

DS-600 Data Mining – Spring Semester 2018
Jane Cheng

General Information:

Instruction method: Lecture with laboratory, 2.5 hours per week
Times: 1/18/2017-05/10/2017 - Wednesday 06:00PM - 08:30PM
Location: Pope Hall, Room 212
Estimated preparation time: approx. 5-7 hours per week
Office Hours: upon appointment scheduled by email.

Email: zcheng@saintpeters.edu; janecheng@yahoo.com
Office Phone: 908.268.0083

Course Outline:

The course gives an introduction Data Mining theory and Business usage using latest tools and technology

Lab/Tools/Topics includes:

- 1) SQL review with **PostgreSQL**
- 2) Python usage in data mining and application
- 3) Introduction to AWS cloud computing
- 4) Implement Hadoop using AWS,
- 5) Lab on EC2, S3, EBS, RDS, EWR, AWS Cli
- 6) Introduction to Data Mining task and sample
- 7) Basic Classification technique
- 8) Bayesian Classification in depth
- 9) Clustering technique
- 10) Advanced clustering example using K-means algorithm using python
- 11) Python Libraries (Using NumPy and SciPy)
- 12) Association rule
- 13) Hadoop and Map Reduce

Pre-requisites:

- Introductory undergraduate courses in science or equivalent professional experience
- Computing experience using PC, Mac or Linux computers

Preparation:

- Complete required reading prior to the weekly lecture (see section Tentative Topics below).
- Install software required

Reading Material:

1. Text Book

<http://www.cambridge.org/us/catalogue/catalogue.asp?isbn=9781107077232>
(Chapter 1, 2, 6 , 7,)

reference text book

<http://www.amazon.com/Introduction-Data-Mining-Pang-Ning-Tan/dp/0321321367>

ISBN-13: 978-0321321367

ISBN-10: 0321321367

2. Supplement document Data Mining: Introductory and Advanced Topics

http://www.pearsonhighered.com/pearsonhigheredus/educator/product/products_detail.page?isbn=013088923&forced_logout=forced_logged_out

Margaret H. Dunham

(chapter 3, 4)

3. http://www.tutorialspoint.com/mysql/mysql_tutorial.pdf

4. <http://docs.aws.amazon.com/gettingstarted/latest/awsgsg-intro/gsg-aws-intro.html>

5. <http://aws.amazon.com/>

6. <http://betterexplained.com/articles/an-intuitive-and-short-explanation-of-bayes-theorem/>

7. https://www.amazon.com/Python-Crash-Course-Hands--Project-Based/dp/1593276036/ref=sr_1_1?ie=UTF8&qid=1472777841&sr=8-1&keywords=python+books

Personal computers:

- A computer running OS X and with 8 GB of RAM minimum: see www.saintpeters.edu/data-science for student discount information for a computer pre-installed with OS X, or install OS X on your existing computer (follow instructions on internet - advanced IT expertise required)

University computers:

Lab sessions will be performed on MacBook computers with 16 GB of RAM once the lab installation will be completed.

Attendance Policy:

Attendance is required and will count towards your grade. You are allowed to a maximum of 3 absences. Reasons need to be provided by email to the instructor at most 1 week after the absence. No food in classroom

Learning Objective/Outcomes:

- Learn the latest tools used in software development
- Understand the Big data and concept of data mining
- Using AWS lab to understand the how to store, share and process big data
- Deep illustration of data mining task –Bayesian classification

Outcome Measure:

20% Labs

30% Assignments

15% Participation and Teamwork

35% Final Exam

Tentative Topics:

- Introduction to Data mining , Data mining vs. database Basic data mining tasks,

PostgreSQL /HQL difference (SQL/Python Lab)

- Introduction to Big Data, Introduction to MapReduce and Hadoop, Big Data Querying with Hive (AWS)
- Define data mining and KDD, Data mining development
- Data Classification, Clustering, Clustering, Association Rules

Assignments:

Students will need to install on their personal computers MySql and PHP. All assignments need to be submitted electronically through Blackboard with the result The report needs to include title, name, date and one section heading for each question in the assignment.

Collaboration with Classmates:

- Students are encouraged to discuss with their classmates about the assignments. However, assignment must be produced individually. Students must create their own charts and figures to illustrate their own reports.
- Collaboration between classmates is encouraged and will count towards your teamwork grade.

Grading:

- A Exceptional
- A- Excellent
- B+ Very good
- B Good
- B- Average
- C+ and below Improvement needed

Campus Services:

- IT Support: Information Technology Services (ITS) are located in the basement of Loyola Hall.
<http://www.saintpeters.edu/its/>
- Graduate Scholarships and Fellowships: schedule appointment with Dr. Daniel Murphy,
<http://www.saintpeters.edu/graduate-scholarship-professional-studies/fellowships/>
- Career Services: schedule an appointment with a Saint Peter's Career Officer,
<http://www.saintpeters.edu/internships-careers/>
- OPT/CPT for International Students: Schedule an appointment with Mr. Tushar Trivedi,
<http://www.saintpeters.edu/contact/trivedi-tushar-2/>