TO: The Faculty of the College of Engineering

FROM: Elmore Family School of Electrical and Computer Engineering

RE: New Graduate Course, ECE 50836 Introduction to Data Mining

The faculty of the School of Electrical and Computer Engineering has approved the following new course. This action is now submitted to the Engineering Faculty with a recommendation for approval.

ECE 50836   Introduction to Data Mining
Sem. 1, Lecture 3, Cr. 3.
Prerequisite: MA 26500, ECE 20875, ECE 30200
Prerequisite by Topic: Linear Algebra, Statistics, and Python programming

Description: This course introduces fundamental techniques in data mining, i.e., the techniques that extract useful knowledge from a large amount of data. Topics include data preprocessing, exploratory data analysis, association rule mining, clustering, classification, anomaly detection, recommendation and graph analysis. Students are expected to gain the skills to formulate data mining problems, solve the problems using data mining techniques and interpret the output.

Reason: This course is an elective course on a specific subject, which will be an essential supplement to the core courses offered in the School of ECE. It will introduce the state-of-the-art data mining techniques. The ability to manage and analyze big data has become a very useful skill for the graduate students who are interested in pursuing industry or research positions after graduation.

Course History: Fall 2022 – 88, Fall 2021 - 30

Milind Kulkarni,
Associate Head for Teaching and Learning
Elmore Family School of Electrical and Computer Engineering

Course Information

- **Course number and title:** ECE 50836 Introduction to Data Mining
- **CRN:** 25678
- **Meeting day(s) and time(s)**: 3:30PM-4:20PM MWF Wang Hall 2599 (on-campus students), asynchronous online learning (online students).
- **Course credit hours:** 3 credits
- **Prerequisites (if any)**: Linear Algebra, Probability and Statistics, and Python Programming.

Instructor(s) Contact Information

- **Name of the instructor:** Jing Gao
- **Office Location:** EE 334A
- **Office Phone Number:** (765)494-7223
- **Purdue Email Address:** jinggao@purdue.edu
- **Student consultation hours, times, and location**:

    - **Platform for QA:** submit a public or private post on piazza.com.
    - **Announcement:** made by a public post on piazza.com.
    - **Office hour:** 7-8PM Friday via Zoom.
    - **Email:** Please include "ECE 595" in the subject line.

Course Description

This course introduces fundamental techniques in data mining, i.e., the techniques that extract useful knowledge from a large amount of data. Topics include data processing, exploratory data analysis, association rule mining, clustering, classification, anomaly detection, recommendation and graph analysis. Students are expected to gain the skills to formulate data mining problems, solve the problems using data mining techniques and interpret the output.

Learning Resources, Technology & Texts

- **Textbooks**
    - **Required:** *Introduction to Data Mining* (2nd Edition). Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar, Pearson.
    - **Recommended:** *Data Mining: Concepts and Techniques* (3rd Edition). Jiawei Han, Micheline Kamber, and Jian Pei, Morgan Kaufmann Publishers.
- **Software requirement:** Python 3
- **Class websites:**
    - **Brightspace:** Lecture videos, quizzes, assignments and grades.
    - **Piazza:** Lecture slides, question answering.

Learning Outcomes

At the end of this course, you'll be able to:

- describe and explain the process of data mining;
- formulate problems in real-world applications into data mining tasks and solve the problems using data mining techniques;
- implement software programs that conduct data mining and evaluate the output of data mining programs;
- present data mining solutions to people in scientific and other disciplines.

Assignments

Your learning will be assessed through a combination of online quizzes, programming assignments, final exam and project spread throughout the semester. Details on these assignments and exam, including a schedule of due dates and rubrics to guide evaluation will be posted on the course website.

- **Quizzes:** Biweekly quizzes will be given to test your knowledge on data mining process, concepts, procedure and techniques. They will be hosted on Brightspace.
- **Programming assignments:** Four programming assignments (roughly one assignment every three weeks) will be given to evaluate your understanding of data mining techniques, and ability of implementing data mining software and evaluating the output of data mining software. Specifically, the assignments include 1) data pre-processing, 2) association rule mining, 3) clustering and 4) classification. Python-based template will be provided.
- **Final exam:** Final exam is to test your understanding of data mining process, concepts, procedure and techniques and your ability of applying data mining to real world.
- **Project:** The project is to evaluate your ability to formulate real world problems into data mining tasks, select and apply appropriate data mining algorithms, and present solutions.

Grading Scale

In this class, grades reflect the sum of your achievement throughout the semester. You will accumulate points as described in the four aforementioned categories, with each assignment graded according to a rubric. At the end of the semester, a final score will be calculated by adding the points earned in each category, normalizing these points into [0,100] and conducting a weighted combination of the points as follows:

- **Quizzes:** 5%
- **Programming assignments:** 40%
- **Final exam:** 40%
- **Project:** 15%

The final score (out of 100) will be translated into the following letters (there will be no partial points or rounding):

- **A+:** [98,100]
- **A**: [94,98)
- **A-**: [90,94)
- **B+**: [88,90)
- **B**: [84,88)
- **B-**: [80,84)
- **C+**: [78,80)
- **C**: [74,78)
- **C-**: [70,73)
- **D+**: [68,70)
- **D**: [64,68)
- **D-**: [60,64)
- **F**: [0,60)

Attendance Policy

This course follows Purdue's academic regulations regarding attendance, which states that students are expected to be present for every meeting of the classes in which they are enrolled. When conflicts or absences can be anticipated, such as for many University-sponsored activities and religious observations, the student should inform the instructor of the situation as far in advance as possible. For unanticipated or emergency absences when advance notification to the instructor is not possible, the student should contact the instructor as soon as possible. Under academic regulations, excused absences may be granted for cases of grief/bereavement, military service, jury duty, and parenting leave. For details, see the Academic Regulations & Student Conduct section of the University Catalog website.

Guidance on class attendance related to COVID-19 are outlined in the Protect Purdue Pledge on the Protect Purdue website.

Academic Guidance in the Event a Student is Quarantined/Isolated

If you must miss class at any point in time during the semester, please reach out to me via Purdue email so that we can communicate about how you can maintain your academic progress. If you find yourself too sick to progress in the course, notify me via email. We will make arrangements based on your particular situation.

Course Schedule

A tentative schedule is as follows.

Number of weeks	Content
0.5	Background and Introduction
1.5	Data—Type of data, data quality, data preprocessing, measure of similarity and dissimilarity, data exploration and visualization
2.5	Association analysis—Frequent itemset generation, rule generation, compact representation of frequent itemsets and evaluation
1	Association analysis on special data types—Relational data, sequences and graphs
1.5	Clustering—K-means, hierarchical clustering, spectral clustering and density-based clustering
1	Clustering on special data types—Text clustering, subspace clustering and clustering spatial-temporal data
2	Classification—Decision tree, rule-based classifier, nearest-neighbor classifier, support vector machines, and ensemble methods
1	Classification on special data types--Classification in a network setting and sequence labeling
1	Anomaly detection—Statistical, distance-based, density-based and clustering-based approaches
1	Recommendation--Collaborative filtering, matrix factorization and applications
1	Graph analysis--Node ranking, link prediction and graph embedding

Classroom Guidance Regarding Protect Purdue

Any student who has substantial reason to believe that another person is threatening the safety of others by not complying with Protect Purdue protocols is encouraged to report the behavior to and discuss the next steps with their instructor. Students also have the option of reporting the behavior to the Office of the Student Rights and Responsibilities. See also Purdue University Bill of Student Rights and the Violent Behavior Policy under University Resources in Brightspace.

Academic Integrity

Incidents of academic misconduct in this course will be addressed by the course instructor and referred to the Office of Student Rights and Responsibilities (OSRR) for review at the university level. Any violation of course policies as it relates to academic integrity will result minimally in a failing or zero grade for that particular assignment, and at the instructor's discretion may result in a failing grade for the course. In addition, all incidents of academic misconduct will be forwarded to OSRR, where university penalties, including removal from the university, may be considered.

Academic integrity is one of the highest values that Purdue University holds. Individuals are encouraged to alert university officials to potential breaches of this value by either emailing integrity@purdue.edu or by calling 765-494-8778. While information may be submitted anonymously, the more information is submitted the greater the opportunity for the university to investigate the concern. More details are available on our course Brightspace table of contents, under University Policies.

Nondiscrimination Statement

Purdue University is committed to maintaining a community that recognizes and values the inherent worth and dignity of every person; fosters tolerance, sensitivity, understanding, and mutual respect among its members; and encourages

each individual to strive to reach his or her potential. In pursuit of its goal of academic excellence, the University seeks to develop and nurture diversity. The University believes that diversity among its many members strengthens the institution, stimulates creativity, promotes the exchange of ideas, and enriches campus life. A hyperlink to Purdue's full Nondiscrimination Policy Statement is included in our course Brightspace under University Policies.

Accessibility

Purdue University is committed to making learning experiences accessible. If you anticipate or experience physical or academic barriers based on disability, you are welcome to let me know so that we can discuss options. You are also encouraged to contact the Disability Resource Center at: drc@purdue.edu or by phone: 765-494-1247.

Mental Health/Wellness Statement

If you find yourself beginning to feel some stress, anxiety and/or feeling slightly overwhelmed, try WellTrack**.** Sign in and find information and tools at your fingertips, available to you at any time.

If you need support and information about options and resources, please contact or see the Office of the Dean of Students. Call 765-494-1747. Hours of operation are M-F, 8 am- 5 pm.

If you find yourself struggling to find a healthy balance between academics, social life, stress, etc., sign up for free one-on-one virtual or in-person sessions with a Purdue Wellness Coach at RecWell. Student coaches can help you navigate through barriers and challenges toward your goals throughout the semester. Sign up is completely free and can be done on BoilerConnect. If you have any questions, please contact Purdue Wellness at evans240@purdue.edu.

If you're struggling and need mental health services: Purdue University is committed to advancing the mental health and well-being of its students. If you or someone you know is feeling overwhelmed, depressed, and/or in need of mental health support, services are available. For help, such individuals should contact Counseling and Psychological Services (CAPS) at 765-494-6995 during and after hours, on weekends and holidays, or by going to the CAPS office on the second floor of the Purdue University Student Health Center (PUSH) during business hours.

CAPS also offers resources specific to COVID-19 on its website. Topics range from "Adjusting to the New Normal" to "How to Talk with Professors about Personal Matters."

Basic Needs Security

Any student who faces challenges securing their food or housing and believes this may affect their performance in the course is urged to contact the Dean of Students for support. There is no appointment needed and Student Support Services is available to serve students 8 a.m.-5 p.m. Monday through Friday. Considering the significant disruptions caused by the current global crisis as it relates to COVID-19, students may submit requests for emergency assistance from the Critical Need Fund

Emergency Preparation

In the event of a major campus emergency, course requirements, deadlines and grading percentages are subject to changes that may be necessitated by a revised semester calendar or other circumstances beyond the instructor's control. Relevant changes to this course will be posted onto the course website or can be obtained by contacting the instructor. You are expected to read your @purdue.edu email on a frequent basis.

Disclaimer: This syllabus is subject to change. You will be notified of any changes as far in advance as possible via an announcement on Piazza. Monitor your Purdue email daily for updates.