## 2.0 Introduction

When presenting, make sure to connect back to Yuxin and module 1—a lot of this is filling in items that were mentioned but not discussed (e.g. linear regression, clustering techniques beyond k-means), extending topics that were discussed (PCA), and turning to networking and cybersecurity proper.

The story arc of the module may also require a little explanation—the shift from Machine Learning to basic networking, and then back to machine learning.

**Slide 9: Learning Objectives**

Suggestion around reframing these as measurable outcomes:

Describe the basic architecture of the internet and the mechanics of networked communications

Acquire and represent data for modeling and analysis

Implement simple supervised machine-learning models for cybersecurity applications

Analyze network traffic using unsupervised learning techniques, such as PCA and clustering

## 2.1 Overview

**Slide 3: Learning Objectives**

Might take care to situate this overview in the context of the larger goal (ML for cybersecurity); maybe describe how this knowledge might help them address a larger problem or challenge in this space—where will this enter their pipeline?

Some rephrased objectives to start from:

Describe the basic infrastructure and architecture of the internet

Explain how packets are routed through networks to communicate information

Explain the role of domain name servers in routing traffic to IP addresses

## 2.2 Measurement

Some rephrased objectives to start from:

Describe different considerations and approaches to measuring internet traffic in a sound way

Collect network traffic using a variety of industry-standard tools and packages

*Represent the network data that has been collected in ways that can be inputted into machine learning models*

*Load the data into software libraries for analysis and modeling*

These objectives that I put in italics don’t actually seem to be addressed in this segment (I think they’re covered in the next one)—should they be excised?

**Slides 15, 16, 18, 19:** Need to reformat slightly to stop graphic from overlapping footer; it’s hard to read the black on maroon text

## 2.3 Data Representation

Some rephrased objectives to start from:

Describe the considerations and challenges around choosing how to represent network data for analysis and modeling

Load network traffic that has been collected into software libraries

Represent your network data in ways that can be inputted into machine learning models

## 2.4 Supervised Learning

Might reiterate here the connection back to module 1. No learning objectives given?

Is it worth mentioning polynomial basis functions here? These appear in the notebook

Would it be possible to conclude with some sketch of where or in what use cases or scenarios these linear models might be good or bad choices? Or why they might want to use these instead of tree-based models, or the deep-learning- based models that Yuxin described in module 1?

## 2.4 Unsupervised Learning

No learning objectives given?

Likewise, some concluding summary or overview about where they might find themselves using these different clustering techniques would be valuable

## Notebooks

Basic Analysis of Network Traffic Traces – both seem like nice, gentle intros. Might require a little cleaning up to keep them aligned with module content, but that will be light work.

Will there be examples corresponding to the content around PCA and clustering?