Electrical & Computer ENGINEERING

Carnegie Mellon

Course Syllabus

18-787: Data Analytics
Spring 2023

Instructor: Prof. Patrick McSharry
Email Address: mcsharry@cmu.edu
Course Management Assistant: Stephanie Bolla
Email Address: sbolla@andrew.cmu.edu
Office Location: HH 1113
Waitlist Manager: Megan Oliver
Email Address: mvoliver@andrew.cmu.edu
Office Location: HH 1113B
Course Website: https://diaml.github.io/18787/

Teaching Assistants:

Name:	Email:	Zoom Meeting Id:
Christian Iradukunda	Christianiradukunda2@gmail.com	https://cmu.zoom.us/my/christian.iradukunda
Seth Abayo	sabayo@andrew.cmu.edu	https://cmu.zoom.us/my/sabayo
Remy Patrick Mwizerwa	rmwizerw@andrew.cmu.edu	https://cmu.zoom.us/my/rmwizerw
Gabriel Ntwari	ntwarig@andrew.cmu.edu	https://cmu.zoom.us/my/Gabriel Ntwari
Innocent Mukoki	imukoki@andrew.cmu.edu	https://cmu.zoom.us/my/imukoki

Course Discipline: Electrical and Computer Engineering
Course Level: Graduate
Type of Course: Onsite - Remote
Course Streamed to Pittsburgh: Yes
Course Concentration: Applied Machine Learning
Core/Elective: Core
Units: 6
Lecture/Lab/Rep Hours/Week: 3 Lecture Hours Per Week; 1 Lab/Rep Hours Per Week
Semester/Year Offered: Spring, All Years

Pre-Requisites: 18-785
Lecture Zoom details:
ID: 950 7844 3041
Passcode: 810320
Direct url: https://cmu.zoom.us/j/95078443041?pwd=alRQNE9qUUtTMVZNdXk2TStmN3Y3QT09

Recitation Zoom details:
ID: 910 0808 2602
Passcode: 571926
Direct url: https://cmu.zoom.us/j/91008082602?pwd=SjMzam9XTFl0Yjg0RWdmQ1llbXBZQT09

Class Schedule:
Pittsburgh: *Remote*
Kigali: *F305 - Remote*

Lectures: *1 hour and 20 minutes*

Week	A	Pittsburgh	Kigali	B	Pittsburgh	Kigali
1	Tues, Jan 17	08:00 ET	15:00 CAT	Thurs, Jan 19	08:00 ET	15:00 CAT
2	Fri, Jan 20	08:00 ET	15:00 CAT	Tue, Jan 24	08:00 ET	15:00 CAT
3	Tues, Jan 31	08:00 ET	15:00 CAT	Thurs, Feb 02	08:00 ET	15:00 CAT
4	Tues, Feb 07	08:00 ET	15:00 CAT	Thurs, Feb 09	08:00 ET	15:00 CAT
5	Tues, Feb 14	08:00 ET	15:00 CAT	Thurs, Feb 16	08:00 ET	15:00 CAT
6	Tues, Feb 21	08:00 ET	15:00 CAT	Thurs, Feb 23	08:00 ET	15:00 CAT

Course Description:
Data analytics is the process of analyzing datasets in order to draw conclusions, generate insights, make predictions and support decision-making. This course will take a practical approach to solving challenges in the public and private sectors using data analytics. A number of different themes will be explored as case studies in order to demonstrate how data-driven decision-making has widespread applications. The course will examine how the question being posed, the available data and the selected modelling approach all come together to arrive at a feasible solution. A range of quantitative techniques, involving both linear and nonlinear methods will be presented for dealing with numerical structured datasets. Substantial emphasis will be placed on the process of delivering data analytics via a dashboard to facilitate decision-making and policy-making. The course content will be structured to provide a roadmap for carrying out the necessary procedures and will be illustrated using case studies, reading material and previously published models. Participants will obtain hands-on experience by working on specific challenges with real-world data through a carefully structured set of assignments.

Learning Objectives:
The objective of this course is to provide students with practical experience of undertaking a data analytics project to address a challenge of their choice. This process involves: (1) proposing a feasible research plan based on a challenge that can be addressed with available data; (2) exploratory analysis of the data; (3) construction and evaluation of a quantitative model; and (4) communication of the model output to decision-makers. This will involve identifying a relevant key performance indicator and collecting appropriate explanatory variables. The ability to write a report that documents the challenge, data, analysis, model, evaluation and feasibility of deploying the analytics in a real-world setting will be assessed. Applications will include forecasting, classification, novelty detection and risk monitoring.

Outcomes

After completing this course, students should be able to:

- Design a data analytics project in response to a specific challenge
- Download and organize data for addressing the challenge
- Explore the dataset using visualization techniques
- Apply a range of quantitative techniques
- Discuss the advantages and disadvantages of different models
- Select an approach that is optimal for meeting the objective
- Present conclusions and recommendations
- Communicate model output to decision-makers

Grading

The grades for this course will be based on students' performance on three homework assignments, Canvas quizzes, a final exam and class participation. Homework assignments will be done individually and turned in via Canvas by the designated due date. Late work will be acceptable until 24 hours past the deadline, but it will lose 10%. The assignments will be graded based on both a writing report and code used to achieve results presented in the report. Class participation will be evaluated based on the student's contribution to discussions both in-class and on the Piazza Discussion Board. When posting or reacting to online discussion threads, students are expected to use their own words and the post should be relevant to the topic under discussion. Make sure to introduce, summarize and explain the article in your own words to enlighten the audience on the point the article is making.

The following is the weight distribution of the grades:

Class participation	**5%**
Canvas quizzes	**2.5%**
Piazza participation	**2.5%**
Assignment 1	**20%**
Assignment 2	**25%**
Assignment 3	**30%**
Final Exam (Multiple Choice)	**15%**

Important Dates:

Description	Release Date	Due Date	Days	Grade Date
Assignment 1	Tue, Jan 17	Mon, Jan 30	14	Fri, Feb 03
Assignment 2	Tue, Jan 31	Mon, Feb 13	14	Fri, Feb 17
Assignment 3	Tue, Feb 14	Mon, Feb 27	14	Fri, Mar 03
Final Grades				Wed, Mar 13

* The final exam date is to be announced but likely to be the first week of March.

Deadlines:

It is expected that deadlines are respected and met on time. Missing a deadline by between 0 and 24 hours will result in the deduction of 10% of the marks assigned. A submission made later than 24 hours after the deadline, will be regarded as unclassified and result in zero marks. In the case of a situation preventing a student from meeting the

deadline (such as a medical condition), a student is required to write to the professor and TAs of the course in advance, copy his/her advisor and submit supporting evidence.

MATLAB:
One option is to use MATLAB software. Download MATLAB software for your computer operating system from the CMU download website. After unzipping the file, read the Matlab-Licence_Instructions.pdf file for instructions for connecting to MATLAB server and running Matlab.

Python
Another option is to use Python. Google Colaboratory (Colab) is an excellent platform for building notebooks which allow you to combine executable code and rich text in a single document, along with images, HTML, LaTeX.

Canvas and Piazza:
Canvas will be used for posting supplementary course materials and turning in assignments. Please familiarize yourself with navigating, uploading and downloading. Piazza will be used for questions and discussion among students, TAs and the instructor.

Topic Outline (Weeks 1-6)

Data Analytics 18-787			
Week	Activity	Lecture A	Lecture B
1	**Theme**	**Data Analytics**	**Weather forecasting**
	Challenge	Human versus machine	Weather forecasting
	Discussion	Course objectives and timeliness	History of weather forecasting
	Case Study	Data scientist	Michael Fish's hurricane
	Analysis	Statistician versus data scientist	Communicating forecasts
	Demo	Roadmap	Evaluating forecasts
2	**Theme**	**Renewable energy**	**Wind energy**
	Challenge	Supply variability	Wind energy integration
	Discussion	Forecasting supply	Wind energy growth
	Case Study	Ensemble forecasts	Wind farm forecasting
	Analysis	Electricity prices	Probabilistic forecasting
	Demo	Renewable policy	Variability indices
3	**Theme**	**Solar energy**	**Demand forecasting**
	Challenge	Storage and price	Seasonality
	Discussion	PV and thermal	Demand and weather
	Case Study	Actual solar efficiency	Peak scenarios
	Analysis	Efficiency versus irradiance	Short-term forecasts
	Demo	Smoothing techniques	Electricity consumption
4	**Theme**	**Risk**	**Extreme events**
	Challenge	Risk management	Floods
	Discussion	Global risks	Extreme values
	Case Study	Natural disasters	Climate change and flood risk
	Analysis	Disaster trends	Wind farm risk metric

	Demo	Risk communication	Estimating V50
5	**Theme**	**Health**	**Biomedicine**
	Challenge	Data-driven medicine	Detecting disorders
	Discussion	History of healthcare	Epilepsy
	Case Study	Growth charts	ECG analysis
	Analysis	Electronic records	Parkinson's disease
	Demo	Rwandan growth chart	Voice disorders
6	**Theme**	**Early warning systems**	**Economic forecasting**
	Challenge	Preventing disaster	Economic activity
	Discussion	Components of EWS	Expansion and recession
	Case Study	Parkinson's disease	US GNP forecasting
	Analysis	Nuclear reactor clogging	Regime switching
	Demo	UK Met Office	Parsimonious models

Grading Scale:

A : > 93% A - :]93% , 90%]	C + :]80%, 77%] C :]77%, 73 %] C - :]73%, 70%]
B + :]90% , 87%] B :]87% , 83 %] B - :]83% , 80%]	D + :]70%, 67%] D :] 67%, 60%] F : < 60 %

Education Objectives (Relationship of Course to Program Outcomes)
The ECE department is accredited by ABET to ensure the quality of your education. ABET defines 7 Educational Objectives that are fulfilled by the sum total of all the courses you take. The following list describes which objectives are fulfilled by this course and in what manner they are fulfilled. The objectives are numbered from "1" through "7" in the standard ABET parlance. Those objectives not fulfilled by this course have been omitted from the following list:

1) **an ability to identify, formulate, and solve complex engineering problems by applying principles of engineering, science, and mathematics:** The course poses many problems in homework assignments involving data from complex real-world systems that will require students to think critically, analyse, model and solve using good engineering practice.
2) **an ability to apply engineering design to produce solutions that meet specified needs with consideration of public health, safety, and welfare, as well as global, cultural, social, environmental, and economic factors:** The course introduces the use of analytics and machine learning to address specific challenges and facilitate improved decision-making for practitioners and managers in public and private sectors.
3) **an ability to communicate effectively with a range of audiences:** Students can practice communication skills remotely (Piazza Q&A), during class (quizzes, polls, questions and answers) and especially during recitations where challenges are discussed and volunteers with solutions are welcome to present.

4) **an ability to recognize ethical and professional responsibilities in engineering situations and make informed judgments, which must consider the impact of engineering solutions in global, economic, environmental, and societal contexts:** The ethical aspects of analytics and machine learning are discussed in relation to data sampling biases, model error, risk and uncertainty. Case studies are used to demonstrate the potential adverse consequences of AI.
5) **an ability to function effectively on a team whose members together provide leadership, create a collaborative and inclusive environment, establish goals, plan tasks, and meet objectives:** Quizzes and polls are used as a mechanism to encourage group discussion and motivate debate about the process of interpreting data and arriving at implementable solutions.
6) **an ability to develop and conduct appropriate experimentation, analyze and interpret data, and use engineering judgment to draw conclusions:** The course focuses extensively on the interpretation and analysis of data to predict outcomes and generate actionable insights based on transparent processes that can be documented and explained.
7) **an ability to acquire and apply new knowledge as needed, using appropriate learning strategies:** Students will tackle problems using real-world datasets and will select from amongst a range of techniques and models to undertake statistical learning and arrive at a robust and predictive solution.

ECE Academic Integrity Policy (http://www.ece.cmu.edu/programs-admissions/masters/academic-integrity.html**):**

The Department of Electrical and Computer Engineering adheres to the academic integrity policies set forth by Carnegie Mellon University and by the College of Engineering. ECE students should review fully and carefully Carnegie Mellon University's policies regarding Cheating and Plagiarism; Undergraduate Academic Discipline; and Graduate Academic Discipline. ECE graduate student should further review the Penalties for Graduate Student Academic Integrity Violations in CIT outlined in the CIT Policy on Graduate Student Academic Integrity Violations. In addition to the above university and college-level policies, it is ECE's policy that an ECE graduate student may not drop a course in which a disciplinary action is assessed or pending without the course instructor's explicit approval. Further, an ECE course instructor may set his/her own course-specific academic integrity policies that do not conflict with university and college-level policies; course-specific policies should be made available to the students in writing in the first week of class.

This policy applies, in all respects, to this course.

CMU Academic Integrity Policy (http://www.cmu.edu/academic-integrity/index.html**):**
In the midst of self-exploration, the high demands of a challenging academic environment can create situations where some students have difficulty exercising good judgment. Academic challenges can provide many opportunities for high standards to evolve if students actively reflect on these challenges and if the community supports discussions to aid in this process. It is the responsibility of the entire community to establish and maintain the integrity of our university.

This site is offered as a comprehensive and accessible resource compiling and organizing the multitude of information pertaining to academic integrity that is available from across the university. These pages include practical information concerning policies, protocols and best practices as well as articulations of the institutional values from which the policies and protocols grew. The Carnegie Mellon Code, while not formally an honor code, serves as the foundation of these values and frames the expectations of our community with regard to personal integrity.
This policy applies, in all respects, to this course.

The Carnegie Mellon Code

Students at Carnegie Mellon, because they are members of an academic community dedicated to the achievement of excellence, are expected to meet the highest standards of personal, ethical and moral conduct possible.
These standards require personal integrity, a commitment to honesty without compromise, as well as truth without equivocation and a willingness to place the good of the community above the good of the self. Obligations once undertaken must be met, commitments kept.
As members of the Carnegie Mellon community, individuals are expected to uphold the standards of the community in addition to holding others accountable for said standards. It is rare that the life of a student in an academic community can be so private that it will not affect the community as a whole or that the above standards do not apply.
The discovery, advancement and communication of knowledge are not possible without a commitment to these standards. Creativity cannot exist without acknowledgment of the creativity of others. New knowledge cannot be developed without credit for prior knowledge. Without the ability to trust that these principles will be observed, an academic community cannot exist.
The commitment of its faculty, staff and students to these standards contributes to the high respect in which the Carnegie Mellon degree is held. Students must not destroy that respect by their failure to meet these standards. Students who cannot meet them should voluntarily withdraw from the university.

This policy applies, in all respects, to this course.

Carnegie Mellon University's Policy on Cheating

(http://www.cmu.edu/academic-integrity/cheating/index.html) states the following:
According to the University Policy on Academic Integrity, cheating "occurs when a student avails her/himself of an unfair or disallowed advantage which includes but is not limited to:

- Theft of or unauthorized access to an exam, answer key or other graded work from previous course offerings.
- Use of an alternate, stand-in or proxy during an examination.
- Copying from the examination or work of another person or source.
- Submission or use of falsified data.
- Using false statements to obtain additional time or other accommodation.
- Falsification of academic credentials."

This policy applies, in all respects, to this course.

Carnegie Mellon University's Policy on Plagiarism
(http://www.cmu.edu/academic-integrity/plagiarism/index.html) states the following: According to the University Policy on Academic Integrity, plagiarism "is defined as the use of work or concepts contributed by other individuals without proper attribution or citation. Unique ideas or materials taken from another source for either written or oral use must be fully acknowledged in academic work to be graded. Examples of sources expected to be referenced include but are not limited to:

- Text, either written or spoken, quoted directly or paraphrased.
- Graphic elements.
- Passages of music, existing either as sound or as notation.
- Mathematical proofs.
- Scientific data.
- Concepts or material derived from the work, published or unpublished, of another person."

This policy applies, in all respects, to this course.

Carnegie Mellon University's Policy on Unauthorized Assistance
(http://www.cmu.edu/academic-integrity/collaboration/index.html) states the following: According to the University Policy on Academic Integrity, unauthorized assistance "refers to the use of sources of support that have not been specifically authorized in this policy statement or by the course instructor(s) in the completion of academic work to be graded. Such sources of support may include but are not limited to advice or help provided by another individual, published or unpublished written sources, and electronic sources. Examples of unauthorized assistance include but are not limited to:

- Collaboration on any assignment beyond the standards authorized by this policy statement and the course instructor(s).
- Submission of work completed or edited in whole or in part by another person.
- Supplying or communicating unauthorized information or materials, including graded work and answer keys from previous course offerings, in any way to another student.
- Use of unauthorized information or materials, including graded work and answer keys from previous course offerings.
- Use of unauthorized devices.
- Submission for credit of previously completed graded work in a second course without first obtaining permission from the instructor(s) of the second course. In the case of concurrent courses, permission to submit the same work for credit in two courses must be obtained from the instructors of both courses."

This policy applies, in all respects, to this course.

Carnegie Mellon University's Policy on Research Misconduct
(http://www.cmu.edu/academic-integrity/research/index.html) states the following:

According to the University Policy for Handling Alleged Misconduct in Research, “Carnegie Mellon University is responsible for the integrity of research conducted at the university. As a community of scholars, in which truth and integrity are fundamental, the university must establish procedures for the investigation of allegations of misconduct of research with due care to protect the rights of those accused, those making the

allegations, and the university. Furthermore, federal regulations require the university to have explicit procedures for addressing incidents in which there are allegations of misconduct in research."

The policy goes on to note that "misconduct means:

- fabrication, falsification, plagiarism, or other serious deviation from accepted practices in proposing, carrying out, or reporting results from research;

- material failure to comply with Federal requirements for the protection of researchers, human subjects, or the public or for ensuring the welfare of laboratory animals; or

- failure to meet other material legal requirements governing research."

"To be deemed misconduct for the purposes of this policy, a 'material failure to comply with Federal requirements' or a 'failure to meet other material legal requirements' must be intentional or grossly negligent."

To become familiar with the expectations around the responsible conduct of research, please review the guidelines for Research Ethics published by the Office of Research Integrity and Compliance.

This policy applies, in all respects, to this course.

Take care of yourself. Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

All of us benefit from support during times of struggle. You are not alone. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful.

If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. Counseling and Psychological Services (CaPS) is here to help: call 412-268-2922 and visit their website at http://www.cmu.edu/counseling/. Consider reaching out to a friend, faculty or family member you trust for help getting connected to the support that can help.

If you have questions about this or your coursework, please let me know.

Every individual must be treated with respect. The ways we are diverse are many and are critical to excellence and an inclusive community. They include but are not limited to: race, color, national origin, sex, disability, age, sexual orientation, gender identity, religion, creed, ancestry, belief, veteran status, or genetic information. We at CMU, will

work to promote diversity, equity and inclusion because it is just and necessary for innovation. Therefore, while we are imperfect, we will work inside and outside of our classrooms, to increase our commitment to build and sustain a community that embraces these values.

It is the responsibility of each of us to create a safer and more inclusive environment. Bias incidents, whether intentional or unintentional in their occurrence, contribute to creating an unwelcoming environment for individuals and groups at the university. If you experience or observe unfair or hostile treatment on the basis of identity, we encourage you to speak out for justice and support in the moment and and/or share your experience anonymously using the following resources:
Center for Student Diversity and Inclusion: csdi@andrew.cmu.edu, (412) 268-2150, www.cmu.edu/student-diversity
Report-It online anonymous reporting platform: www.reportit.net username**:** *tartans* password**:** *plaid*

All reports will be acknowledged, documented and a determination will be made regarding a course of action." All experiences shared will be used to transform the campus climate.