

Valparaiso University
IT-533-OL Data Mining

Syllabus

Instructor: Saso Poposki

Email: saso.poposki@valpo.edu

Office: Meier Hall 145-D. (by appointment)

Session: August 20 – December 6, 2018

Textbook & Materials:

1) *Python for Data Analysis* (1st Edition).

DataWrangling with Pandas, NumPy, and IPython

Author: Wes McKinney

ISBN: 9781449319793

O'Reilly Media 2013.

2) *Data Science from Scratch: First Principles with Python* (1st Edition).

Author: Grus

ISBN: 9781491901427

O'Reilly Media 2015.

3) *A Programmer's Guide to Data Mining* (The Ancient Art of the Numerati)

Ron Zacharski.

- free online data mining textbook: <http://guidetodatamining.com/>

4) *Data Mining : Concepts and Techniques* (3rd Edition).

Han, Jiawei, Kamber, Micheline, and Pei, Jian.

ISBN: 9780123814791

Saint Louis, MO, USA: Morgan Kaufmann, 2011. ProQuest ebrary.
<http://site.ebrary.com/lib/valpo/detail.action?docID=10483440&p00=han+jiawei>

Available as eBook through the Valpo Library URL above. Login required.

Software:

- 1) Anaconda
<https://www.continuum.io/downloads>
- 2) Datasets from the UCI Machine Learning Repository at
<http://archive.ics.uci.edu/ml/datasets.html>

Course Information: Data mining is a broad area that integrates techniques from several fields including machine learning, statistics, pattern recognition, artificial intelligence, and database systems, for the analysis of large volumes of data. This course gives a wide exposition of these techniques and their software tools. This course will illustrate data mining process and how the technology works with sample live applications of data mining. The following topics will be covered:

- Basic concepts, applications and trends in data mining
- Relationship between data mining, data warehouse, and query tools
- Data preparation for the data mining process
- Model building, algorithms and technology:
- Supervised learning:
 - Classification and Prediction (Decision Trees & Bayesian Networks)
- Unsupervised learning
 - Clustering (k-means and hierarchical clustering)
- Evaluation of the data mining model; comparisons of different data mining models
- Visualization using Data Mining

Data mining aims at finding useful regularities in large data sets. Interest in the field is motivated by the growth of computerized data collections which are routinely kept by many organizations and commercial enterprises, and by the high potential value of patterns discovered in those collections.

Course 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
- Describe legal, ethical and public relations implications of data mining

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.
 - 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.

Attendance Policy: Attendance/ participation at all classes and labs is required

Homework

This 3-credit course requires significant research and teamwork. You will be completing the following tasks:

- **Attendance & Preparation:** Depending on the material, each week, you will be completing a lab, a quiz, or answering a short question in addition to the weekly discussion board post. Your attendance and participation will be measured by the quality of your answers. Answers are due by 11:59 pm CST each Sunday evening of the week .

- Homework Assignments: There are five complex, work-intensive homework assignments, which consist of theoretical/ mathematical or programming exercises. Homework must be posted on Blackboard by 11:59 pm CST on Sunday evening.
- Final Project (100 points): The final project requires that pursues an interesting question about a dataset which you have found (there are numerous sources on the internet). You will draft the project in stages and apply everything that you learn in this course, including preprocessing (with Excel, Python, or MatLab), algorithm selection and application, visualization, and interpretation.

Tests: Tests and quizzes must be taken on the announced dates and times. The VU Honor Code applies to all tests and quizzes.

Grading System: Your grade will be based upon the following;

Attendance/Discussion	10%
Project	15%
Quizzes	20%
Assignments	25%
Exams	30%
Total	100%

The final grading scale will be no stricter than:

93% >= A, 83% >= B, 73% >= C, 60% >= D

Honor Code: You must adhere to and sign the University's Honor Code on all your work.

Valparaiso University operates under a student initiated Honor Code. Each student studying at Valparaiso University agrees to conform to that Code. All new students should read the following explanation, since they must sign their registration form indicating that they will abide by the Honor Code. The Honor Code is an integral part of VU and permits students to do their academic work in an atmosphere of responsible freedom. The Honor Code is based on the highest principles of Christian ethics and morality and presumes every student is willing to maintain honesty in all academic work, as well as other phases of university living. Students are required to sign the pledge, indicating that they have submitted honest work and have not allowed the dishonesty of others to erode the integrity of the Honor System:

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

The Honor Code applies to all students registered for academic credit at Valparaiso University. Students have the responsibility for not using, giving, or tolerating unauthorized aid. When the definition of unauthorized aid is in question, students should

as the instructor to interpret the application of the Honor Code. In case of further doubt, students share the responsibility of clarifying the definition of unauthorized aid. Ignorance is not acceptable as a valid excuse for violations of the Honor Code. Students should report suspected violations to the Honor Council.

Support Services

Disability Support Services

Please contact Dr. Sherry DeMik, Director of Disability Support Services, at x6956, if you believe you have a disability that might require a reasonable accommodation in order for you to perform as expected in class. Dr. DeMik will work with you and me to make sure you receive any reasonable accommodations needed as a result of a disability.

Academic Support

To get help with this course, the best place to start is to work with your professor during office hours and to ask your professor if there are any Help Sessions or department-level tutoring offered for this course. Another step is to use the Academic Success Center (ASC) online directory (www.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 to 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.
- Language Resource Center: Provides tutoring and other resources for language study as well as opportunities for authentic language use through conversation programs, enrichment activities and other exchanges.
- Academic Success Center: Provides referral service to help connect students with appropriate resources on campus to support their academic achievements.

Class Cancellations and Other Announcements

Notifications of class cancellations and other announcements will be made through Blackboard with as much advance notice as possible. It will be both posted in Blackboard and sent to your Valpo email address. If you don't check your Valpo email

regularly or have set up to be forwarded to your preferred email account, you may not get the message. Please check Blackboard and your Valpo email before coming to class.