**Title:** Actor Threat Profiling

Our Final Year Project (FYP) focuses on detecting and extracting Advanced Persistent Threats (APTs) through Actor Threat Profiling by using text mining and machine learning techniques. We aim to create a system that can identify and analyze APTs, which are complex, long-term cyberattacks planned by skilled threat actors. These attacks typically target specific organizations or industries to steal sensitive information or disrupt operations.

The main goal is to design and implement tools capable of recognizing patterns of APTs from network traffic and system logs. This involves developing algorithms that use text mining to filter through large volumes of data and machine learning to classify and recognize threat patterns. Text mining will process and analyze textual data from sources to uncover hidden threats. Machine learning algorithms will learn from this data to detect patterns and predict future threats.

By examining the behavior and strategies of threat actors, the system will provide detailed insights into their methods, origins, and potential targets. This involves analyzing various indicators of compromise (IOCs), such as unusual network activities, specific malware signatures, and phishing attempts, to build a profile of the attackers. Understanding these elements helps in predicting their moves.

The integration of text mining and machine learning will allow the system to improve its detection capabilities over time. As it encounters new data, it will adapt and become more effective at identifying new threats. This adaptive approach is essential for staying ahead of cybercriminals who constantly evolve their tactics to bypass traditional security measures.

Our aim is to strengthen organizational defenses against persistent and complex cyberattacks. By providing a detailed profiling mechanism, the system will help security teams understand the threats they face and develop effective countermeasures. This proactive defense strategy is crucial for protecting sensitive data and critical infrastructure, ensuring secure operations in a digital world.

**TimeLine:**

**Month 1: Data Extraction and Storage**

**Week 1-2:** Set up data extraction pipelines.(Graph DB or MySQL)

**Week 3:** Implement DOM parsing for extracting structured data from web documents.

**Week 4:** Validate extracted data and identify IOCs and IOC risk score.

**Tools Used**: Scrapy for data extraction.

**Platforms:** Web sources, forums, security blogs for APT-related data.

**Month 2: Name, Identity Extraction**

**Week 1-2**: Identify and Develop algorithms for name and identity extraction.

**Week 3:** Implement machine learning (ML) models or regex for pattern matching and extraction.

**Week 4:** Validate and refine extraction algorithms.

**Tools Used:** NLTK (Natural Language Toolkit), SpaCy for text processing, regex for pattern matching, sickit-learn and pandas .

**Platform**: \_\_\_\_\_\_\_\_\_\_\_\_\_

**Month 3: Text Mining**

**Week 1-3:** Design and implement text mining algorithms.

**Week 4:** Validate and optimize text mining processes.

**Tools Used:** TF-IDF (Term Frequency-Inverse Document Frequency), LDA (Latent Dirichlet Allocation) for topic modeling.

**Platform:** Textual data source (Crwalled)

**Month4 : APT Data Extraction**

**Week 1-3:** Develop algorithms for APT data extraction from network traffic and system logs.

**Week 4:** Validate and integrate APT data extraction processes.

**Tools Used:**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Platforms:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Month 5: Machine Learning for Zero-Day APT Detection**

**Week 1-3:** Train machine learning models for APT detection.

**Week 4:** Validate and fine-tune models for zero-day APT detection.

**Tools Used:** Scikit-learn, TensorFlow, or PyTorch for machine learning.

**Platform:** Textual and network data sources.

**Month 6: Integration and Finalization**

**Week 1-2:** Integrate all components into a comprehensive system.

**Week 3:** Conduct thorough testing and validation.

**Week 4:** Prepare documentation and final project presentation.