



## IT-533 STEM Summer I 2023 Data Mining

<b>Dates: Summer I</b> 05/16/2023- 06/23/2023	<b>Mode: ONLINE</b> • Expected Activity: MWF	<b>Location:</b> <a href="http://blackboard.valpo.edu">http://blackboard.valpo.edu</a>
<b>Instructor:</b> Sonja Streuber	<b>Office Hours:</b> M-F 6-8 pm CST on Google Meet	<b>Contact:</b> <a href="mailto:sonja.streuber@valpo.edu">sonja.streuber@valpo.edu</a>

### Introduction



Welcome to IT-533 STEM, Data Mining! Motivated by the growth of data collections routinely kept by many organizations, and by the high potential value of patterns discovered in those collections, Data Mining focuses on identifying useful regularities in large data sets, turning these regularities into models, and using these models to forecast future data behavior. For instance, bar code and loyalty card readers at supermarket checkouts generate data that can be used to predict future shopping trends among a certain

demographic, and daylight length, temperature, and rainfall data collected at strategic locations is used for weather forecasting. This makes Data Mining a broad area that integrates techniques from machine learning, statistics, artificial intelligence, and database systems, for the analysis of large volumes of data. This course gives a summary exposition of these techniques and their tools and is built for teachers delivering the knowledge to middle- and high-school learners.

### Learning Objectives

Students who successfully complete this course will gain a strategic and analytical toolkit essential to Data Mining. Students will be able to:

- Analyze large sets of data and uncover patterns within the data
- Use mathematical algorithms to uncover patterns in both web and regular data
- Predict data based on the patterns discovered previously
- Use tools and statistical analysis to analyze data, including MS Excel and Python
- Apply the learned techniques to common data mining problems
- Describe legal, ethical and public relations implications of data mining
- Plan to teach data mining concepts

### Topics Covered

- Big Data and Statistics with spreadsheets and Python
- Preprocessing: Data Preparation for Data Mining
- Linear Modelling to predict a continuous-valued attribute associated with an object--including linear and logistic regression
- Classification to identify to which category an object belongs--including Decision Trees, kNearest Neighbor and Random Forest; probability-based models such as Naïve Bayes; and Neural Networks/ Deep Learning

- Applications in Data Mining: Search Engines and Text Retrieval with Python
- Applications in Data Mining: Social Network Mining
- Applications in Data Mining: Data Visualization
- Applications in Data Mining: Ethics and Electronic Profiling

## Course Format

---

### Instructional Method: ONLINE ASYNCHRONOUS.

This course runs very similar to a flipped classroom/inverted learning design. Lecture content will be video recordings and readings posted on Sunday at midnight, when the weekly Content folder opens on Blackboard – these readings or videos must be completed prior to attending an office hour or the Tuesday Walkthrough session associated with that material. A weekly announcement on Blackboard will lay out the exact schedule for the week on the previous Thursday. Please bring questions that you have from the lecture material to the Tuesday Walkthrough so that we can work on them "in class." If you cannot attend the Tuesday Walkthrough, please bring your questions to the office hour sessions, in which I expect to work with you through lecture concepts, syntax problems, and the bigger coding assignments.

### Estimated Time on Task:

This is a 3-credit course. Students are expected to spend about 9 hours of work on this course each week. These 9 hours are distributed as follows:

- Monday/ Tuesday study of concepts based on instructor-provided material: 2 hours
- Tuesday attendance at online lecture (synchronous) or watching of video recording: 1.5 hours
- Wednesday Work (demonstration that weekly material has been studied): 1-2 hours
- Friday Lab Assignment (application of studied weekly material): 3-4 hours
- Assignment work is including, but not limited to:
  - Reading
  - Writing
  - Computation
  - Exploration/research
  - Building student assignments and lesson plans
  - Test-taking time
  - Time spent on administrative tasks (e.g. learning software, uploading files)

## Textbooks & Materials

---

- **Tools:**
  - A computer (Windows, Mac, or Linux) with at least 4GB RAM and a working webcam
  - Any kind of spreadsheet software—EXCEPT for Apple Numbers (free for VU students)
  - A @valpo.edu Google Account to use the Google Colab Platform (free)
  - A teacher license to Tableau (free for 1 year)
  - Free account on Kahoot!, Gimkit, Edpuzzle, or another K-12 content delivery platform

- **Books and other Materials (free):**
  - All materials are instructor-generated and available free of charge for use in this course

### Workload and Grading

---

This 3-credit course requires SIGNIFICANT individual and teamwork (plan on 5-9 hrs per week). You will be completing the following tasks every week:

- **Weekly Technical Work ((7\*10) =70 points):** Every Wednesday, students complete an assignment of a technical nature in a piece of spreadsheet software, in Python, or in Tableau and display their results in a quiz; this will prepare them for the pedagogy lab on Friday. Due to their timebound nature, these assignments **CANNOT BE MADE UP**.
- **Weekly Pedagogy Assignment ((6\*10) = 60 points):** Every Thursday, students complete the following for their end-of-course portfolio: One lesson plan outlining the week's classroom activities, including one or more student-centered assignments and one Flipgrid video presenting their work for the week. These assignments **CANNOT BE MADE UP**.
- **Weekly Peer Review ((6\*5) = 30 points):** Every Friday, students review one of their peers' pedagogy materials and write a constructive one-paragraph evaluation. These assignments **can be made up for 50% credit**, but must be submitted by the Wednesday BEFORE the end of the term.
- **Final Portfolio (40 points):** The final portfolio consists of all Weekly Pedagogy Assignments, revised based on the Peer Reviews obtained, with a course syllabus and a learner-centered video introduction of the course. The goal of this portfolio is to give the teacher-student the opportunity to leave this course with their own course package, ready for implementation in their classroom.

Students can earn up to 200 points in this course. No extra credit assignments will be given.

### Letter Grade Conversion:

A	>93%	A-	90 to 93%	B+	87 to <90%	B	83 to <87%	B-	80 to <83%
C+	77 to <80%	C	73 to < 77%	C-	70 to <73%	F	<70%		

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

---

- **Assignment Submission:** All Assignments must be submitted on Blackboard. **No emailed Assignments will be accepted.**
- **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 late Peer Reviews may be made up for 50% credit.
- **Academic Honesty:** All work you submit for this course must be your own. You may NOT use anyone else's words (from blogs, webpages, purchased solutions, etc.) without giving a clear source citation. If you are unsure, consult <http://www.plagiarism.org/> or the Writing Center. 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>

**Any work suspected in noncompliance the Valparaiso University Honor Code will receive 0 points and be referred to the Graduate School for adjudication.**

---

## 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) 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:

- [Graduate Tutoring Lab](#): Serves the academic needs of 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](#): Primarily serves the needs of undergraduate students, but is also available for Graduate students. Writing Consultants provide proofreading and editing assistance for papers and assignments.

### **Library Services**

---

The librarian best able to help you navigate information resources for independent research or additional reading is listed on the library research guide for our department. Click the link to Library Guides within the Blackboard table of contents for this course.

### **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.

## Schedule

Week	Start Date (all 2018)	Topic	Graded Work Due <ul style="list-style-type: none"> <li>• <b>W=Wed 11:59 pm CST</b></li> <li>• <b>R=Thu 11:59 pm CST</b></li> <li>• <b>F=Fri 11:59 pm CST</b></li> </ul>
1	05/16	Big Data and Basic Statistics Review  Introduction to Python with Google Colab	<b>R</b> Introduction Video with Background  <b>F</b> Technical Work 1: Basic Statistics in Python with Google Colab
2	05/21	Advanced Statistics in Python: Correlation, Cosine Similarity, Chi Square, Simple Regression	<b>W</b> Technical Work 2: Interpret an assigned dataset with Measures of Association; predict simple values  <b>R</b> Pedagogy Lab 1: Lesson Plan & Assignment & Teaching Video  <b>F</b> Peer Review 1
3	05/28	Supervised Machine Learning: Logistic Regression and simple Classification with k Nearest Neighbor, and Naïve Bayes	<b>W</b> Technical Work 3: Build Logistic Regression, k Nearest Neighbor, and Naïve Bayes on an assigned dataset & interpret the results  <b>R</b> Pedagogy Lab 2: Lesson Plan & Assignment & Teaching Video  <b>F</b> Peer Review 2
4	06/04	Classification: From Random Forest to Deep Learning with Neural Networks	<b>W</b> Technical Work 4: Use ensemble learning techniques with Random Forest and simple Neural Networks  <b>R</b> Pedagogy Lab 3: Lesson Plan & Assignment & Teaching Video  <b>F</b> Peer Review 3
5	06/11	Live Text Analysis and Deep Learning: Neural Networks with real-time data	<b>W</b> Technical Work 5: Configure a real-world API to access and process text data with Deep Learning  <b>R</b> Pedagogy Lab 4: Lesson Plan & Assignment & Teaching Video  <b>F</b> Peer Review 4

6	06/18	Data Visualization and Ethics	<b>W</b> Technical Work 6: Build useful visuals around data & evaluate your analysis based on ethical guidelines  <b>R</b> Pedagogy Lab 5: Lesson Plan & Assignment
	06/23	FINAL PORTFOLIO DUE	<b>F FINAL PORTFOLIO</b>

**Final Project (8-15 pages) (100 points)—Completed in stages; check the syllabus for due dates**

---

This course requires a major project, which you will complete in your assigned groups. For this project, you will **be assigned** a useful real-world dataset and **conduct** your own data mining activities as prompted through the weekly questions.

The report must be 8-15 pages, excluding source code but including figures, and should be well written. **It must have the following 9 sections (please copy the section headings into your final writeup):**

1. What does the dataset describe? -- 10 pts
2. What preprocessing was needed for what attributes? What replacement strategies for missing values were used and why? --10 pts
3. Which are the three most important attributes? Provide their basic statistical measures (see weeks 1 and 2) and explain what these measures tell you about the entire dataset. – 10 pts
4. Show how the three most important attributes relate to each other and to the class attribute (if it exists in your dataset) using correlation and regression analysis. What does this analysis tell you about the entire dataset? – 10 pts
5. Which association algorithms and tuning techniques did you use and what information did they give you about your dataset? – 10 pts
6. Which classification algorithms and tuning techniques did you use and what information did they give you about your dataset? – 10 pts
7. Which clustering algorithms and tuning techniques did you use and what information did they give you about your dataset? – 10 pts
8. Interpretation of results: Which data mining method used in points 3-7 provides the most useful and important information about your dataset? —10 pts
9. What action would you recommend management to take as a result of your data analysis, and where are the limitations of your analysis? Use at least 5 sentences to explain how you arrived at your recommendation, which managerial conclusions from your analysis would be useful, and which ones would be erroneous. – 20 pts

Please use the Final\_Project\_n submissions that you have worked on all semester as the basis for your Final Project submission and polish them into a professional-looking analysis paper. This includes proper English grammar, vocabulary, and expressions, and proper formatting.

**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.