that works with students to provide access and accommodations in cases of diagnosed mental or emotional health issues, attentional or learning disabilities, vision or hearing limitations, chronic diseases, or allergies. You can contact the office at [aarc@valpo.edu](mailto:aarc@valpo.edu) or 219.464.5206. Students who need, or think they may need, accommodations due to a diagnosis, or who think they have a diagnosis, are invited to contact AARC to arrange a confidential discussion with the AARC office. Further, students who are registered with AARC **are required to contact their professor(s) if they wish to exercise the accommodations** outlined in their letter from the AARC.

### Academic Support

---

To get help, use the [Academic Success Center \(ASC\) online directory](http://valpo.edu/academicsuccess) (valpo.edu/academicsuccess) or contact the ASC ([academic.success@valpo.edu](mailto:academic.success@valpo.edu)) to help point you in the right direction for academic support resources for this course. Valpo's learning centers offer a variety of programs and services that provide group and individual learning assistance for many subject areas. These learning centers include:

- [Tutoring Lab](#): Serves the academic needs of undergraduate and graduate students – tutors offer suggestions on organization of papers, assist in research and citations, and help in understanding difficult assignments. Additional one on one tutoring is also available.
- [Writing Center](#): Writing Consultants provide proofreading and editing assistance for papers and assignments.

### Class Cancellations

---

Notifications of class cancellations will be made through Blackboard with as much advance notice as possible. It will be both posted on Blackboard and sent to your Valpo e-mail address. If you don't check your Valpo e-mail account regularly or have it set-up to be forwarded to your preferred e-mail account,

you may not get the message. Please check Blackboard and your Valpo e-mail (or the e-mail address it forwards to) before coming to class.

### **Emergencies**

---

VU's Emergency Notification System (ENS) uses multiple forms of communication, including e-mail, building alarms, outdoor sirens, message boards, computer alerts, Twitter, and public address messaging. Please review the specific procedures for this class found in Blackboard. Remember: "Siren inside, GO outside; Siren outside, GO inside." To evacuate, gather your personal belongings quickly and proceed to the nearest exit. Do not use the elevator. To shelter in place, move away from the windows and stay low to the ground; lock or barricade the door if there is a threat of violence.

**APPENDIX B: Learning Objectives****Student Learning Objectives—Valparaiso University**

---

1. Demonstrate theoretical and practical knowledge as well as the intellectual skills and creative capacities pertinent to their respective fields of study.
2. Solve both conceptual and applied problems by integrating broad-based knowledge, evidence-based reasoning, and informational literacy.
3. Practice experiential, interdisciplinary, and collaborative learning in both academic and co-curricular pursuits.
4. Communicate effectively in oral, written, and digital forms in increasingly complex contexts.
5. Engage in cross-cultural dialogue and experiences with the requisite knowledge to succeed in a diverse, global community.
6. Develop character, integrity, and wisdom as they discern their vocations and prepare to ethically lead and serve church and society.

<http://www.valpo.edu/institutional-effectiveness/files/2017/08/University-wide-SLOs.pdf>

**Student Learning Objectives—Computer Science Majors (Undergraduate)**

---

1. To understand and practice methods of inquiry and strategies of interpretation within the student's field of study.
  - a. Students will master several programming environments.
  - b. Students will learn to identify and isolate problems.
2. To master the knowledge and skills pertinent to the student's field of study.
  - a. Students will acquire an extensive technological vocabulary.
  - b. Students will become comfortable with a wide range of technology environments.
3. To effectively articulate the ideas, concepts, and methods through written and oral presentation.
  - a. Students will be taught how to make formal oral presentations and be required to give 6 such presentations during their program.
  - b. Students will write numerous thorough papers requiring extensive research. They will be required to use the services of the writing center.
4. To understand the connection between their knowledge and skills on one hand and their professional identity, responsibilities, and demands on the other.
  - a. Students will understand the implications of legal and professional regulations as they relate to information technology.
  - b. Students will study how technology can be made available to people that are traditionally less advantaged.
5. To integrate knowledge and methods of their study with cognates and other disciplines.
  - a. Students will learn techniques of modeling data from other disciplines.
  - b. Students will study human factors in IT.
6. To practice ethical and cultural sensitivity as it relates to professional and personal responsibility.
  - a. Students will examine a wide range of ethical issues related to technology and the potential side effects on people and the environment.
  - b. Students will explore the relationship between IT and ethnic and cultural diversity.

**Student Learning Objectives—Graduate School**

---

1. Students will understand and practice methods of inquiry and strategies of interpretation within the student's field of study.
2. Students will master the knowledge and skills pertinent to the student's field of study.
3. Students will effectively articulate the ideas, concepts, and methods through written and oral presentation.
4. Students will understand the connection between their knowledge and skills on the one hand and their professional identity, responsibilities, and demands on the other.
5. Students will integrate knowledge and methods of their study with cognates and other disciplines.
6. Students will study, reflect upon, and practice ethical behavior and cultural sensitivity as they relate to professional and personal responsibility.

**Student Learning Objectives—Information Technology Program (Graduate)**

---

1. To understand and practice methods of inquiry and strategies of interpretation within the student's field of study.
  - 1A. Students will master several programming environments.
  - 1B. Students will learn to identify and isolate problems.
2. To master the knowledge and skills pertinent to the student's field of study.
  - 2A. Students will acquire an extensive technology related vocabulary.
  - 2B. Students will become comfortable using a wide range of technology environments.
3. To effectively articulate the ideas, concepts, and methods through written and oral presentation.
  - 3A. Students will be taught how to make formal, oral presentations and be required to give 6 such presentations during their program.
  - 3B. Students will write numerous, thorough papers requiring extensive research. They will be required to use the services on the writing center.
4. To understand the connection between their knowledge and skills on one hand and their professional identity, responsibilities, and demands on the other.
  - 4A. Students will understand the implications of legal and professional regulations as they relate to information technology.
  - 4B. Students will study how technology can be made available to people that are traditionally less advantaged.
5. To integrate knowledge and methods of their study with cognates and other disciplines.
  - 5A. Students will learn techniques of modeling data from other disciplines.
  - 5B. Students will study human factors in IT.
6. To practice ethical and cultural sensitivity as it relates to professional and personal responsibility.
  - 6A. Students will examine a wide range of ethical issues related to technology and the potential effects on people and the environment.
  - 6B. Students will explore the relationship between IT and ethnic and cultural diversity.

**APPENDIX****Student Learning Objectives—Graduate School**

---

1. Students will understand and practice methods of inquiry and strategies of interpretation within the student's field of study.
2. Students will master the knowledge and skills pertinent to the student's field of study.
3. Students will effectively articulate the ideas, concepts, and methods through written and oral presentation.
4. Students will understand the connection between their knowledge and skills on the one hand and their professional identity, responsibilities, and demands on the other.
5. Students will integrate knowledge and methods of their study with cognates and other disciplines.
6. Students will study, reflect upon, and practice ethical behavior and cultural sensitivity as they relate to professional and personal responsibility.

**Student Learning Objectives—Information Technology Program**

---

1. To understand and practice methods of inquiry and strategies of interpretation within the student's field of study.
  - 1A. Students will master several programming environments.
  - 1B. Students will learn to identify and isolate problems.
2. To master the knowledge and skills pertinent to the student's field of study.
  - 2A. Students will acquire an extensive technology related vocabulary.
  - 2B. Students will become comfortable using a wide range of technology environments.
3. To effectively articulate the ideas, concepts, and methods through written and oral presentation.
  - 3A. Students will be taught how to make formal, oral presentations and be required to give 6 such presentations during their program.
  - 3B. Students will write numerous, thorough papers requiring extensive research. They will be required to use the services on the writing center.
4. To understand the connection between their knowledge and skills on one hand and their professional identity, responsibilities, and demands on the other.
  - 4A. Students will understand the implications of legal and professional regulations as they relate to information technology.
  - 4B. Students will study how technology can be made available to people that are traditionally less advantaged.
5. To integrate knowledge and methods of their study with cognates and other disciplines.
  - 5A. Students will learn techniques of modeling data from other disciplines.
  - 5B. Students will study human factors in IT.
6. To practice ethical and cultural sensitivity as it relates to professional and personal responsibility.
  - 6A. Students will examine a wide range of ethical issues related to technology and the potential effects on people and the environment.
  - 6B. Students will explore the relationship between IT and ethnic and cultural diversity.