ILLINOIS TECH
College of Computing

ITMD 522 Syllabus

Data Mining & Machine Learning

Department of Information Technology and Management

Fall, 2022

Professor: **Yong Zheng**

Address: Perlstein Hall, Room 221; 10 W 33rd St, Chicago IL 60616
Telephone: 312 567 3575; *Email:* yzheng66@iit.edu
Office(s): Main Campus - Perlstein Hall, Room 221 D
Office Hours: Fridays 1:00 – 3:00 PM or google meet/zoom by appointment

Note: This syllabus is applicable to the following sections

- ITMD 522-01, IT-D 870-01; Live Section
- ITMD 522-02, IT-D 870-02; Online Section
- ITMD 522-03; Remote Students from India
- ITMD 522-04; Remote Students from China

Course Catalog Description: Data mining is a useful tool to uncover patterns and underlying relationships in large data by using data analytics and knowledge discovery techniques. Machine learning algorithms additionally learn from the data and make predictions or decisions by different optimization methods. This course is a graduate level survey of concepts, principles and techniques related to data mining and machine learning. Students will be familiar with data preprocessing skills and the popular data mining and machine learning techniques, including the supervised learning (regressions and classification) and unsupervised learning(clustering and association rules analysis), as well as semi-supervised learning and ensemble learning. Students will learn Python programming for data mining and machine learning and be able to handle real-world data or applications.. **Prerequisites or Co-requisites:** ITMD/ITMS 514. **Credit:** 3-0-3

Course Outcome: At the completion of the course, each student will have the capability to deal with real-world data mining problems by using standard knowledge discovery in databases (KDD) process. More specifically, students are able to perform data selection, data preprocessing and data mining on real-world data sets. Students will have demonstrated their knowledge and skills in classical data mining techniques, including classification, clustering, association rule analysis, etc. Students are able to use related tools (Python) to apply these data mining and machine learning techniques.

Lecture Days, Time & Place:

Time: Tuesdays & Thursdays, 11:25 – 12:40 PM; Place: SB-107

Important Dates:

August 22	Fall Courses Begin
August 27	Last Day to Add/Drop for ID A Session Courses with No Tuition Charges
September 2	Last Day to Add/Drop for ID Full Semester Courses with No Tuition Charges
September 3	Last Day to Add/Drop for Full Semester Courses with No Tuition Charges
September 5	Labor Day—No Classes
September 9	Last Day to Request Late Registration
September 12	Fall Degree Conferral Applications Due
September 19	Last Day to Withdraw for ID A Session Courses
October 3	Spring and Summer Incomplete Grades Due
October 10	Fall Break Day—No Classes

October 12	Midterm Grading Begins
October 17	ID B Session Courses Begin
October 19	Fall Final Exam Schedule Published Online
October 21	Midterm Grades Due
October 24	Spring and Summer Course Schedules Published Online
October 24	Last Day to Add/Drop for ID B Session Courses with No Tuition Charges
October 30	Last Day to Withdraw for ID Full Semester Courses
October 31	Last Day to Withdraw for Full Semester Courses
November 7	Spring and Summer Registration Begins
November 15	Spring Reinstatement Applications Due for Undergraduate Students
November 18	Last Day to Withdraw for ID B Session Courses
November 23–27	Thanksgiving Break—No Classes
December 3	Last Day of Fall Courses
December 4	Last Day to Request an Incomplete Grade
December 5–10	Final Exam Week/Final Grading Begins on Dec 5
December 14	Final Grades Due at Noon (CST)

Schedule of Topics: *Updates to the schedule, readings and assignments will be posted to Blackboard on a weekly basis and announced in class. The schedule below is tentative and subject to change.*

Session	Topic
1	Intro: Data Science & AWS Summit
2	Data Preprocessing, Python
3	Classification by KNN
4	Classification by Naïve Bayes
5	Classification by Decision Trees
6	Classification by Logistic Regression & SVM
7	Ensemble Methods
8	Multi-Label Classifications
9	Classifications: Advanced Techniques & Python Practice
10	Clustering Techniques
11	Clustering Techniques
12	Feature Selection and Reductions
13	Exam, Association Rules
14	Outlier Detection & Thanks' Giving
15	Advanced Topics: Neural Networks
16	Final Project Presentation

Precautions: Visit iit.edu/COVID-19 for details on Illinois Tech's response to coronavirus (COVID-19). For information from government authorities, please see the Centers for Disease Control and Prevention website at cdc.gov.

Textbook: The textbook for this course is NOT mandatory, but highly suggested and recommended.

- Peter Flach. "Machine Learning: The Art and Science of Algorithms that Make Sense of Data", Cambridge University Press; 1 edition; ISBN-10: 1107422221, ISBN-13: 978-1107422223
- Jake VanderPlas. "Python Data Science Handbook: Essential Tools for Working with Data 1st Edition", O'Reilly Media; 1 edition; ISBN-10: 1491912057, ISBN-13: 978-1491912058

Online Resources: A list of helpful online resources can be found below.

- KDnuggets, Data Mining Community's Top Resource for Data Mining and Analytics info. http://www.kdnuggets.com/
- Ron Zacharski. A Programmer's Guide to Data Mining, http://guidetodatamining.com/
- Kaggle, A Data Science Community, https://www.kaggle.com/

Readings/Videos: Readings for the class will be assigned from the textbook as well as in the form of online reading. Online resources and videos will be linked from or embedded in a Blackboard page. It is essential that you do all readings and/or view the videos before coming to class on the assigned date. These materials are a necessary and integral part of the class and will form the basis for any class discussions on the topic. Specific readings are assigned by topic above.

ILLINOIS TECH
College of Computing

ITMD 522 Syllabus

Data Mining & Machine Learning

Department of Information Technology and Management — Fall, 2022

Course Objectives: At the conclusion of this course, each successful student will able to:

- Understand what is data mining and how useful it is
- Understand and describe real-world data mining techniques and applications
- Understand and distinguish related terms: artificial intelligence, data mining, machine learning, etc
- Understand and be familiar with standard KDD process
- Perform popular data preprocessing: data selection, data cleaning, data transformation, etc
- Understand the difference between classification and clustering
- Perform and evaluate different classification algorithms over real-world data sets
- Perform and evaluate different clustering algorithms over real-world data sets
- Understand how association rule analysis works and how it can be used to assist business intelligence
- Perform and evaluate associate rule mining to discover useful patterns over real-world data sets
- Be able to use data mining tools (e.g., Python) to perform data mining tasks
- Understand the basic evaluation protocols, challenges (e.g., overfitting, imbalance issues) and corresponding solutions in different data mining tasks
- Understand the pros and cons of each techniques for specific data mining tasks. For example, among different classification techniques, what are their pros and cons
- Students will learn many techniques, and finally they must be very clear about which techniques should be used in which situations for which specific problems.
- They must have the capability to identify problems and figure our multiple appropriate techniques to solve the problems.

Course Notes: Copies of the course lecture notes in the form of a PDF of the PowerPoint presentation accompanying each lecture will be provided for each student on Blackboard. This should be useful if you must miss a class. You should be aware that note taking is encouraged and should help your understanding of the material.

Course Web Site: http://blackboard.iit.edu/

Blackboard: The course will make intensive use of Blackboard (http://blackboard.iit.edu/) for communications, assignment submissions, group project coordination, providing online resources and administering examinations. All remote students will view the course lectures online via Blackboard, and online readings will be found on Blackboard.

Guest Lectures: Guest lecturers may be featured as part of course topics. When a guest speaker is expected you should make an extra effort to be seated and ready prior to class time. Guest lectures may be in the evening in which case class will not be held during a scheduled morning period. A question & answer/ discussion period will be held at the end of each lecturer's presentation.

Attendance: If you are in a live section of the class and will not be able to attend class, please notify me via email prior to class time. Post-notifications will not be considered. Online students are welcome to attend the live sections if there are available seats. Roster sheets will be randomly distributed to the class, and students in the live sections are responsible to sign. For online sections, the online roster sheets will be randomly distributed on Tuesdays or Wednesdays, and online students should sign by Friday, 11:59 PM on that week.

Assignments: There will be three types of assignments: 1). *Written assignments* which examine students' knowledge and skills in related topics, such as concepts/problems/challenges/solutions/algorithms/ evaluations/Python programming, etc; 2). *Paper reviews* which extend students' vision in addition to classical data mining techniques; 3). *Final individual or team projects*. No late submissions are allowed for final projects. Note that, all assignments except the final project are individual assignments, you cannot work together and submit the same or similar answers.

Late submissions: 1). For late submissions, a 15% penalty based on the grades they earned will be applied. For example, the total score is 100 for an assignment, and one student finally got a score of 90. But his submission was late and passed the due date. Just apply a 15% penalty on 90. His or her final score will

be 90 * 0.85 = 76.5. 2). The submission will be discarded if it was submitted later than a week. For example, due date is Sep 1st, 11:59 PM. Students can submit late work, but their submission must be priori to Sep 8th, 11:59 PM. Otherwise, he or she will get a ZERO score. Special note: students are responsible for their submissions. It will be considered as late submission, if a student re-submitted their work after the due date and time, even if the student claimed that they submitted the wrong documents. Therefore, students should double-check their submitted documents by themselves.

Blog and Quizzes: Each student will have access to the discussion board on the Blackboard system. Students are encouraged to post questions and discussions in the forums. Students can help each other in the following topics: understanding assignments, discussing challenges in assignments, questions about lectures or learnings, sharing learning materials and job positions, helping debugging, and so forth. Students are NOT allowed to share answers in assignments or exams.

I may give quizzes at my discretion and may use them for verification that you have completed assigned course readings or have read the blog entries. Quizzes may be online via Blackboard. As they are discretionary, the weight of quizzes in grading is also left to my discretion and will be included in your class participation grade. If I see a regular pattern of comments on other student's blog entries, I will not need to give any quizzes in this area.

Examinations: At least one exam will be offered in the class. The Final will be a team project.

Academic Honesty:

Plagiarism: All work you submit in this course **must be your own**. You must fully attribute **all** material directly quoted in papers and you must document all sources used in the preparation of the paper using complete, APA-style bibliographic entries. Including directly quoted material in an assignment without attribution is always plagiarism and will always be treated as such by me. No more than thirty-three percent of material included in any paper may be direct quotes.

For any concept questions in the assignments, such as "what is classification?. You can learn by searching answers from Internet, but you cannot simply copy the original texts online in your assignments. You should use your own language/texts as answers based on your understandings.

I will process all written, coding and project submissions using automated utilities to check for potential violations in assignments, including programming projects. Students who have copied codes, texts, or answers from other students or the internet without attribution in their assignments, exams or final projects are considered as plagiarism.

For the first time of plagiarism, the assignment or exam will be graded as a zero score. Students' behaviors will be reported to the department. The student will receive a warning by the instructor and department. If the plagiarism happens more than one time by a specific student, an Academic Honesty Violation Report (AHVR) will be filed, and the student will automatically receive a final grade of E for this course as per the IIT and ITM academic honesty policies. **There is no excuse for not understanding this policy** and if you do not understand it please let me know and I will be happy to discuss it with you until you do.

Grading: Grading criteria for this course will be as follows:

- **A** *Outstanding work reflecting substantial effort* 90-100%
- **B** *Adequate work fully meeting that expected of a graduate student* 75-89.99%
- **C** *Satisfactory work meeting minimum expectations* 60-74.99%
- **F** *Unsatisfactory work* 0-59.99%

The final grade for the class will be calculated as follows:

Regular Assignments	**28%**
Exam	**30%**
Final Project Presentations	**40%**
Class Attendance	**2%**

Class Participation: See the Section of "Attendance"

Graduate and Undergraduate Students: Parts of the assignments and the questions in the exams may be different for the graduate and undergraduate students in the class. The degree of difficulties will be lowered for the undergraduate students.

Other Class Resources: Online readings and other class resources may be found at on Blackboard.

Our Contract: This syllabus is my contract with you as to what I will deliver and what I expect from you. If I change the syllabus, I will issue a revised version of the syllabus; the latest version will always be available on Blackboard. Revisions to readings and assignments will be communicated via Blackboard.

Disabilities: Reasonable accommodations will be made for students with documented disabilities. In order to receive accommodations, students must obtain a letter of accommodation from the Center for Disability Resources and make an appointment to speak with me as soon as possible. My office hours are listed on the first page of the syllabus. The Center for Disability Resources (CDR) is located in 3424 S. State St., room 1C3-2 (on the first floor), telephone 312.567.5744 or disabilities@iit.edu.

Illinois Tech's Sexual Harassment and Discrimination Information: Illinois Tech prohibits all sexual harassment, sexual misconduct, and gender discrimination by any member of our community. This includes harassment among students, staff, or faculty. Sexual harassment of a student by a faculty member or sexual harassment of an employee by a supervisor is particularly serious. Such conduct may easily create an intimidating, hostile, or offensive environment.

Illinois Tech encourages anyone experiencing sexual harassment or sexual misconduct to speak with the Office of Title IX Compliance for information on support options and the resolution process.

You can report sexual harassment electronically at iit.edu/incidentreport, which may be completed anonymously. You may additionally report by contacting the Title IX Coordinator, Virginia Foster at foster@iit.edu or the Deputy Title IX Coordinator at eespeland@iit.edu.

For confidential support, you may reach Illinois Tech's Confidential Advisor at (773) 907-1062. You can also contact a licensed practitioner in Illinois Tech's Student Health and Wellness Center at student.health@iit.edu or (312)567-7550

For a comprehensive list of resources regarding counseling services, medical assistance, legal assistance and visa and immigration services, you can visit the Office of Title IX Compliance website at https://www.iit.edu/title-ix/resources.