1. Problem Statement

- Develop a syllabus to study cyber threats and vulnerabilities using big data and secure analytics.
- Develop course materials, instructor notes, interactive videos and lab exercises that will inform and train students in applying learned skills to cyber defense and trusted analytics.
- Leverage current existing research work in big data, cyber security and data processing using trusted hardware extensions (TEE).

2. Course Materials

**Developed:**
- Machine Learning and Analytics
- Big Data Analytics
- Secure Big Data Analytics

**In progress:**
- Cyber Threats and Vulnerability Data Processing
- Actionable Insight from Cyber Threat and Vulnerable Data
- High Level knowledge Extraction and Kill Chain Inference

**Machine Learning and Analytics**

- Module focuses on different machine learning algorithms so that students get an understanding of the emerging techniques.
- Contains three set of PowerPoint slides for the lessons describing text mining, text feature extraction techniques and some efficient classification algorithms such as Support Vector Machines (SVM) and k-Nearest Neighbor Methods (KNN).
- Contains some lab tasks that give empirical knowledge on different classification and feature extraction processes.

**Big Data Analytics**

- Focuses on data mining & machine learning algorithms for analyzing very large amount of data with the help of tools/frameworks such as Hadoop MapReduce and Spark.
- Contains four modules:
  - Hadoop MapReduce Basics: contains three set of PowerPoint slides for the lessons describing the functionality and mechanisms of MapReduce framework.
  - Hadoop Setup and Programming: contains two set of PowerPoint slides explaining how to setup and program in MapReduce framework and some lab tasks that teach techniques to analyze massive amounts of data using MapReduce.
  - Spark Basics: contains three set of PowerPoint slides for the lessons that discuss Spark, Spark SQL and it’s advantage over Hadoop.
  - Spark Setup and Programming: contains two set of PowerPoint slides explaining how to setup and program in Spark framework and some lab tasks that show techniques for processing vast data using Spark.

**Secure Big Data Analytics**

- Module focuses on securing big data analytics by using the Intel Software Guard Extension (SGX) as a secure hardware solution.
- Contains the PowerPoint slides on the application of SGX in securing logging operations, the design of SGX-enabled log server, programs and protocols.
- Contains lab tasks that provide comprehensive knowledge to effectively run a new system designed to secure system logs using SGX-based log server.

3. Study Cyber Threat Reports and Vulnerabilities

**Learning Objectives**

- Allow students to gain actionable insights from cyber threat and vulnerable data repositories.
- Help students gain hands-on experience using techniques from NLP and Big Data Analytics.
- Prepare and instruct students in state-of-the-art skills needed to defend against attack techniques and challenges in software security.

**Sample Problem for Lab**

**Challenges:**

- Computer Security Threat reports are in unstructured text format.
- Large scale generation of threat reports due to increase in computer security breaches.
- Requires human analyst to interpret reports and extract meaningful information.
- Limited training data due to limited labeled threat reports.
- Need for automated system in extracting information and classifying threat reports to tactics and techniques

**Problem Setting:**

- Given a threat report, obtain the tactics and techniques used by the attacker as described in the report.
- After classifying to tactics, students need to identify the techniques used out of the available techniques under the tactics.
- For Technique classification, follow Top-Down approach: classify to tactics and then classify to techniques.
- Requires prunin the techniques under the tactics to identify relevant techniques.

**Proposed Solution**

- Download threat reports from different computer security vendors.
- Data Cleaning and Filtering to remove noise.
- Apply natural language processing techniques to extract features from unstructured text.
- Apply Classification method such as SVM classifier.

4. Materials Delivered So Far

**Machine Learning and Analytics Module Outline**

<table>
<thead>
<tr>
<th>Lectures</th>
<th>No. of Slides</th>
<th>Labs</th>
<th>Quiz</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Text Mining</td>
<td>26</td>
<td>2 Lab tasks</td>
<td>Assessment Quiz containing 13 questions</td>
</tr>
<tr>
<td>2. Text Feature Extraction</td>
<td>31</td>
<td>2 Lab tasks</td>
<td></td>
</tr>
<tr>
<td>3. Classification Algorithms</td>
<td>47</td>
<td>1 Lab task</td>
<td></td>
</tr>
</tbody>
</table>

**Big Data Analytics Unit Outline**

<table>
<thead>
<tr>
<th>Module</th>
<th>Lectures</th>
<th>No. of Slides</th>
<th>Labs/Qizzes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hadoop MapReduce Basics</td>
<td>1. Introduction to Hadoop MapReduce</td>
<td>56</td>
<td>Assessment Quiz containing 20 questions</td>
</tr>
<tr>
<td></td>
<td>2. MapReduce Algorithm Design</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Processing Relational Data with MapReduce</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Hadoop Setup and Programming</td>
<td>1. Hadoop Setup</td>
<td>21</td>
<td>4 Lab tasks</td>
</tr>
<tr>
<td></td>
<td>2. Hadoop Program</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>Spark Basics</td>
<td>1. Introduction to Spark</td>
<td>52</td>
<td>Assessment Quiz containing 10 questions</td>
</tr>
<tr>
<td></td>
<td>2. Programming with Spark</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Spark SQL</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Spark Setup and Programming</td>
<td>1. Spark Setup</td>
<td>9</td>
<td>4 Lab tasks</td>
</tr>
<tr>
<td></td>
<td>2. Spark Program</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

**Secure Big Data Analytics Module Outline**

<table>
<thead>
<tr>
<th>Lectures</th>
<th>No. of Slides</th>
<th>Labs</th>
<th>Quiz</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Installing and Running</td>
<td>N/A</td>
<td>1 Lab task</td>
<td>Assessment Quiz containing 8 questions</td>
</tr>
<tr>
<td>2. Understanding Theory</td>
<td>87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Programs and Protocols</td>
<td>27</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>