|  |  |
| --- | --- |
|  | **CIT 52600 Applied Data Analytics/42100 Big Data Analytics** |
|  | Fall 2022 (Aug 22, 2022– Dec 18, 2022) |

**Instructor Information**

Instructor: Dr. Xiao Luo

E-mail: [luo25@iupui.edu](mailto:luo25@iupui.edu)

Class Time: Mondays, Wednesdays 1:30-2:45pm

Class Location: ET317

Office Hours: Wednesdays 3:00-4:30pm (Or by appointment to meet online)

Office Location: ET309D

## Course Description

This course will cover both the fundamentals and concepts of the data analytics and the advanced data analytics techniques. The focus is on emerging advanced data analytics techniques and their applications to practical problems for different disciplines, such as IT, health care, and economics. Both machine learning algorithms and distributed computing environment, such as Hadoop will be explored. Students will apply these advanced techniques in labs and a project to resolve an applied problem and identify scientific findings by using public data sets. A project report is required at the end of the course.

## ****Course Outcomes****

* Demonstrate the solid understanding of the fundamentals of data analytics concepts, principles and life cycle [c]
* Apply the data analytics cycles for effective problem solving [b]
* Design state-of-the art data analytics techniques to apply to practical work [c]
* Explain the Hadoop architecture [m]
* Develop and test a MapReduce application [i]

## PREREQUISITES:

# Students are required to have previous knowledge of data structures and programming experience.

Programming skill: Some homework and projects are in JAVA, Python or Weka. A brief introduction of Python and Weka will be given in class. You are expected to learn the languages and Weka on your own.

## TEXT(S):

# Recommended

Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data

EMC Education Services (Editor)

ISBN: 978-1-118-87613-8, 2015

Hadoop the Definitive Guide

By Tom White

ISBN: 978-1-491-90163-2

# 

## Course Evaluation

Student achievement will be assessed, and grades will be given according to class participation, Lab projects and reports, database security project, mid-term test and final project and presentation. Grades will be determined upon the following distribution:

|  |  |
| --- | --- |
| Class participation and In-class discussion: | 5% |
| Assignments (total 8, 5% each) | 40% |
| Project and presentation | 35% |
| Final exam | 20% |

Letter grades will be based on the following grading scale:

93 - 100% = A 90 - 92% = A- 87 - 89% = B+ 83 - 86% = B 80 - 82% = B-

75 - 79% = C+ 70 - 75% = C

< 70% = F ‘FN’, ‘FNN’, ‘W’, ‘I’ grades shall be granted per school policy

## TENTATIVE WEEKLY SCHEDULE (subjected to change)

Week 1: Course introduction and Overview of big data and data analytics

Week 2: Data Analytics Lifecycle, Overview of Data Analytics Methodologies

Week 3: Intro to Data Analytics Packages in Python, JAVA and Weka, Data Preprocessing

Week 4: Feature Selection, Data Visualization

Week 5: Supervised Learning Algorithms 1

Week 6: Supervised Learning Algorithms 2

Week 7: Unsupervised Learning Algorithms

Week 8: Association Rule Mining

Week 9: Natural Language Processing and Text Mining

Week 10: Hadoop

Week 11: HDFS Hadoop Map Reduce

Week 12: NoSQL Overview

Week 13: HBase

Week 14: Project presentations

Week 15: Final Exam (Dec 14, 2022)

## EXPECTATIONS, GUIDELINES, AND POLICIES

# Attendance:

Class attendance is required for zoom-based courses. It entails being present and attentive for the entire class period. The instructor is required to submit to the Registrar a record of student attendance, and action shall be taken if the record conveys a trend of absenteeism. Illness or a death in the immediate family is usually the only acceptable excuse for absence from class. Absences must be explained to the satisfaction of the instructor, who will decide whether omitted work may be made up. To protect your privacy, doctor’s excuses should exclude the nature of the condition and focus instead on how the condition effects on your coursework.

Missing class reduces your grade through the following grade reduction policy: 1% will be taken off for each unexcused absence. If you have 5 unexcused absence, you will loss all your 5% participation score, and you will be given an ‘F’. Generally, a doctor’s note is required for an absence.

# Deliverables:

You are responsible for completing each deliverable (e.g., assignment, quiz) by its deadline and submitting it by the specified method. Deadlines are given in class or in supplementary documents accessible through Canvas. Should you miss a class, you are still responsible for completing the deliverable and for finding out what was covered in class, including any new or modified deliverable. In fairness to the instructor and students who completed their work on time, a grade on a deliverable shall be reduced 10% if it is submitted late but within 24-hour period. No submission will be accepted after the 24 hours passed the deadline.

## project and presentation:

The project can be a group project of 2 people or individual. The requirements are below.

Please submit the project proposal and data set to use by Oct 14th, please include the members in your group in the proposal. The proposal needs to include an introduction of your project, your hypothesis and an overview of the data set you would use to validate your hypothesis. If you have any questions about choosing an interesting topic, please book an appointment with instructor or come to see the instructor during the office hour, so that we can have a discussion on the topics and finalize it for you.

You are responsible to finish the tasks that are associated with the project. The breakdown of the tasks and percentage towards the project grade is as following:

|  |  |
| --- | --- |
| Project Proposal | 5% |
| Implementation and analysis | 15% |
| Final Report and presentation | 15% |

The presentations will be scheduled on Dec 5 and Dec 7, 2022.

The report and code are due at 11:59pm on Dec 12, 2022. You are required to submit PDF version of your report. The report needs to follow the format of general research paper which has all the basic sections: Title, author(s), abstract, introductions, a short literature summary of the topic, system design, specific methodologies, use cases, conclusion and references.

**Tips:**

Make sure that the scope of the topic that you choose is reasonable. Do not choose a topic that is too broad or too narrow. Try to find some materials that are most interesting for you to read or related to your work experiences

Proofread your report thoroughly and use a spell checker. The writing does affect your score.

## Grade Standards

|  |  |
| --- | --- |
| A | Represents the highest grade possible and indicates outstanding achievement.  This grade is not automatically given to the top student performance but instead indicates student work which demonstrates complete mastery of course learning objectives or evinces a level of creativity or originality which far exceeds course expectations.  The grade indicates the student works independently and with strong initiative, seeking knowledge outside the normal framework of the course. |
| B | Represents achievement considerably above expectations.  Student performance demonstrates thorough understanding of course learning objectives and a high level of creativity or originality. |
| C | Student performance meets designated course requirements and demonstrates understanding of the course material and attainment of the course learning objectives. This is the grade that may be expected of a student who puts forth a reasonable amount of time and effort and completes all requirements. |
|  |  |
| F | This grade indicates serious deficiency in understanding course learning objectives and failure to complete requirements of the course. |

## Withdrawing from This Class Using eDrop (Fall 2022)

To find withdrawal deadlines for other session lengths, see the calendar at <https://facultystaffcentral.iupui.edu/calendars/official-calendar.html> and select your session of interest and the “Register, drop/add, refunds” category from the dropdown menus under “Choose your view”.)

If you choose to withdraw from this class, here are important dates and information:

* Aug 29 – Oct 23 Withdrawal with automatic grade of “W.” Requires advisor approval.
* Oct 24 – Dec 8 Withdrawal with a grade of “W” or “F” begins on Oct 25 and extends through Dec 9. Requires advisor, instructor, and dean’s approvals. **Beginning on this date (Oct 24), drops will be approved only if you have serious, extenuating circumstances.** **You must provide documentation (or evidence) of the serious, extenuating circumstance to the Associate Dean for Academic Affairs and Undergraduate Programs in ET 215. However, if your advisor is in the New Student Academic Advising Center in ET 101, you must see your advisor to initiate an eDrop. You must provide documentation (or evidence) of the serious, extenuating circumstances to your advisor who will forward it to the Associate Dean.**

**NOTE**: After you submit the electronic eDrop request, others must approve it. ***You are not withdrawn until you receive confirmation your eDrop has been approved and processed.*** You should attend class while you wait for the approvals and remind your instructor to approve your electronic eDrop. If all approvals are not completed in 14 days, the Office of the Registrar cancels your eDrop request; if this happens, you are still officially enrolled in the class.

## LIST OF SCHOOL/CAMPUS RESOURCES FOR TUTORING such as the TCM Writing Center, Math Labs, etc

The School of Engineering & Technology supports the Technical Communication (TCM) Writing Center (ET 199D, 1st floor of the ET building diagonally across the hall from ET Student Services). You may schedule individual tutoring on campus or online through the Writing Center for writing or speaking assignments. Hours vary by semester. Walk-ins are taken (on-campus only) on a first-come/first-served basis. For more information or to schedule an appointment, go to <https://et.iupui.edu/students/tcmwriting/.>

Another excellent resource for communication skills is The Purdue Online Writing Lab (OWL). <http://owl.english.purdue.edu/>

The Math Assistance Center (<https://science.iupui.edu/math/academics/academic-support/mac.html>) provides support for most 100- and 200-level math courses as well as STAT 35000. The Bepko Learning Center (<https://blc.iupui.edul>) is available to help student connect with academic mentoring, tutoring, and success coaching opportunities.

## STATEMENT ABOUT CIVILITY IN THE CLASSROOM

IUPUI Everyone deserves to be treated with dignity and respect. Within this classroom, civility requires mutual respect for all class members and instructor(s) and their knowledge and expertise. All members of this class (instructor and students) are responsible for and expected to promote respectful and courteous language, demeanor, and actions. Behavior that is offensive, disruptive, intimidating, dismissive, or condescending **will not be tolerated** from either the instructor or the student.

## IUPUI Policy on Academic Integrity

The IU Code of Student Rights, Responsibilities, and Conduct states that students must uphold and maintain academic and professional honesty and integrity; the code defines academic misconduct as any activity that tends to undermine the academic integrity of the institution. Students engaging in academic misconduct may therefore receive penalties from their course instructor and disciplinary action from the university. Policies against academic misconduct apply to all course-, department-, school-, and university-related activities. Academic misconduct may involve human, hard-copy, or electronic resources and includes but is not limited to the following: cheating, fabrication, plagiarism, interference, violation of course rules, and facilitating academic dishonesty. For definitions of these activities, visit the [Definitions](https://studentcode.iu.edu/appendices/definitions.html) appendix on the Student Code website For information on how faculty and students are expected to handle cases involving academic misconduct, visit [Academic Misconduct](https://studentcentral.iupui.edu/grades-progress/academic-misconduct.html) on the Student Code website. Additional information about the rights and responsibilities of IU students is available in the [Code of Student Rights, Responsibilities, & Conduct](https://studentcode.iu.edu/).

## IUPUI Policy on Sexual Misconduct

What you should know about sexual misconduct: Title IX and IU’s Sexual Misconduct Policy prohibit sexual misconduct in any form, including sexual harassment, sexual assault, stalking, and dating and domestic violence. If you have experienced sexual misconduct, or know someone who has, the university can help.

If you are seeking help and would like to speak to someone confidentially, you can make an appointment with:

* Counseling & Psychological Services (CAPS) at 317-274-2548 (counseling services)
* Confidential Advocacy Resources at 317-274-5715 or [saadv@iupui.edu](mailto:saadv@iupui.edu)
* IUPUI Student Health Center at 317-274-2274 (University Blvd. location) or 317-274-8214 (West Michigan St. location) (health and medical services)

It is also important that you know that Title IX and university policy require faculty to share any information brought to them about potential sexual misconduct with the campus Deputy Title IX Coordinator(s) or IU’s Title IX Coordinator to ensure that appropriate measures are taken and resources are made available. Protecting student privacy is of utmost concern, and information will only be shared with those that need to know to ensure the university can respond and assist.

Find more information about sexual violence, including campus and community resources on the [IU's Stop Sexual Violence website](https://stopsexualviolence.iu.edu/index.html).

Other resources:

* [Stop Sexual Violence: Resources for Employees](https://stopsexualviolence.iu.edu/employee/index.html)
* [Stop Sexual Violence:  Frequently Asked Questions](https://stopsexualviolence.iu.edu/employee/faq.html)

## IUPUI Non-Discrimination Policy

Indiana University prohibits discrimination on the basis of age, color, disability, ethnicity, sex, gender identity, gender expression, genetic information, marital status, national origin, race, religion, sexual orientation, or veteran status. If you feel like you have been discriminated against, please contact IUPUI’s Office of Equal Opportunity (OEO). OEO is located in Lockefield Village, LV 4443. You can reach the office by calling 317-274-2306.

IUPUI does not tolerate acts that are damaging to our safe, civil, and inclusive community–and neither should you. If you experience or witness an incident of bias, you should report it.  For more information, see [Student Incident Reporting](https://reportincident.iu.edu/index.html).

## STATEMENT ABOUT THE TWO-STEP LOGIN (DUO) FOR ALL CAS LOGINS

IUPUI students are required to enroll in Two-Step Login (Duo) to gain access to sensitive documents and Canvas using IU login credentials. The security of student information is critical. Be sure to bring your primary device (like a cellphone or tablet) to class, so you can log in to secure IU systems. Also, make sure you have a backup device like a hardware token or Google Voice. If you get stuck without a working device, the UITS Support Center can give you a bypass code, but you will need to verify your identity.

To learn more about or get help with two-step login, consult the following resources:

* [Help for Two-Step Login (Duo)](https://kb.iu.edu/d/aluu)
* [Two-Step Login (Duo) device recommendations](https://kb.iu.edu/d/anfl)
* [Contact your campus IT Support Center](https://kb.iu.edu/d/abxl) (for locations of UITS Support Centers and phone numbers and a chat feature).