

# CS685: DATA MINING COURSE METADATA

Arnab Bhattacharya  
arnabb@cse.iitk.ac.in

Computer Science and Engineering,  
Indian Institute of Technology, Kanpur  
<http://web.cse.iitk.ac.in/~cs685/>

1<sup>st</sup> semester, 2020-21  
Mon 1030-1200 (online)

# Rules

- No strict pre-requisites except general aptitude
  - Data structure, algorithms, probability, statistics, databases, linear algebra expected
- Email `arnabb@cse.iitk.ac.in`
  - Put “CS685” in the subject so that automatic filter catches it

# Rules

- No strict pre-requisites except general aptitude
  - Data structure, algorithms, probability, statistics, databases, linear algebra expected
- Email `arnabb@cse.iitk.ac.in`
  - Put “CS685” in the subject so that automatic filter catches it
- Participate
  - Attend online classes, if possible
  - Clear doubts
  - Answer questions
- Do assignments *individually*

- No strict pre-requisites except general aptitude
  - Data structure, algorithms, probability, statistics, databases, linear algebra expected
- Email `arnabb@cse.iitk.ac.in`
  - Put “CS685” in the subject so that automatic filter catches it
- Participate
  - Attend online classes, if possible
  - Clear doubts
  - Answer questions
- Do assignments *individually*
- *No* extension of deadlines
  - Since issues can happen online, aim to submit early
- If *you* are unwell, follow standard IITK procedure

# Grading Policy

- Exams: 20-30%
- Assignments: 20-30%
- Project: 50-60%

# Grading Policy

- Exams: 20-30%
- Assignments: 20-30%
- Project: 50-60%
- Things may be changed by mutual consent after discussion in class

# Project Details

- Form your *own* idea
- Just implementation or survey will *not* be enough
  - Back it up with analysis
- Groups of at most 5
- Deadlines
  - 1 Groups for project: Sep 14
  - 2 Initial one paragraph write-up: Sep 21
  - 3 Mid-term report: Oct 19
  - 4 Final report: Nov 23

# Course Material

- Slides
- Book: no text book
- Conference proceedings and journal articles
  - KDD, ICDM, SDM, PKDD, PAKDD, etc.
  - TKDE, KDD, DMKD, etc.



# Course Contents I

- ① What is data mining?
  - Connection to machine learning, statistics, databases
  - What is *not* data mining?
- ② Data pre-processing
  - Data extraction
  - Data cleaning
  - Data transformation
- ③ Data warehousing and data cube
  - Multi-dimensional data model
  - OLAP: on-line analytical processing
- ④ Itemset mining
  - Frequent itemsets
  - Association rule mining
- ⑤ Classification
  - Tree-based classification
  - Bayesian classification

# Course Contents II

- Rule-based classification
- Support vector machines
- Artificial neural networks

## 6 Prediction

- Regression

## 7 Clustering

- Partition-based methods
- Hierarchical methods
- Model-based methods

## 8 Anomaly detection

- Rule-based methods
- Statistical methods

## 9 Mining special kinds of data (if time and interests permit)

- Graph mining
- Text mining
- Image analysis
- Biological data