Security Data Analytics

MISM Course F24-95-747 A2

Fall 2024

Carnegie Mellon University

Instructor: Dan Costa
Office hours: By appointment, email to schedule

Phone: 412-268-8006
E-mail: dlcosta@andrew.cmu.edu
Website: https://cmu.instructure.com/

Textbooks (CMU library-licensed resources)

- Chio, Freeman. *Machine Learning and Security*. O'Reilly Media, Inc. 2018 ISBN: 9781491979907 (CF)
- Bird, Klein, Loper. *Natural Language Processing with Python*. http://www.nltk.org/book/ (BKL)
- Scifo. *Hands-On Graph Analytics with Neo4j*. O'Reilly Media, Inc. 2020. ISBN: 9781839212611 (S)

Prerequisites and Requirements

Prerequisites:

- 95-796 (Statistics for IT Managers) or 90-711 (Statistical Reasoning with R)
- 95-752 (Introduction to Information Security Management)
- 90-812 (Introduction to Programming with Python) or 95-888 (Data Focused Python) or 95-898 (Introduction to Python)

Requirement: Students *MUST* have a computer with the ability to install Python and various Python libraries.

Course Description

Modern information security is full of big data problems. Discovering patterns and trends in cybersecurity incident data, detecting anomalous network or host traffic, predicting the likelihood of an email message containing malicious attachments or links – these are all examples of combining data sources and analytic techniques to preserve the confidentiality, integrity, and availability of information and information systems. In this course, we will cover analytic techniques such as clustering, classification, and anomaly detection, in the context of their applicability to the information security domain. We will explore the data sources that can be mined for security information. We will use hands-on labs to provide practical experience applying analytic techniques to these data sources. Finally, we will present strategies that can be used ensure the outputs of information security analytics are accurate, understandable, and actionable by security practitioners and business decision makers alike.

Learning Objectives

Learning Objective	**How Assessed**
Demonstrate competency in identifying effective combinations of data sources and analytic techniques that can be used to solve information security problems.	Final Project, Labs, Assignments
Demonstrate competency in applying analytic techniques to information security data sets to identify patterns and anomalies, make predictions, and support security practitioner decision making.	Final Project, Labs, Assignments
Demonstrate competency in measuring the effectiveness of information security analytics.	Final Project, Labs, Assignments

Schedule (subject to change during semester)

Date	Lecture / Lab	Readings / References
October 21	Course Introduction / Decisions and Data	CF Chapter 1
October 28	Classification and Clustering *Clustering and Classification Lab*	CF: Chapter 2
November 4	Anomaly Detection *Anomaly Detection Lab*	CF: Chapter 3
November 11	Text Analytics *Text Analytics Lab*	BKL: Chapters 0,1
November 18	Graph Analytics *Graph Analytics Lab*	S: Chapter 1
November 25	Deep Learning and Adversarial ML *Deep Learning Lab*	CF: Chapter 8
December 2	Security and Data at Scale	CF: Chapter 7

Labs / Assignments / Final Project

There will be in-class labs for weeks 2-6 of this class. There will be 2 assignments based on topics covered in lectures, additional assigned readings, and your work in the lab sessions. In addition, there will be a final project in which students will select a topic germane to security data analytics for independent research, submit a research project proposal for approval, and develop a 5-8-page report on their chosen topic. Following is a list of due dates for each assignment:

Item	Due Date
Lab 1 – Clustering and Classification	October 29 @ 11:59 PM EST
Lab 2 – Anomaly Detection	November 5 @ 11:59 PM EST
Homework 1	November 11 @ 11:59PM EST
Lab 3 – Text Analytics	November 12 @ 11:59PM EST
Final Project Proposal	November 11 @ 11:59PM EST
Lab 4 – Graph Analytics	November 19 @ 11:59 PM EST
Lab 5 – Deep Learning	November 26 @ 11:59 PM EST
Homework 2	December 2 @ 11:59PM EST
Final Project	December 6 @ 11:59 PM EST

Final Grade Evaluation Method

Labs: 30%
Homework: 30%
Final Project: 40%

Grading Scale

100 – 98	A+
97 – 92	A
91 – 90	A-
89 – 88	B+
87 – 82	B

81 – 80	B-
79 – 78	C+
77 – 72	C
71 – 70	C-
Below 70	R

Grade Distribution

I plan on using the Heinz School guidelines in deciding on the overall grade distribution. Accordingly, the average grade will be an A-. However, I grade on an absolute scale. If every student does well in the class, each will get an A+ regardless of the recommended grading scale. The same holds true on the other end of the scale. ***Students will only have 2 weeks after an assignment or exam is returned to question or challenge a grade.*** After the two-week challenge period, the grade will not be changed. Please contact the instructor if you wish to question a grade.

Late assignment policy

All assignments are due at 11:59 PM EST on the assigned due date. I WILL NOT accept late homework unless the student has made arrangements with me prior to the assignment's due date. PRIOR ARRANGEMENTS MUST BE MADE NO LATER THAN 12 PM ON THE DUE DATE.

Policy on cheating and plagiarism

This course follows Heinz School and Carnegie Mellon policies for student conduct, including policies that address inappropriate student collaboration and plagiarism. Each student is responsible for handing in their own work. For any assignment found to be the partial or complete result of cheating or plagiarism, your grade for that assignment will be zero. Cheating is defined as inappropriate collaboration among students on an assignment. This can include copying someone else's work with or without alteration. When students are found to be collaborating in this way, BOTH will pay the penalty regardless of who originated the work.

Accommodations for Students with Disabilities

If you have a disability and have an accommodations letter from the Disability Resources office, I encourage you to discuss your accommodations and needs with me as early in the semester as possible. I will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources, I encourage you to contact them at access@andrew.cmu.edu.

Student Wellness

Take care of yourself, and each other! 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.

How to Succeed in This Class

1. Get started on the final project as soon as possible. Even just spending some time in week 1 thinking about topics you may be interested in learning more about can make a significant difference.
2. Do not get stuck on syntax when working on labs and homework assignments. If you find yourself spending large amounts of time troubleshooting syntax errors for a query, reach out to me.
3. Complete the assigned readings for each week, if not before the class for which they are assigned, then after that class and before the next.
4. Use office hours productively. Get feedback on your final project, suggestions for references, tips on homework or lab questions, and clarifications of lecture materials. Make sure you are comfortable applying the concepts and lessons learned from course case studies to new requirements or use cases.