**CLASSIFYING TOR TRAFFIC ENCRYPTED PAYLOAD USING MACHINE LEARNING**

**ABSTRACT**

The Tor Traffic Classification Using Machine Learning project aims to develop an advanced framework for identifying and classifying encrypted Tor network traffic using machine learning techniques. With the increasing use of Tor for both privacy-focused activities and potential cyber threats, network security professionals face challenges in distinguishing between benign and malicious traffic without compromising user privacy. This project addresses these challenges by leveraging network metadata—such as source and destination ports, protocol type, flow duration, and inter-arrival times— as input features for classification.

This documentation provides a detailed analysis of network traffic characteristics, machine learning methodologies, and performance evaluation of different models. A thorough examination of network attributes enables the selection of key features that improve classification accuracy while maintaining non-invasive detection methods. The research implements and compares three machine learning models—Decision Tree, Logistic Regression, and XGBoost—to assess their effectiveness in identifying Tor traffic patterns. Model performance is evaluated based on predictive accuracy and computational efficiency, providing insights into the most suitable approach for real time encrypted traffic analysis.

The development process follows a structured machine learning pipeline, including data preprocessing, feature selection, model training, and evaluation, ensuring scalability and adaptability for future advancements in cybersecurity. By prioritizing privacy preserving classification techniques, this project contributes to the development of more secure network environments while respecting user anonymity. The findings offer valuable insights for network administrators and security professionals, enabling them to enhance threat detection mechanisms without compromising legitimate privacy focused use cases.

**KEYWORDS**: Decision Tree, Logistic Regression, XG Boost, Random Forest, Tor Traffic Classification.