




Intrusion Detection Based System (IDS)

Behnam Sobhani Nadri


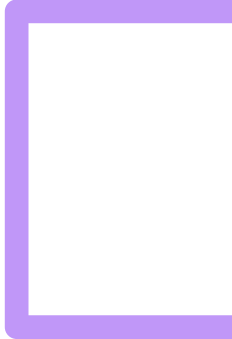


IDS

- Introduction
 - Methodology
 - Design
 - Results
 - Challenges
 - Conclusion
- 



Intrusion Detection System (IDS)

- An intrusion detection system investigates the connections in a network and analyze the data based on pretrained models to detect and classify malicious activities.
 - Connections can be of TCP/UDP, ICMP protocol and more. Other features also such as duration, protocol type and the data volume (bytes) are important to consider when there is a malicious activity in a network.
 - The objective of this project is to develop an IDS by using several ML models to improve the accuracy of detecting a malicious behavior by choosing the most optimum ML algorithm.
- 
- 



Methodology

A Comparative Analysis of ML Models

Machine Learning Approach in IDS

Machine Learning Models are powerful tools to analyze and classify the captured data from a Real or Simulated Environments. Some of those were applied to analyze the performance of and IDS.

- Logistic Regression (LR)
- Naive Bayes (NB)
- Support Vector Machine (SVM)
- Principal Component Analysis (PCA) : Dimensionality Reduction

Data preprocessing techniques helps us to clean and prepare a dataset for the ML model. Having a well-designed dataset can help the ML model to perform better and classify more accurately. Some of the preprocessing techniques are:

- Correlation Analysis
- Normalization and Standardization
- Feature Selection (Removing the unnecessary features)
- Cross Validation (75%-25% with random selection)

Results

- Logistic Regression:
95.04% Accuracy
- Support Vector Machine (SVM) :
 - Linear Kernel
Accuracy **95.5%**
 - Sigmoid kernel
Accuracy **90.34%**
 - Polynomial kernel
Accuracy **98.06 %**
- Naïve Bayes: **92.1%**
Accuracy

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression (LR)	95.04%	95.67%	93.58%	94.61%
SVM with polynomial kernel	98.06%	99.05%	96.75%	97.89%
SVM with linear kernel	95.42%	97.17%	92.86%	94.97%
SVM with sigmoid	90.34%	89.30%	90.03%	89.66%
Gaussian Naïve Bayes	92.12%	88.70%	95.18%	91.83%

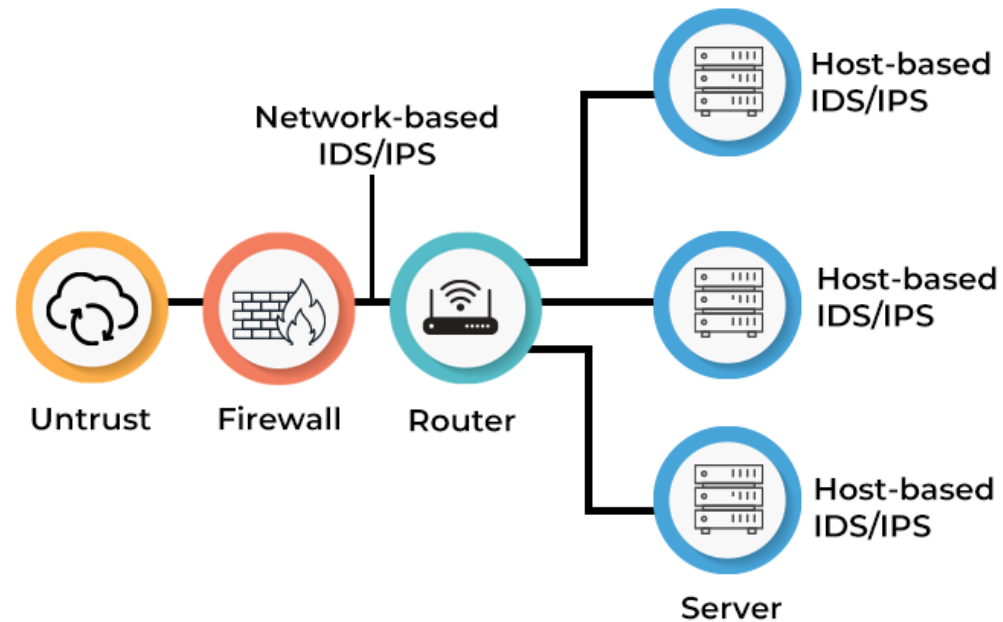
Results

After applying PCA combined with the Supervised Model:

- Logistic Regression: **95.04%** Accuracy (K=18)
- Support Vector Machine (SVM)
 - Linear kernel: **95.2%** Accuracy
 - Sigmoid kernel: **86.4%** Accuracy (K=18)
 - Polynomial kernel: **98.8%**
- Naïve Bayes: **92.1%** Accuracy (K=21)

Model with PCA	Accuracy	Precision	Recall	F1-score
Logistic Regression (LR)	95.04%	95.67%	93.58%	94.61%
SVM with polynomial kernel	98.80%	98.80%	98.62%	98.71%
SVM with linear kernel	95.20%	96.70%	92.78%	94.70%
SVM with sigmoid	86.44%	85.06%	85.70%	85.38%
Gaussian Naïve Bayes	92.12%	88.70%	95.18%	91.83%

Conclusion



- SVM performed the best classification to detect anomalies in the IDS dataset.
- Logistic Regression has demonstrated consistent results before and after applying the PCA
- Gaussian NB has the lowest accuracy among all three models
- PCA does not improve the classification accuracy



Thank you

Behnam Sobhani Nadri

bsobhani@charlotte.edu

github.com/behnamsn/IntroML