

Stevens Institute of Technology
Khasha Dehnad
kdehnad@stevens.edu
Teaching Assistant
Posted on CANVAS

Course Requirements

Recommended:

- **Familiarity with the principals of statistics and probabilities; for example, completion of MGT 502 (no credit).**
-

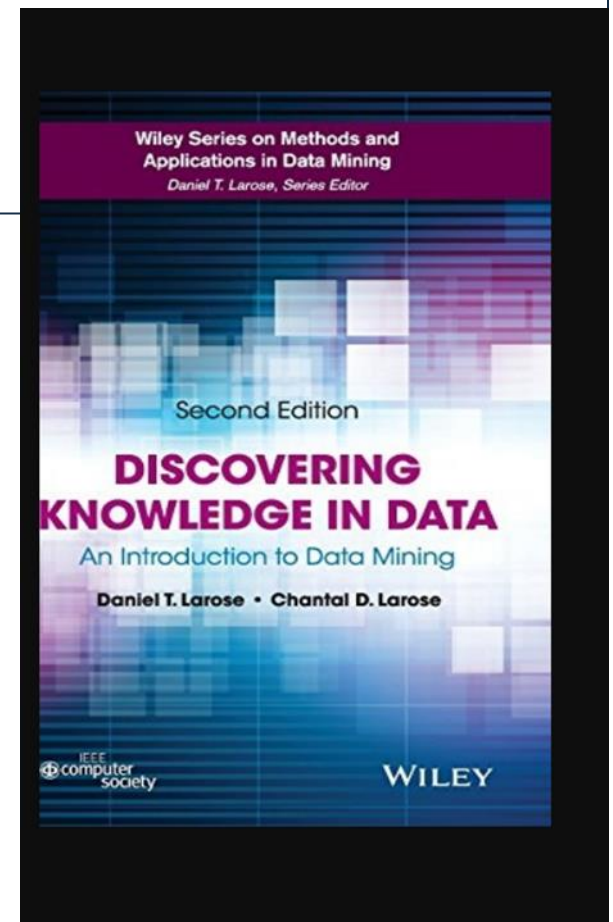
Hardware and Software:

- **Laptop with Spreadsheet**
- **Internet access and ability to install software (admin rights). Students will be installing R and Python (our use COLAB) on their computers**

Course Requirements

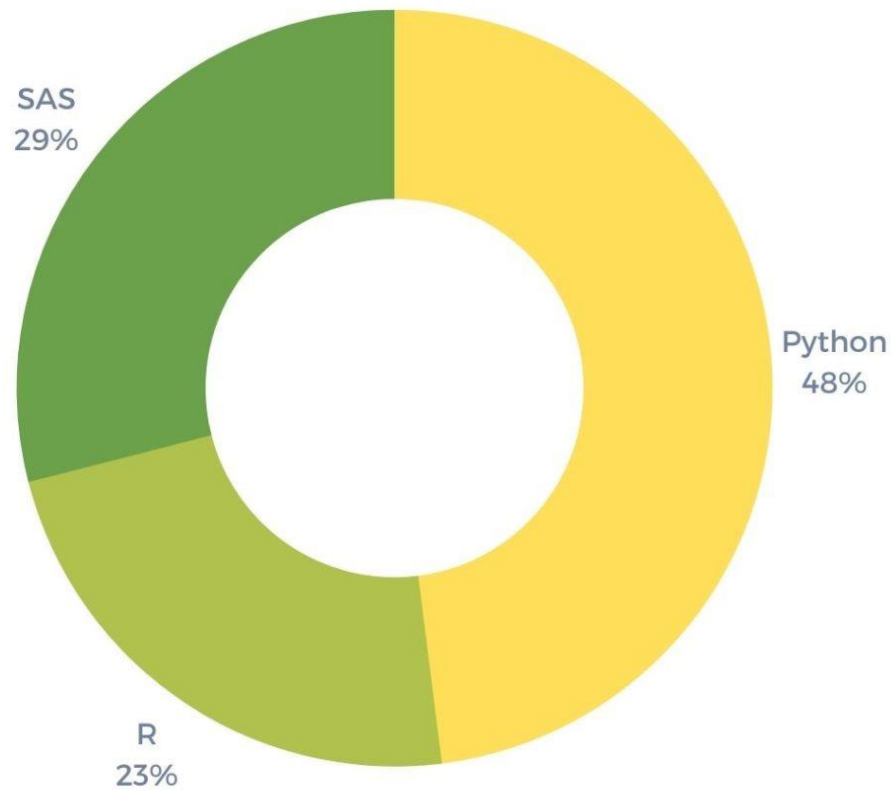
Books, Notes, and Manuals:

- **Discovering Knowledge in Data: An introduction to Data Mining, Daniel T. Larose, John Wiley, latest edition**
- **Lecture Notes and Handouts**
- **Internet Based Papers, Manuals and Documentation**



Course Requirements

SAS, R, or Python 2021 Overall Results



Course Requirements

AI policy (e.g., chatGPT):

- Utilize AI as you would any other external assistance.
- Clearly mark and identify code that has been generated by AI.
- Supply both the AI prompt and the email address used for code generation.

Course Requirements

```
# Course      : Data Mining
# First Name  : Khasha
# Last Name   : Dehnad
# Id          : 12345
# purpose     : apply knn to iris dataset
```

Course Schedule

Housekeeping, Schedule, Intro DM

Week 1

Probability Review

Week 2

Python (review basics)

**DM Lifecycle: Six Phases, Five Case Studies
& Data Preprocessing**

Week 3

**Deriving Rules from Data: ML Algorithms
(Data Preprocessing)**

Week 4

**Data Transformation &
Exploratory data analysis**

Week 5

Course Schedule (Continued)

**k-Nearest Neighbor Algorithm
& Case Study**

Week 6

Naive Bayes classifier

Week 7

**Decision Trees: CART & C4.5 Algorithm
Model Performance Measurements**

Week 8

Random Forest

Artificial Neural Networks (ANN) Deep Learning

Week 9

Hierarchical Clustering

Week 10

**k- Means Clustering Algorithm
& Case Study (Guest speaker)**

Week 11

Special Topics(Intro to Large Language Models)

Week 12

Student Projects and Presentations

Week 13 & 14

Assignments and Grading

Assignments	Grade Percent
Exercises (Best 10 out of 12)	20%
Mid-term	20%
Final	20%
Class Participation	10%
Final project /research paper	30%
Total Grade	100%

Course Grade

Your course grade is based on **your rank in the class**, which is mostly **determined by your project performance and class participation**.

The course is designed to maximize learning and practice through the use of many take home assignments and exams, with plenty of time for answering questions, and you are free to get help from other students and the Tas/CAs. As such, most students should do similarly well on these aspects of the course. Again, the differentiating focus for your grade is placed on project performance and class participation.

Remember, getting **help is different than plagiarism**. **Students that plagiarize will fail the course.**

Project Case Study

Project:

A real-world data mining project (problem statement, data, methodology/algorithm), software, execution and analysis, references, documentation, and presentation). The problem statement, sample data, relevant methodology/algorithm).

Case Study:

A case study from literature/books, prepare and deliver a comprehensive presentation including, problem statement ('profound question'), data source(s), methodology, data mining, result, suggestions for future work, and references.

Project Performance Measurements

- **The novelty of the project idea(s).**
- **Techniques used.**
- **Comparison of the results of the above techniques applied to the data.**
- **Uniqueness of the data source(s). For example, UCI data gets lower ranking**
- **Additional techniques extending those studied in the class**
- **Quality of the presentation material and presentations.**
- **Timing/sequence of the presentation. (Week1 vs Week2)**