

# **Knowledge Discovery in Databases**

**Stevens Institute of Technology**

**Khasha Dehnad**

**kdehnad@stevens.edu**

**Teaching Assistant**

**TBD**

# Course Requirements

## Prerequisites:

- **Familiarity with the principals of statistics and probabilities; for example, completion of MGT 502 (no credit).**
- **Instructor's permission is required for this course.**

## Hardware and Software:

- **Lap top with Excel.**
- **Internet access and ability to install software (admin rights). Students will be installing R on their computers**

## 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 Schedule

**Housekeeping, Schedule, and Probability Review** **Week 1**

**Probability Review** **Week 2**

**Introduction to R  
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)

**Naive Bayes classifier**

**Week 6**

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

**Week 7**

**Decision Trees: CART & C4.5 Algorithm  
& Case Study**

**Week 8**

**Random Forest  
Artificial Neural Networks (ANN)**

**Week 9**

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

**Week 11**

**Special Topics**

**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%
Final project /research paper	40%
<b>Total Grade</b>	<b>100%</b>

# 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.

- **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)**