Problem Statement Worksheet (Hypothesis Formation)

Cyber Threat Detection: Train a machine learning model to identify and classify various types of cyber threats based on network traffic data and textual content.

H

1 Context

Using a dataset with a comprehensive collection of data for detecting, diagnosing, and mitigating cyber threats using network traffic data, textual content, and entity relationships. We will be training a machine learning model to identify various types of cyber threats, understand their underlying patterns, and recommend appropriate solutions.

2 Criteria for success

Accurately determining which textual content is a cyber threat or attack vector and which content is not.

3 Scope of solution space

Using the textual content transferred over the network, this column may contain descriptions of potential cyber threats or attack vectors along with other safe content, and its various relationships and diagnosis to identify cyber threat characteristics and patterns.

4 Constraints within solution space

Unable to train our model to detect other cyber threats or attack vectors which are not included in the dataset.

5 Stakeholders to provide key insight

- **IT Professionals**
- **Security Analysts**
- **Risk Management Exports**
- **Legal Advisors**
- **Senior Executives**

6 Key data sources

https://www.kaggle.com/datasets/ramoliyafenil/text-based-cyber-threat-detection

Citation: F. Ramoliya, R. Kakkar, R. Gupta, S. Tanwar and S. Agrawal, "SEAM: Deep Learning-based Secure Message Exchange Framework For Autonomous EVs," 2023 IEEE Globecom Workshops (GC Wkshps), Kuala Lumpur, Malaysia, 2023, pp. 80-85, doi: 10.1109/GCWkshps58843.2023.10465168.