

COURSE OUTLINE
CSE6444: Mining Software Engineering Data
Winter 2018

Instructor:

Dr. Zhen Ming (Jack) Jiang

Office Hours: Wednesdays 04:30 - 05:30 pm or by appointment

Email: zmjiang at cse dot yorku dot ca

Location: LAS 1012E

For course related questions, please post to the forum. For personal issues (e.g., grading etc.), please email with the subject "CSE6444 -".

Lectures:

Mondays and Wednesday: 05:30 - 07:00 pm

Location: R S501

Course Description:

Software engineering data (such as source code repositories, execution logs, performance counters, developer mailing lists and bug databases) contains a wealth of information about a project's status and history. Applying data mining techniques on such data, researchers can gain empirically based understanding of software development practices, and practitioners can better manage, maintain and evolve complex software projects.

Course Requirements:

For breadth requirements, this course belongs to the following two groups:

- Group 3: Systems: Software and Hardware, and
- Group 4: Computer Systems Engineering

Students are expected to have some background in programming and software development. Data mining knowledge will be beneficial but not expected.

Topics Covered:

- Automated test analysis
- Bug detections
- Bug predictions
- Building and extracting large software engineering data repositories
- Leveraging big data frameworks to mine software engineering data
- Mining software processes
- Visualizing and understanding large-scale software engineering data

Course Assessment:

This is a research-oriented course. The first part of the course will introduce basic concepts and approaches to mining software engineering data. The second part will

focus on student review and presentation of relevant journal and conference papers. The third part is the preparation and in-class presentation of a publication-quality research project.

- **Class participation (10%)**
 - Students are expected to read all the papers covered in a week, come to class prepared to discuss their thoughts and take part in the discussions.
- **Paper presentation and discussion (20%)**
 - Depending on the size of the class, every student will present (*presenter*) and discuss (*discussant*) at least two papers on a course-related topic. The presentation will last 20 minutes strict and the discussion will last 15 – 20 minutes. Each student should upload the presentation and discussion slides before the class.
 - As a *presenter*, you should not simply repeat the paper's content (remember you only have 20 minutes), instead you should point out the main important findings of the work. You should highlight any novel contributions, any surprises, and other possible applications of the proposed techniques. You should check the authors' other work related to the presented paper. Finally, you should place the work relative other papers covered in the course (especially the papers covered in that particular week).
 - As a *discussant*, you should take an adversarial position by pointing out weak and controversial positions in the paper. You should present a short rebuttal of the paper. You should come prepared with problems and counterexamples for the presented work.
 - Your presentations should have:
 - one slide that lists the main contributions of the paper;
 - one slide that places the paper relative to any recent work done by the authors of the paper;
 - one slide that links places the paper relative to other papers presented that week;
 - one slide (the final slide) that lists of at least three technical points that you liked and three areas that should be improved.
- **Paper critique (10%)**
 - Every week that we have student paper presentations, each student is expected to send a critique for one of the papers that are being presented. The critique should include a brief summary of the paper, at least three strengths of the paper and at least three weaknesses of the paper. You do not need to submit a critique if you are presenting that week.
- **Assignment (20%)**
 - There will be one assignment done in a group of at most 3 students. More details will be provided in class. The assignment will use WEKA or R on some software engineering data.
- **Project (40%)**
 - A large portion of the course deliverables is a course research project. You are expected to work on the course project alone or in pairs. Each group is expected to write a research paper by the end of the semester. The topic is to be discussed with the instructor. Examples include a new contribution on a

particular mining software engineering data topic, a survey paper of a specific topic (typically involves surveying 30 - 60 papers), or building a tool to support some software mining tasks.

- The first deliverable of the course project is a project proposal. The project proposal should be 3 pages in length (IEEE format plus references) and needs to be submitted around 1.5 months before the end of the term. The project proposal should provide a brief motivation of the project, along with a timeline of milestones and expected outcome. Make sure you cite at least 3 papers in your proposal.
- The final report should be 10 pages in length (IEEE format) and be submitted by the end of the term. The final submission is expected to be of publishable quality. If the paper is deemed publishable, the instructor will work with the student to make appropriate changes to the final report and submit the paper for publication.
- You can refer to the following link for some advice on the proposal and report writing:
 - <http://www.cse.yorku.ca/~zmjiang/teaching/cse6444/HPPProjectProposalTemplate.pdf>

Significant Dates:

First Class	January 08, 2018
Reading Week	February 17 - 23, 2018
Family Day	February 19, 2018
Drop Date	March 09, 2018
Good Friday	March 30, 2018
Class Ends	April 06, 2018

Academic Honesty and Cheating:

Cheating, plagiarism and other forms of academic fraud are taken very seriously by the University, the Faculty, and the teaching staff. Consult York's policy on academic honesty: <http://www.cse.yorku.ca/admin/coscOnAcadHonesty.html>