Network Intrusion Detection

DeadLine: 25 Tir 1402 Project

Project Description

In this machine learning project, the goal is to develop an effective network intrusion detection system using the UNSW-NB15 dataset. The project focuses on applying the techniques and methodologies learned during the course to achieve high performance in classifying network traffic. You're provided with an article[1], to compare your results with.
In evaluating your project, we will consider both the performance metrics you achieve and the sophistication of your approach and models. We encourage you to explore various models, techniques, and methodologies in order to deepen your understanding and showcase your expertise. Remember, the goal is not just to obtain good results using baseline or basic models, but rather to demonstrate your understanding of the underlying principles and apply more advanced techniques where appropriate.

Dataset Overview

UNSW-NB15 is a network intrusion dataset. This dataset is a widely used cybersecurity dataset that contains network traffic data captured in a controlled environment. It contains nine different attacks, including DoS, worms, Backdoor, and Fuzzers. All samples are labeled with one of the nine categories mentioned in the "attack_cat" column. The dataset comprises 42 features arranged in columns between the id(first column) and the attack cat(last column). The number of records in the training set is 175,341 and the testing set has 82,332 records from the different types, attack and normal.

Instructions

1. Data Exploration and Preprocessing:
    - Perform exploratory data analysis to gain insights into the dataset.
    - Handle missing values, if any, and preprocess the data for further analysis.
    - Visualize the distributions of different features and explore any patterns.
2. Feature Processing:
    - Apply feature processing techniques to enhance the quality and relevance of the features for the classification task.
    - This may involve techniques such as feature scaling, normalization, or transformation to ensure the features are in a suitable range or distribution for the models.
    - Consider exploring techniques such as dimensionality reduction to extract meaningful features.
3. Model Selection and Training:
    - Train and evaluate different models using appropriate evaluation metrics (e.g., accuracy, precision, recall, F1-score) on the dataset.

[1] IGRF-RFE: a hybrid feature selection method for MLP-based network intrusion detection on UNSW-NB15 dataset

- Utilize the models that are appropriate for this task and it is recommended to employ more advanced models.
- Tune hyperparameters to optimize model performance.

4. Performance Evaluation and Comparison:
   - Compare the performance of different models and techniques employed.
   - Analyze the strengths and weaknesses of each approach.
   - Report the achieved results and make sure to compare them with those presented in the paper.
5. Documentation and Reporting:
   - Document the entire project, including the steps taken, methods used, and code implementation.
   - Provide an analysis of the obtained results and compare them with the results presented in the paper.
   - If you manage to achieve better results compared to those documented in the paper, you're encouraged to reach out to the professor before commencing your report writing. This will enable you to collaborate with the professor to produce a comprehensive and detailed report.

Additional Guidance

- Make sure your code is in .ipynb format.
- Along with your code, please include a report file that thoroughly analyzes your results and compare them with the results presented in the paper.
- Use appropriate visualizations and statistics to support your analysis and conclusions.