New Course EFD Template

PURDUE UNIVERSITY | College of Engineering

Engineering Faculty Document
No.: [EFD #]
October 23, 2023

TO: The Engineering Faculty

FROM: The Faculty of the Lyles School of Civil Engineering

RE: New graduate course – CE 56401: Data Science for Smart Cities

The Faculty of the Lyles School of Civil Engineering has approved the following new graduate course. This action is now submitted to the Engineering Faculty with a recommendation for approval.

FROM:

CE 59700 Data Science for Smart Cities, 3 credits

Fall 2023 (18); Spring 2021 (4), Fall 2020 (13)

No Prerequisites

TO:

CE 56401: Data Science for Smart Cities

Spring or Fall

Three total credits

No Prerequisites

Course Description: The availability of low-cost and ubiquitous sensors in city infrastructure provides high granular data at unprecedented spatiotemporal scales. "Smart Cities" is a concept to utilize this data to provide a healthy, happy and sustainable urban ecosystem by integrating the information and communication technology, Internet of things, and citizen participation to effectively manage and utilize city's infrastructure and services. 'Data science' is an interdisciplinary field of scientific methods, processes, algorithms, and systems to extract knowledge from data in various forms. This course will introduce techniques that will allow the analysis, inference, and prediction of large-scale data (e.g., GPS trajectories, individual social networks, etc.) that are present in urban networks. The course will focus both on the methods and their application to smart city problems. Python will be used to demonstrate the application of each method on datasets available to the instructor. Examples of applications that will be discussed as an example of applications of data science for smart city applications include: ridesharing platforms, energy modeling, smart and energy-efficient buildings, evacuation modeling, decision-making during extreme events, and urban resilience.

RATIONALE:

The course addresses an urgent and growing need for learners and professionals who navigate the complexities of smart city planning and management using data science techniques. It contributes to existing majors and concentrations in computer science, data science, civil engineering, urban planning, and public policy by offering an interdisciplinary approach to tackling real-world issues. It complements other graduate courses in data analytics, civil engineering, and computer science by integrating these disciplines and applying them to the pressing needs in cities.

Head/Director of the Lyles School of Civil Engineering

Link to Curriculog entry: https://purdue.curriculog.com/proposal:25609/form

CE 56401: Data Science for Smart Cities

Classroom: **TBD**

Class hours: **TBD**

Instructors

Name	Prof. Satish Ukkusuri
Affiliation	Reilley Professor of Civil Engineering
Office location	HAMP G167D
Office hours	By appointment
Email	sukkusur@purdue.edu

Prerequisites

Undergraduate calculus and basic knowledge of statistical analysis.

Course description

The availability of low cost and ubiquitous sensors in city infrastructure provides high granular data at unprecedented spatiotemporal scales. "Smart Cities" is a concept to utilize this data to provide a healthy, happy and sustainable urban ecosystem by integrating the information and communication technology, Internet of things, and citizen participation to effectively manage and utilize city's infrastructure and services. 'Data science' is an interdisciplinary field of scientific methods, processes, algorithms, and systems to extract knowledge from data in various forms. This course will introduce techniques that will allow the analysis, inference, and prediction of large-scale data (e.g., GPS trajectories, individual social networks, etc.) that are present in urban networks. Basics of the data science methods to analyze these datasets will be presented. The course will focus both on the methods and their application to smart city problems. Python will be used to demonstrate the application of each method on datasets available to the instructor. Examples of problems that will be discussed as an example of applications of data science for smart city applications include: ridesharing platforms, energy modeling, smart and energy efficient buildings, evacuation modeling, decision-making during extreme events, and urban resilience.

Course objectives

A student completing this course is expected to be able to:

1. Understand the different types of data generated by smart cities.
2. Understand the basics of various data mining techniques.
3. Understand what type of data mining analysis is appropriate for various smart city applications.
4. Gain basic knowledge of using Python for data analytics and results visualization.
5. Apply the methods and techniques learned in this course to an applied smart city project.

Course readings

Required readings will be posted on a Box folder that will be shared with students.

Required software

We will use a Python-based scientific computing framework for all the classroom exercises, including the following programs and packages. Installation details will be provided in class.

- **Python**, an interpreted high-level programming language for general-purpose programming.
- **Anaconda** or **Miniconda**, a package environment manager for scientific computing.
- **Jupyter Notebook** or **Jupyter Lab**, a notebook-style code editor.
- Python packages for scientific computing, such as `numpy`, `pandas`, `matplotlib`, `scikit-learn`, `scipy`, etc.

Course requirements and evaluation

Homework

Problem sets will be given, and the analysis of these assignments will be the basis for some class discussion. Problem sets are due at the beginning of class on designated days. For the problem sets, you may (are encouraged to) discuss with other students but the final written solution should be your own work. The exam will be open class notes.

Exam

There will be one in-class examination in which you will be tested on readings, materials, and discussions covered in class.

Final project and paper

Students are expected to work in groups of 2 students on a problem related to data science application or algorithm implementation.

Class participation

Throughout the semester, classes will utilize a combination of lecture, discussion, activities, and case analyses. Although all lectures will be virtual, I hope to engage with students actively. Please keep your videos on and engage in the discussion both in and outside the class. Students should prepare for class by reading all required articles in advance and engage in informed listening and frequent contribution to the discussions. There will be several short quizzes randomly throughout the semester at the beginning of class. Completing assigned readings before class is essential to the success of your quizzes. If you miss a quiz, you will receive a zero for it. The quizzes will be used as a proxy for attendance and class participation. They will not be returned.

Grading scheme

The final grade for this course is based on the following weighted components:

- Homework 25%
- Quizzes 5%
- Exams 35%
- Final project and paper 35%

Final grades will be assigned as follows (subject to change at the instructors discretion):

A+	97–100%	B+	87–90%	C+	77–80%	D+	67–70%	F	≤59%
A	93–96%	B	83–86%	C	73–76%	D	63–66%		
A–	90–92%	B–	80–82%	C–	70–72%	D–	60–62%		

Course policies

Attendance

Attendance is required. We will not record attendance at each class. More than two unexcused absences will lead to reduction of your class participation grade. Patterns of tardiness will also be a concern. If you are absent for a class, you are responsible for all class materials and information mentioned in the class.

Personal technology

Please do not use a cell phone, laptop, or other personal devices for any purpose unrelated to the class. Instances of distracting behaviors such as texting and web-surfing will lead to reduction of your class participation grade without notice. Cell phones may be left on vibrate for emergency notification purposes. If you are expecting an important phone call, please inform me beforehand and I will understand if you leave the classroom to take a call.

Late assignments

All assignments need to be turned in by due date. Late assignments will incur a penalty of 20% of the assignment grade per day, including weekends and holidays. Assignments turned in late due to computer and other technical problems will not be excused unless in emergencies. No extension will be given unless you provide me with official documentation (e.g., doctor's note in case of illness) in advance. It is your responsibility to keep track of the due dates indicated in the syllabus.

Email and communication

You are responsible for reading and responding to email messages from instructors. Please expect a response to your emails within 48 hours during weekdays. If you have not heard back within 48 hours, please follow up to ensure that we have received your email.

Grief absence policy

Purdue University recognizes that a time of bereavement is very difficult for a student. The University therefore provides the following rights to students facing the loss of a family member through the Grief Absence Policy for Students (GAPS). GAPS Policy: Students will be excused for funeral leave and given the opportunity to earn equivalent credit and to demonstrate evidence of meeting the learning outcomes for misses assignments or assessments in the event of the death of a member of the student's family. See the University's website for additional information: http://www.purdue.edu/studentregulations/regulations_procedures/classes.html.

Accessibility and accommodations

Purdue University strives to make learning experiences as accessible as possible. 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. See Disability Resource Center's website for additional information: https://www.purdue.edu/drc.

Mental health

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 support, services are available. For help, such individuals should contact Counseling and Psychological Services (CAPS) at (765)494-6995 and http://www.purdue.edu/caps/ during and after hours, on weekends and holidays, or through its counselors physically located in the Purdue University Student Health Center (PUSH) during business hours.

Violent behavior policy

Purdue University is committed to providing a safe and secure campus environment for members of the university community. Purdue strives to create an educational environment for students and a work environment for employees that promote educational and career goals. Violent Behavior impedes such goals. Therefore, Violent Behavior is prohibited in or on any University Facility or while participating in any university activity. See the University's website for additional information: http://www.purdue.edu/policies/facilities-safety/iva3.html.

Nondiscrimination

Purdue University is committed to maintaining a community which 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 own 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.

Purdue's nondiscrimination policy can be found at:

http://www.purdue.edu/purdue/ea_eou_statement.html

Ethics and academic integrity

"As a boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do. Accountable together - we are Purdue." Here is a link to a web page for Purdue's Honor Pledge.

Purdue prohibits "dishonesty in connection with any University activity. Cheating, plagiarism, or knowingly furnishing false information to the University are examples of dishonesty." [Part 5, Section III-B-2-a, Student Regulations] Furthermore, the University Senate has stipulated that "the commitment of acts of cheating, lying, and deceit in any of their diverse forms (such as the use of substitutes for taking examinations, the use of illegal cribs, plagiarism, and copying during examinations) is dishonest and must not be tolerated. Moreover, knowingly to aid and abet, directly or indirectly, other parties in committing dishonest acts is in itself dishonest." [University Senate Document 72-18, December 15, 1972]

Academic integrity is one of the highest values that Purdue University holds. Individuals are encouraged to alert university officials to potential breeches of this value by either emailing integrity@purdue.edu or by calling 765-494-8778. While information may be submitted anonymously, the more information that is submitted provides the greatest opportunity for the university to investigate the concern.

See Purdue's student guide for academic integrity (https://www.purdue.edu/odos/academic-integrity) for more information.

Campus emergencies

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 instructors or TAs via email or phone. You are expected to read your @purdue.edu email on a frequent basis.

When we hear fire alarms or are instructed to leave the building, we will use stairways to immediately evacuate the building and gather at the emergency assembly area location: XX

When we hear all hazards sirens, immediately seek shelter (Shelter-In-Place) inside the building. This course of action may need to be taken during a tornado, earthquake, release of hazardous materials in the outside air, or a civil disturbance. When you hear the sirens immediately go inside a building to a safe location and use all communication means available to find out more details about the emergency. Remain in place until police, fire, or other emergency response personnel provide additional guidance or tell you it is safe to leave.

For additional information, see: https://www.purdue.edu/ehps/emergency_preparedness

Disclaimer

I reserve the right to change readings and assignments based on the progress of the class. All changes to the syllabus will be announced in advance.

Course schedule, readings, and assignments

Date	Topic	Readings/assignments due
	UNIT 1. Introduction to data mining	
Week 1	Part 1: Introduction to the course & syllabus • Instructor introduction • Introduction to data mining • Project overview- outline and potential topics • Student introduction Part 2: Review of Statistical methods • Introduction • Modeling uncertainty • Random variables • Population and samples • Statistical inference	Bayesian vs frequentist inference Review of random variables (Sections 1.1.1, 1.1.2, 1.1.6, 1.2) McGill Paul Ch. 4 - Statistical inference
Week 2	Part 1: Optimization • Introduction • Optimization - Basic concepts • Optimization problem formulation • Optimization algorithms Part 2: Data pre-processing • Data and measurement • Types of datasets • Data quality • Data pre-processing • Task identification Part 3: Intro. to Python (Installation and basics)	Sheffi book (Ch. 2 and 4) Morgan-Kauffman (2011) (Ch. 3) **Assignment #1**
Week 3	Part 1: Project Discussion/Introduction to Python • Introduction • Python for data mining Part 2: Project Discussion/Introduction to Python • Optimization using Python • Data pre-processing using Python Part 3: Python exercises	Intro to ML in Python Scipy documentation Scikit-learn preprocessing
	UNIT 2. Data mining tasks	
Week 4	Part 1: Regression analysis • Introduction • Linear regression • Logistic Regression Models • Poisson Regression Models	Intro to regression Logistic regression Association analysis

	• Applications of regression analysis to smart cities Part 2: Association rule mining • Introduction • Association rule mining applications to Urban systems • Association rule mining approaches Part 3: Python exercises	**Identify teams for the project**
Week 5	Part 1: Association rule mining (contd...) • F-P growth algorithm • ECLAT (Equivalence Class Clustering and bottom up Lattice Traversal) • Evaluation methods Part 2: Statistical classification • Introduction to the classification problem • Logistic regression • Naive Bayes Classifier • Bayesian Network Classifier Part 3: Python tutorial	Association analysis Naive Bayes and logistic regression Bayesian networks **Assignment #2** **Identify project topics**
Week 6	Part 1: Decision Trees • Introduction • Decision tree training • Decision tree algorithms • Practical issues with decision trees Part 2: Support Vector Machines • Introduction • Support Vector Machines • Ensemble Classifiers • Classifier performance evaluation Part 3: Python tutorial	Decision trees Ensemble methods
Week 7	Part 1: Introduction to data clustering • Introduction • (Dis)similarity measures • Distribution (Model)-based clustering algorithms Part 2: Clustering algorithms: Partitional and Hierarchical • Types of Clustering algorithms • Partitional clustering (k-means and its variants)	Intro to clustering **Assignment #3**

	• Hierarchical clustering	
Week 8	**Break**	
Week 9	Part 1: Other clustering approaches • Density based clustering algorithms • Cluster validity • Characteristics of Data, Clusters, and Clustering Algorithms Part 2: Anomaly detection (**instructed by Dr. Mondal**) • Introduction • The anomaly detection problem • Anomaly detection techniques	Intro to clustering Anomaly detection
Week 9	Part 1: **Review of material and project check up** Part 2: Avoiding false discoveries • Introduction • Statistical significance testing • Hypothesis testing • Multiple hypothesis testing	False discoveries **Assignment #4**
UNIT 3. Advanced data mining techniques		
Week 10	Neural Networks • Introduction • A Neuron model • Learning an ANN model • Multi-layer feed-forward ANNs • ANN application to land use prediction Part 2: Deep learning (**instructed by Dr. Mondal**) • Introduction to deep learning • Deep learning for smart cities	Neural networks (Ch. 1, 3, 5) Deep learning (Ch. 6–9) **Assignment #5**
Week 11	Part 1: Reinforcement Learning (RL) • Introduction • Basics of Markov Decision Process (Examples, Markov Property Definition) • Bellman Equation • Policy Evaluation (Coding Examples) via iteration Part 2: Advanced methods in RL • Value Iteration (VI) Algorithm • Policy Iteration (PI) • Coding Examples • Limitations of VI and PI • Intro to Deep RL	Reinforcement learning

Week 12	Part 1: Advanced methods in RL • Neural Networks as Policy Approximator • Classification of DRL Algorithms • Intro to Deep Q-Learning (DQL) • Implementation of DQL • Policy Gradient Theorem • Intro to Policy Gradient (PG) Algo • Implementation of vanilla PG Part 2: Additional methods • Advantage Actor-Critic (A2C) PG • Trust Region Policy Optimization (TRPO) • Deep Deterministic PG (DDPG) • Partially Observable MDP (POMDP) • Hidden Markov Models (HMM)	
Week 13	**Exam (in class)**	
Week 14	**Break**	
Week 15	**Project presentations**	
	Project report due	