CE 597: Data Science for Smart Cities

Classroom: HAMP 1252
Class Hours: 2:30-5:15 pm, Monday

Instructors

Prof. Satish Ukkusuri
Civil Engineering
HAMP G167D, sukkusur@purdue.edu
Office Hours: By Appointment

Dr. Washim Uddin Mondal
Civil Engineering
HAMP XXX,
Office Hours: Wed and Fri from 3-5pm

Teaching Assistant

Rajat Verma
XXX

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 spatio-temporal 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 (ICT), Internet of things (IoT) 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 scientific techniques that will allow the analysis, inference and prediction of large-scale data (e.g. GPS vehicular data, social media data, mobile phone data, individual social network data etc.) that are present in city 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 & 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
Python (interpreted high-level programming language for general-purpose programming) and its libraries for data mining (numpy, scipy, matplotlib, etc.). Available for free download at https://www.python.org/. Details will be provided in class.

Course Requirements & Evaluation
1. Homeworks
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.

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

4. 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.

5. 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.

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

Homeworks	25%
Quizzes	5%
Exam	35%
Final project and paper	35%

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

A+ = 97-100%	C+ = 77-80%

A = 93-96%	C = 73-76%
A- = 90-92%	C- = 70-72%
B+ = 87-90%	D+ = 67-70%
B = 83-86%	D = 63-66%
B- = 80-82%	D- = 60-62%
	F = 59% and below

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 Assignment:

Date	Topic	Readings/Assignments Due
UNIT 1. Introduction to data mining		
Week 1 Aug. 21	Introduction to the course & syllabus • Instructor introduction • Introduction to data mining • Project overview- outline and potential topics • Student introduction	
Week 1 Aug. 21	Review of Statistical methods • Introduction • Modeling uncertainty • Random variables • Population and samples • Statistical inference	https://pdfs.semanticscholar.org/5777/2c52696be0881728ebde18eb84c8397309b8.pdf https://faculty.washington.edu/ezivot/econ424/probreview.pdf (Section 1.1.1, 1.1.2, 1.1.6, 1.2) http://www.cim.mcgill.ca/~paul/StIEs43z.pdf
Week 2 Aug. 28	Optimization • Introduction • Optimization - Basic concepts • Optimization problem formulation • Optimization algorithms	http://web.mit.edu/sheffi/www/selectedMedia/sheffi_urban_trans_networks.pdf (chapter 2 and 4)
Week 2 Aug. 28	Data pre-processing • Data and measurement • Types of datasets • Data quality • Data pre-processing • Task identification Introduction to Python (Installation and Basics)	http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf (chapter 3) **Homework 1**
Week 3 Sept. 4	Project Discussion/Introduction to Python • Introduction • Python for data mining	https://pythonprogramming.net/machine-learning-tutorial-python-introduction/
Week 3	Project Discussion/Introduction to Python	https://docs.scipy.org/doc/scipy/r

Sept. 4	• Optimization using Python • Data pre-processing using Python	eference/tutorial/optimize.html http://scikit-learn.org/stable/modules/preprocessing.html
UNIT 2. Data mining Tasks		
Week 4 Sept. 11	Regression analysis • Introduction • Linear regression • Logistic Regression Models • Poisson Regression Models • Applications of regression analysis to smart cities	http://statlab.stat.yale.edu/workshops/IntroRegression/StatLab-IntroRegressionFa08.pdf http://chem-eng.utoronto.ca/~datamining/Presentations/Logistic_Regression.pdf http://data.princeton.edu/wws509/notes/c4.pdf
Week 4 Sept. 11	Association rule mining • Introduction • Association rule mining applications to Urban systems • Association rule mining approaches	https://www-users.cs.umn.edu/~kumar001/dmbook/ch6.pdf **Identify teams for the project**
Week 5 Sept 18	Association rule mining (contd...) • F-P growth algorithm • ECLAT (Equivalence Class Clustering and bottom up Lattice Traversal) • Evaluation methods **Project Check Up**	https://www-users.cs.umn.edu/~kumar001/dmbook/ch6.pdf https://www.aaai.org/Papers/KDD/1997/KDD97-060.pdf **Homework 2, Identify project topics**
Week 5 Sept. 18	Statistical Classification • Introduction to the classification problem • Logistic regression • Naive Bayes Classifier • Bayesian Network Classifier Python Tutorial	https://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf http://www.cs.technion.ac.il/~dang/journal_papers/friedman1997Bayesian.pdf
Week 6 Sept 25	Decision Trees • Introduction • Decision tree training • Decision tree algorithms	http://www.cs.princeton.edu/courses/archive/spr07/cos424/papers/mitchell-dectrees.pdf

	• Practical issues with decision trees	
Week 6 Sept 25	Support Vector Machines • Introduction • Support Vector Machines • Ensemble Classifiers • Classifier performance evaluation Python Tutorial	http://cs229.stanford.edu/notes/cs229-notes3.pdf http://web.engr.oregonstate.edu/~tgd/publications/mcs-ensembles.pdf
Week 7 Oct 2	Introduction to data clustering • Introduction • (Dis)similarity measures • Distribution (Model)-based clustering algorithms	https://www-users.cs.umn.edu/~kumar001/dmbook/ch8.pdf **Homework 3**
Week 7 Oct 2	Clustering algorithms: Partitional and Hierarchical • Types of Clustering algorithms • Partitional clustering (k-means and its variants) • Hierarchical clustering	https://www-users.cs.umn.edu/~kumar001/dmbook/ch8.pdf
Week 8 Oct 9	**FALL BREAK**	**FALL BREAK**
Week 9 Oct 16	Other Clustering approaches • Density based clustering algorithms • Cluster validity • Characteristics of Data, Clusters, and Clustering Algorithms	https://www-users.cs.umn.edu/~kumar001/dmbook/ch8.pdf
Week 9 Oct. 16	Anomaly detection • Introduction • The anomaly detection problem • Anomaly detection techniques **Instructor: Washim Uddin Mondal**	http://cucis.ece.northwestern.edu/projects/DMS/publications/AnomalyDetection.pdf
Week 9 Oct. 23	**Review of Material and Project Check up**	
Week 9 Oct. 23	Avoiding false discoveries • Introduction • Statistical significance testing • Hypothesis Testing • Multiple hypothesis testing	https://www-users.cs.umn.edu/~kumar001/dmbook/slides/chap10_false_discoveries.pdf **Homework 4**
UNIT 3. Advanced data mining techniques		
Week 10 Oct. 30	Neural Networks • Introduction	http://www.dkriesel.com/_media/science/neuronalenetze-en-

Week 10 Oct 30	• A Neuron model • Learning an ANN model • Multi-layer feed-forward ANNs • ANN application to land use prediction	zeta2-2col-dkrieselcom.pdf **Homework 5**
Week 10 Oct. 30	Deep learning • Introduction • Deep Learning • Deep Learning for smart cities **Instructor: Washim Uddin Mondal**	https://www.deeplearningbook.org/front_matter.pdf (chapter 6-9)
Week 11 Nov. 6	Reinforcement Learning **Instructor: Washim Uddin Mondal**	http://www.dkriesel.com/_media/science/neuronalenetze-en-zeta2-2col-dkrieselcom.pdf (chapter 10)
Week 11 Nov 6	Reinforcement Learning **Instructor: Washim Uddin Mondal**	
Week 12 Nov 13	Reinforcement Learning **Instructor: Washim Uddin Mondal**	
Week 12 Nov 13	Reinforcement Learning **Instructor: Washim Uddin Mondal**	
Week 13 Nov 20	**EXAM (IN CLASS)**	**EXAM (IN CLASS)**
Week 14 Nov 27	BREAK	BREAK
Week 15 Dec 4	PROJECT PRESENTATIONS	PROJECT PRESENTATIONS
Dec 9	Project Report Due	Project Report Due
Week 11 Nov. 6	Hidden Markov Models (HMMs) • Introduction • Markov process • Graphical models • Discrete-State HMMs • Applications of HMMs to Smart Cities	https://web.stanford.edu/~jurafsky/slp3/9.pdf
UNIT 4. Mining Big Data		
Week 12 Nov 10	Mining of Massive Datasets • Map-Reduce and the New Software Stack	http://www.mmds.org/
Week 12 Nov. 10	Data mining algorithms for massive datasets • Frequent itemsets: PCY (Park-Chen-Yu) Algorithm • Clustering: BFR, CURE	http://www.mmds.org/

Addtl Topics	Data mining algorithms for massive datasets • Frequent itemsets: PCY (Park-Chen-Yu) Algorithm • Clustering: BFR, CURE	http://www.mmds.org/
Addtl Topics	Dimensionality reduction • SVD • CUR	http://www.mmds.org/
Addtl Topics	Mining Streaming data • Sampling from a data stream: Reservoir sampling • DGIM algorithm • Filtering data streams	http://infolab.stanford.edu/~ullman/mmds/ch4.pdf
Addtl Topics	Mining social network graphs • Community detection • Spectral clustering	http://www.mmds.org/
Addtl Topics	Mining social network graphs and Wrap up • Overlapping communities • Random walks based methods	http://www.mmds.org/