|  |  |
| --- | --- |
| **Module 1** | November 10, 2020, 7-9pm CT |
| **Topic:** | Foundations of Machine Learning and Data Science for Security |
| **Description:** | A module focused on machine learning fundamentals, with applications to security. This module will offer an introduction to the data science pipeline, and teach fundamental building blocks, from data ingestion and feature engineering to machine learning model selection. |
| **Faculty Leads:** | Yuxin Chen (Lead), Nick Feamster |
| **Asynchronous Content:** | * Pre-course survey to understand student background, familiarity with concepts, what problems/topics they are interested learning about. * Pre-recorded video lectures(~ 20- 45 mins -- to be watched ahead of sync) * Videos are broken up into shorter 5-10 min videos focusing on:   + Industry use cases (motivation for concept/topic)   + Fundamental concepts   + Applications   + Failure cases (real-life examples where the application of ML to security has failed and the dangers of such failures) * Post-video survey/quiz to check for understanding/provide students with opportunity to ask questions that can be answered synchronously in class |
| **Synchronous Content:** | * Group discussion of core concepts and how they relate to students experiences in industry/work * Case studies/group work using jupyter notebooks/dummy data to provide hands on experiences with concepts * Networking -- potentially pair students based on skills/background |

**Module 1 Learning Objectives & Module Outline:**

*Course description*

* A module focused on machine learning fundamentals, with applications to security. This module will offer an introduction to the data science pipeline, and teach fundamental building blocks, from data ingestion and feature engineering to machine learning model selection.

*Learning objectives*:

* Understand basic machine learning ideas & concepts
* Learn to use statistical tools to analyze machine learning models
* Understanding the role of machine learning in data-driven cybersecurity
* Have some experience applying machine learning algorithms on cybersecurity applications

*Asynchronous video outline*:

* (10 min) Introduction to the statistical learning framework; feature, label, risk and loss function.
* (5 min) Bias variance tradeoff, regularization
* (10 min) Introduction to supervised learning, model choices & failure cases;
  + logistic regression
  + neural network
  + Example: malware classification
* (10 min) Introduction to unsupervised anomaly detection
  + Generative/ discriminative models
  + Example: intrusion detection

**Module 1: Synchronous Schedule (via Zoom):**

|  |  |  |
| --- | --- | --- |
| **Topic** | **Time** | **Notes:** |
| Introduction/  Core Concepts Recap | 15 min | * Introduction * Recap of basic concepts in videos * Group discussion about experience with module concepts in industry/work experience |
| Introduction to case study/group work | 5 min |  |
| Simulation/  Case study/Group Work | 40 min | * TBD hands-on lab leveraging virtual case study (Jupyter Notebooks) * Divide students into Zoom breakout rooms * Pair students based on skill levels |
| Break | 10 min |  |
| Discussion/  Wrap up | 20 min |  |
| *Networking Opportunity* | *30 min* | * *TBD Virtual Happy Hour /speed meet-a-thon* * *Potentially led by class facilitator?* |