

## CSC 458 – Data Mining and Predictive Analytics I, Fall 2019 First Day Handout

Wednesday 6-8:50 PM in Old Main 158.

Remote class-time attendance option via Zoom web tools (Real Time Virtual Classroom).

<http://faculty.kutztown.edu/parson> ← Browse from here down to our course page.

**Dr. Dale E. Parson, Old Main 260, [parson@kutztown.edu](mailto:parson@kutztown.edu), 484-646-4296**

**Office Hours Monday 12:30-2:30, Tuesday 3-4, Wednesday 2:30-4:30, or by appointment.**

Many academic and commercial endeavors apply the techniques of data mining and predictive analytics to their data sets. Students taking this course will learn methods and software tools for locating and obtaining data of interest, for preparing data for semi-automated analysis, for interacting with software tools in analyzing data for patterns, for visualizing structural and dynamic patterns in data, and for designing systems that respond to patterns in data. Data cleaning and formatting require some programming in a modern scripting language. Other course activities include learning to use off-the-shelf software tools to accomplish the tasks of data analysis.

**Prerequisite:** C or better in CSC223 AND C or better in a statistics course AND junior status, or acceptance into the KU CSC graduate program.

**Textbook:** *Data Mining: Practical Machine Learning Tools and Techniques*, **Fourth Edition**, Witten, et. al., ISBN 978-0128042915. You can probably buy a discounted copy of the **Third Edition** (better edition). Either edition is fine, since we will not be using the two chapters added to the Fourth.

**Grading** (A = 92:100, A- = 90:91, B+ = 87:89, B = 82:86, B- = 80:81, C+ = 77:79, C = 70:76,

F = 0:69). There is no “D” grade for student grading in 400- and 500-level courses at KU.

<http://app.kutztown.edu/policyregister/policy.aspx?policy=ACA-048>

Projects 80% divided equally among the project assignments.

Exam 20% covering core material, final exam period.

### **Project assignment grading criteria**

Grading criteria will accompany each assignment handout. Please follow them in satisfying all project requirements. Please re-check requirements when you feel ready to turn in an assignment.

**The academic integrity policy is at <http://cs.kutztown.edu/pdfs/AcademicIntegrityPolicy.pdf>**

Your first reading assignment is to read the above policy statement.

You may openly discuss ideas, algorithms, pitfalls, and the use of programming tools.

You may not share code, test drivers or test data except within groups for group projects.

Group projects, when assigned, have documented partitioning of student responsibilities.

There will be 4 project assignments. There will be a 10% per day late penalty for projects that come in after the due date. We will have one final exam worth 20% of the semester grade.

Class attendance is not graded, but I will be teaching using data sources and concepts both inside and outside the scope of the textbook. You are responsible for all material covered in class, including technical information, coding standards and conventions, verbal specification of assignments, and your questions about topics that are not clear to you. Please, there should be no classroom conversations, cell phones, text messaging, eating, sleeping, obscenities, smoking (tobacco or artificial), vaping, listening to music or other disruptions of the class. I will deduct 5% from an assignment for each infraction.

If you have already disclosed a disability to the Disability Services Office (215 Stratton Administration Building) and are seeking accommodations, please feel free to speak with me privately so that I may assist you. If you have an injury sustained during military service including PTSD or TBI, you are also eligible for accommodations under the ADA and should contact the Disability Services Office.

If you have preferred pronouns for yourself, or a name that differs from the MyKU roster, please let me know.

Week	Text chapters	Lecture Topics
1	1,2	Introduction to Zoom for Real Time Virtual Classroom. Scripting in Python.
2		Scripting in Python for data cleaning and formatting. Data file formats. <a href="#">Assn1 out.</a>
3	3	Python example code from previous year(s) and this year. Possible work session.
4	B	Knowledge representation mechanisms for understanding & automated reactions. Possible work session.
5	4.1, 4.3, 4.4	Overview of machine learning and Weka. Finding data. <a href="#">Assn1 due.</a>
6	8	Data transformations within Weka. <a href="#">Classification.</a> <a href="#">Assn2 out.</a>
7		<a href="#">Classification.</a> Possible work session.
8	5.1 – 5.5	Evaluating what has been learned about data relationships. <a href="#">Assn2 due.</a>
9	4.6, model trees	Linear models, model trees, problems with under- and over-fitting. <a href="#">Assn3 out.</a>
10		Possible work session.
11	4.2	Probabilistic methods include Bayesian inference. <a href="#">Assn3 due.</a>
12	4.8	Clustering. <a href="#">Assn4 out.</a>
13		Work session or sessions.
14		Consolidation and review. <a href="#">Assn4 due.</a>
15		<b>Final exam</b>

1. Assignment 1 on data retrieval, cleaning, & formatting using Python may be a pair-programming assignment. Hand out second class of week 3, due end of week 5.
2. Assignment 2 on using Weka with data to extract trees and/or rules, and to evaluate effective of at least two approaches, hand out second class of week 6, due end of week 8.
3. Assignment 3 on using Weka with data to extract linear models, model trees, and possibly other trees, and to evaluate effective of at least three approaches, hand out second class of week 9, due end of week 11.
4. Assignment 4 on using Weka with data to extract Bayesian and cluster models of data, and to evaluate effective of at least two approaches, hand out second class of week 12, due end of week 14.
5. Final exam will take place during final exam week as an on-line, condensed mini-project.

We will be using Zoom for remote attendance during class time. Recorded archives of class sessions will be available. See the following page for instructions. We will go over this in the first class.

Students:

<https://learningtechnologysupport.kutztown.edu/support/solutions/folders/9000185752>