A Technical Evaluation Of The Performance Of Classical Artificial intelligence (AI) And Machine Learning Methods Based On Computational Intelligence (CI) i.e Supervised Learning And Ensemble In Intrusion Detection Systems.

Kudakwashe zvarevashe Harare Institute Of Technology

Presentation Outline

* Abstract
* Introduction
* Types of Intrusion Detection Systems
* Classical Artificial Intelligence Methods
* Machine Learning Methods
* Experiments and Results
* Conclusion

Abstract

* This research is a technical evaluation on some of the Artificial Neural Networks and Machine Learning algorithms when used to detect anomalies in intrusion detection.
* Experiments were done using a few algorithms and the NSL-KDD dataset.

Introduction

* Intrusion Detection is regarded as the second line of defense.
* It is usually done through the detection of unusual traffic which is eventually considered to be an anomaly and several other methods.
* The major classifications of intrusion detection systems are active and passive IDS.

Introduction(Continued)…..

* An active Intrusion Detection Systems (IDS) is sometimes referred to as an **Intrusion Detection and Prevention System (IDPS).**
* It is designed in such a way that malicious traffic will be dropped without the intervention of an operator.
* Intrusion Detection and Prevention System (IDPS) has the benefit of providing real-time remedial action in response to an attack.

Introduction(Continued)…..

* A passive IDS is a system that is configured to only monitor and analyze network traffic activity and alert an operator to potential vulnerabilities and attacks.
* A passive IDS will not perform any protective or corrective action on its own.

Types of Intrusion Detection Systems

* There are five main types of Intrusion Detection Systems:
  - Host Based
  - Network Based
  - Stack Based
  - Signature Based
  - Anomaly Based

Types of Intrusion Detection Systems(Continued….)

* **Host Based** -Intrusion Detection System is installed on a host in the network.
* **Network Based** -Network IDSs (NIDS) are placed in key areas of network infrastructure and monitors the traffic as it flows to other host.

Types of Intrusion Detection Systems(Continued….)

* **Stack Based IDS** -Stack based IDS is latest technology, which works by integrating closely with the TCP/IP stack, allowing packets to be watched as they traverse their way up the OSI layers.
* **Signature Based** -Signature-Based IDS use a rule set to identify intrusions by watching for patterns of events specific to known and documented attacks.

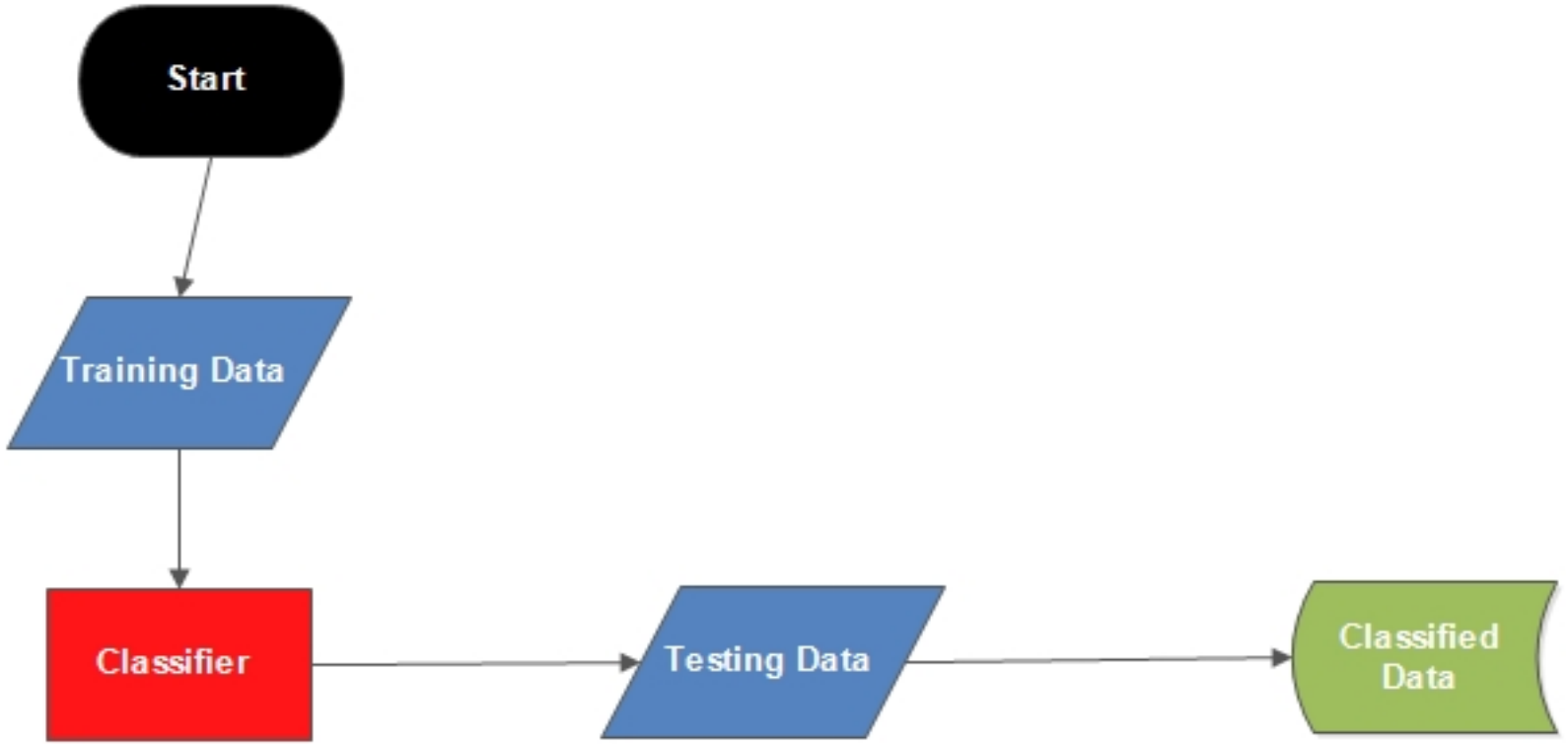
Types of Intrusion Detection Systems(Continued….)

* **Anomaly Based IDS** -Anomaly-Based IDS examines ongoing traffic, activity, transactions and behavior in order to identify intrusions by detecting anomalies.

Supervised Learning

* The algorithms consist of a target / outcome variable (or dependent variable) which is to be predicted from a given set of predictors (independent variables). Eg:
* Artificial Neural Networks
* Ensemble Methods

Classification Model



Classical Artificial Intelligence Methods

* Multi Layer Perceptron
* CHIRP
* Voted Perceptron

Multi Layer Perceptron

* Multi Layer perceptron (MLP) is a feedforward neural network with one or more layers between input and output layer.
* Feedforward means that data flows in one direction from input to output layer (forward).
* This type of network is trained with the backpropagation learning algorithm.
* MLPs are widely used for pattern classification, recognition, prediction and approximation.
* Multi Layer Perceptron can solve problems which are not linearly separable.

Chirp

* This classifier, called CHIRP, is an iterative sequence of three stages (projecting, binning, and covering) that are designed to deal with the curse of dimensionality, computational complexity, and nonlinear separability.

Machine Learning Algorithms

* Random Forest
* Real Adaboost
* Bagging
* SVM
* Naïve Bayes

Random Forest

* Random Forest is a trademark term for an ensemble of decision trees.
* In Random Forest, we've collection of decision trees (so known as "Forest").
* To classify a new object based on attributes, each tree gives a classification and we say the tree "votes" for that class.
* The forest chooses the classification having the most votes (over all the trees in the forest).

Naïve Bayes

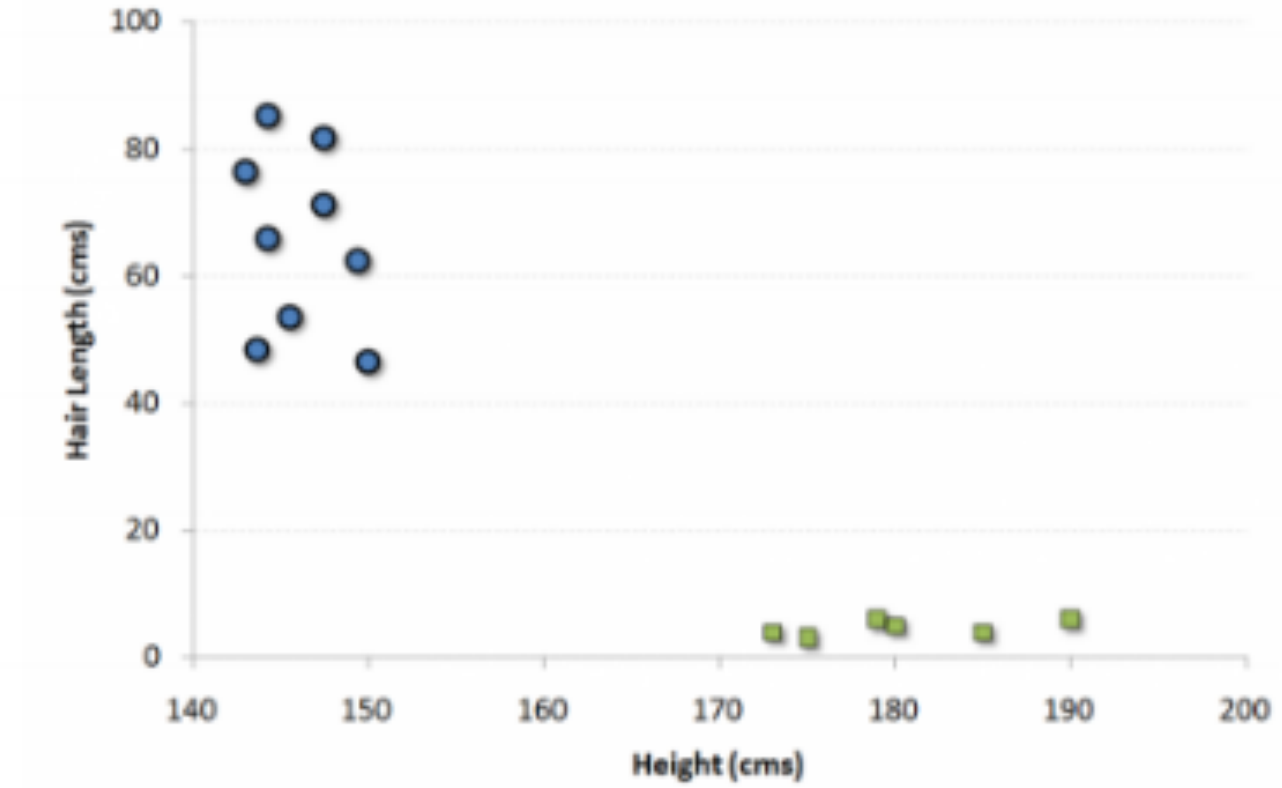
* It is a classification technique based on Bayes' Theorem with an assumption of independence between predictors.
* In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.
* For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter.
* Even if these features depend on each other or upon the existence of the other features, a Naive Bayes classifier would consider all of these properties to independently contribute to the probability that this fruit is an apple.

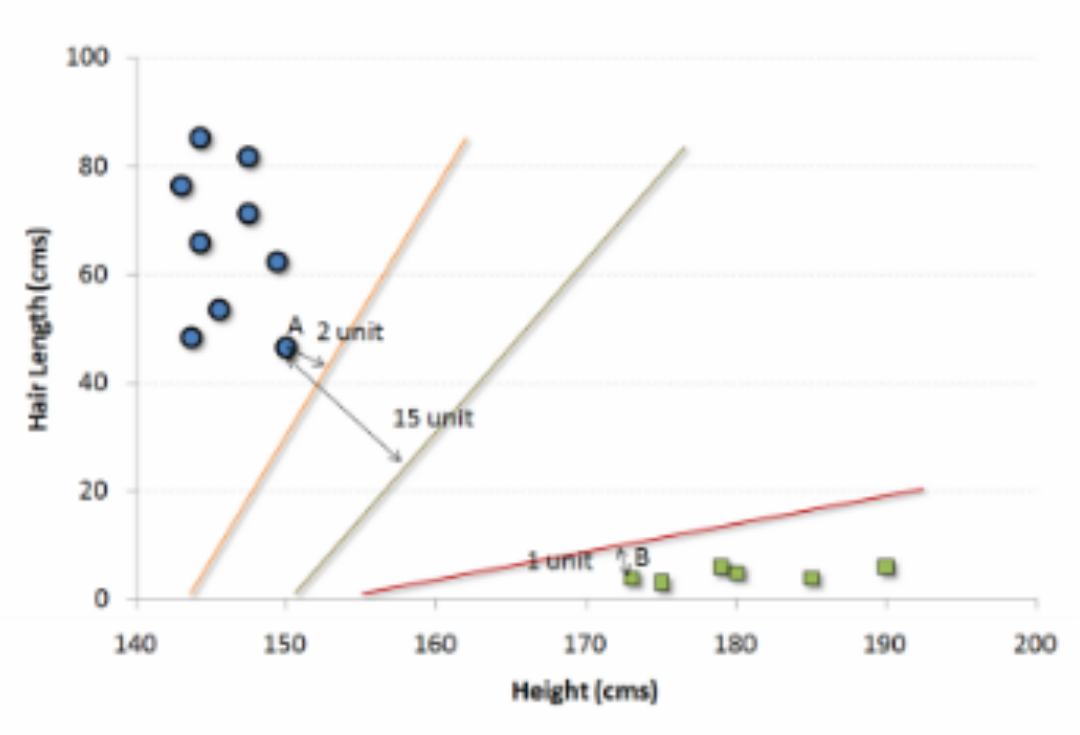
Real AdaBoost

* AdaBoost is a popular boosting technique which helps you combine multiple “weak classifiers” into a single “strong classifier”.
* A weak classifier is simply a classifier that performs poorly, but performs better than random guessing.
* A simple example might be classifying a person as male or female based on their height.
* You could say anyone over 5’ 9” is a male and anyone under that is a female.
* You’ll misclassify a lot of people that way, but your accuracy will still be greater than 50%.

SVM

* In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate.


100
80
60
40
20
0
Hair Length (cms)
140
150
160
170
180
190
200
Height (cms)


100
80
60
40
20
0
Hair Length (cms)
140
150
160
170
180
190
200
Height (cms)
A
2 unit
15 unit
1 unit
B

Experiments and Results

* All experiments were done using WEKA.
* The training dataset used was NSL-KDD dataset which contains **42829** instances with several types of attacks which include:
* DOS, Probe, R2L, U2R etc.
* The testing dataset consisted of **22544** instances

Algorithm	Build Time	Test Time	True Positives
MLP	7695.91s app 2.1 hrs	1.31	77.71
Chirp	429.72s app 7.15mins	0.49s	76.8808
Voted Perceptron	19.43s	54.11s	41.1595
Random Forest	80.39s	0.83s	80.4516
Real Adaboost	8.71s	0.14s	80.4205
Bagging	41.69s	0.15s	82.6295
SVM	21873.47s app 6hrs	396.58s app 6mins	72.3518
Naïve Bayes	0.79s	0.96s	76.1178
Ensemble Selection	106.33s app 1.7mins	4.02	84.3018

Conclusion

* The results of the experiments showed us that Ensemble Selection is a much better option compared to the other algorithms.
* This is because it had more true positives and had a higher precision and recall value.
* The Voted Perceptron yielded the worst results meaning that it might not be the best algorithm to use when it comes to intrusion detection.

Future Work

* We would like to analyse the performance of Immune algorithms.
* Use the winning Algorithm to design a prototype.

The End

* Thank you!