**Real-Time Network Intrusion Detection Using Wireshark and Advanced Ensemble Learning Techniques**

**Project Domain / Category**

Networking/Machine Learning/ Research

**Abstract / Introduction**

With the rapid expansion of internet-connected devices, securing network infrastructures has become a major concern. Real-time network traffic monitoring and analysis are crucial to detect potential cyber threats and intrusions. This project focuses on developing an intrusion detection system (IDS) using real-time network traffic data captured by Wireshark. Students will utilize Wireshark to capture, preprocess, and extract relevant features from the network data. These features will then be used to train advanced ensemble learning models, including TabNet, CatBoost and LightGBM, to identify malicious network activities. The system will also include a web application that allows users to upload network traffic data, analyze it for intrusions, and display results in real-time.

**Functional Requirements:**

1. Students will use Wireshark to capture real-time network traffic data, exporting the captured data to a CSV file for further analysis. The capture will include features such as protocol types, IP addresses, packet sizes, time intervals, and more.
2. Students will preprocess the captured data by cleaning it, encoding categorical variables (e.g., protocol types), and normalizing numerical features (e.g., packet sizes, time intervals).
3. The processed data will then be structured into a CSV dataset, including labels for normal and malicious traffic activities, to be used for training the machine learning models.
4. Students will explore various machine learning classification techniques and select advanced ensemble models, including TabNet, CatBoost, LightGBM, and stacking ensembles. These models will be trained on the preprocessed network traffic dataset.
5. Model evaluation will be performed using state-of-the-art classification metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.
6. A web application will be built using Python Flask or Django, providing an interface where users can upload Wireshark-captured network traffic data for analysis. The application will process the uploaded data, detect potential intrusions using the trained ensemble models, and display the results, along with key network parameters that influenced the classification.

**Tools:**

* Programming Language: Python
* Traffic Capture Tool: Wireshark for capturing network traffic data in real-time and exporting it to a CSV file for analysis.

<https://www.wireshark.org/download.html>

* Machine Learning Libraries: Scikit-learn, XGBoost, LightGBM, pytorch-tabnet, CatBoost
* Web Development Frameworks: Python Flask or Django for building the web application
* Operating System: Any (e.g., Windows, Linux).
* Additional Tools: Jupyter Notebook for data analysis and model training, Matplotlib/Seaborn for data visualization.

**Supervisor:**

Name: Laraib Sana

Email ID: [Laraib.sana@vu.edu.pk](mailto:Laraib.sana@vu.edu.pk)

Skype ID: Laraib.sana