

# DATA MINING (EC – 559)

Dr. Dhaval Patel  
CSE, IIT-Roorkee

# General Information

- Instructor: Dr. Dhaval Patel
  - Email: [patelfec@iitr.ac.in](mailto:patelfec@iitr.ac.in)
  - Tel: (+91)-1332-285700
  - Office: S209
  
- Course Call Number: EC-559
  
- Lecture times & Room: TBA
  
- Course Website: Moodle/
  
- Office hours: 3:00pm-3:30pm, Tuesday & Thursday (or by appointment)

# Course Objective

- What is this course about?
  - To introduce the **foundational concepts and practical** implications of Data Mining Techniques
  - To survey the state-of-the-art **advancements in theories and applications** of Data Mining
- What you will learn from this course?
  - To effectively carry out further **research** on Data Mining techniques for Big Data Analytics
  - To effectively develop new **applications** based on Data Mining Concept

# Course Structure

- The course has three parts:
  - **Lectures** - Introduction to the main topics + In-class data mining laboratories
  - **Programming projects/Assignments**
    - 4 programming assignments.
    - To be demonstrated to me
  - **Research paper reading/Competition**
    - A list of papers will be given
- Lecture slides will be made available at the course web page maintained at Moodle

# Programming projects

- Four programming projects
- To be done in a group of three students or less
- You will write short description about your assignments and demonstrate your programs to me to show that it works
- You will be given a sample dataset and problem description. Your job is to find solution, implement it and obtain result on given sample dataset.

# Grading

- As per as Guideline
  - ▣ Final Exam: 50%
  - ▣ Midterm: 35%
  - ▣ Programming projects: 15%
    - 4 programming assignments.
  - ▣ Research paper reading (some questions from the papers will appear in the final exam).

# Prerequisites

- Knowledge of
  - basic probability theory
  - algorithms
- Programming Languages
  - Java/C++/XML/...
  - R/Matlab/...
  - ...

# Teaching materials

## □ Text

- ▣ Reading materials will be provided before the class & Expected that student read it before they come for class

## □ Reference texts:

- ▣ Data mining: Concepts and Techniques, by Jiawei Han and Micheline Kamber, Morgan Kaufmann, ISBN 1-55860-489-8.
- ▣ Principles of Data Mining, by David Hand, Heikki Mannila, Padhraic Smyth, The MIT Press, ISBN 0-262-08290-X.
- ▣ Introduction to Data Mining, by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Pearson/Addison Wesley, ISBN 0-321-32136-7.
- ▣ Machine Learning, by Tom M. Mitchell, McGraw-Hill, ISBN 0-07-042807-7

## □ Data mining resource site: [KDnuggets Directory](#)



# Course Outline

- Preliminary
  - Introduction to Data Mining
  - Concept of Probability for Data Miner
  - Data pre-processing
  
- Basic Data Mining
  - Frequent Pattern & Association rule mining
  - Classification (supervised learning)
  - Clustering (unsupervised learning)
  - Post-processing of data mining results
  
- Advance Data mining
  - Time Series Data Mining
  - Social Network Analysis
  - Text Mining

# Any questions and suggestions?

- Your feedback is most welcome!
  - ▣ I need it to adapt the course to your needs.
  
- Share your questions and concerns with the class – very likely others may have the same.
  
- No pain no gain – no magic
  - ▣ The more you put in, the more you get
  - ▣ Your grades are proportional to your efforts.

# Rules and Policies

- **Statute of limitations:** No grading questions or complaints, no matter how justified, will be listened to one week after the item in question has been returned.
- **Cheating:** Cheating will not be tolerated. All work you submitted must be entirely your own.
- **Late assignments:** Late assignments will not, in general, be accepted. They will never be accepted if the student has not made special arrangements with me at least one day before the assignment is due. If a late assignment is accepted it is subject to a reduction in score as a late penalty.

# Course Resources

- IVLE
- Login to : <http://192.168.111.173/moodle>
- Register for Course : Data Mining
  - ▣ Course Syllabus
  - ▣ Lectures Notes
  - ▣ Handouts
  - ▣ Assignments
  - ▣ Projects
  - ▣ Discussion Forum

# Book

- Data Mining: Concepts and Techniques, Third Edition
  - ▣ Jiawei Han, ...
- Principles of Data Mining
  - ▣ David J. Hand, Heikki Mannila and Padhraic Smyth
- Introduction to Data Mining
  - ▣ Pang-Ning Tan, ...