Analysis of Ensemble and Hybrid Approaches for Intrusion Detection

Intermediate Project Presentation

By
Abhinav Jaiswal (2016IPG-004)
Mohit Kumar (2016IPG-053)
Surendra Singh Gangwar (2016IPG-107)

June 20, 2019

Introduction

- In recent years, there has been a revolutionary change in the field of networking, use of internet is growing day by day and so is the risk of intrusion, Nowadays, it is very important to have a secure network because of increasing dependability on the internet and to achieve a high level security.
- One of the possible ways to counter this problem is through intrusion detection, which aims to identify various network attacks.
- Advancement in machine learning and deep learning made intrusion detection algorithms result in less error rate and more accurate to classify in less possible time.

Motivation

- The methods used for intrusion detection have proven to be advantageous but classification of different intrusion attack type efficiently is still a major concern
- Data-set used in the past was inconsistent,hence their is a need of consistent data-set to classify the attack accurately.
- Our main work is to combine the ability of different machine learning algorithm to accurately detect a particular attack type, and use them to build a model which is cable of predicting all attack type precisely. .

Literature Review

- Many different types of machine learning approaches has been implemented by researchers to classify the intrusion, some has used single learning algorithms such as Support Vector Machine(SVM) ,Logistic Regression(LR) , K-nearest Neighbors(KNN), Artificial Neural Network (ANN) and some uses multilevel techniques in which they first use some nature inspired algorithm and genetic algorithm to do feature selection and then used single classifier to predict the result.
- The challenges faced in Intrusion detection has been studied by many researchers around the globe .They suggested that their is need of efficient methodology which can identify any kind of intrusion attack precisely with the goal that a specific counter measure could be taken.

- Problem Statement
  - Analysis of ensemble and Hybrid Approaches for Intrusion Detection.
- Thesis Objective
  - Evaluating the performance of many different machine learning algorithms for the intrusion detection.
  - Implementation of mixture of experts technique for intrusion detection.
  - To perform a comparative analysis of ensemble approaches(Bagging,Boosting)with hybrid approach(Mixture of experts).

Methodology Used

- Observation of the KDD 1999 Dataset.
- Data cleaning and preprocessing.
- Division of entire dataset into five input space labeled with Normal, Dos, Probe, U2R, R2L.
- Evaluation of the performance of different base learners with the given dataset.
- Implementation of the Bagging, Boosting and Mixture of experts with the selected base learners.
- Comparative analysis of ensemble and hybrid approaches on the basis of accuracy score, precision, recall and F1-score.

Work Flow

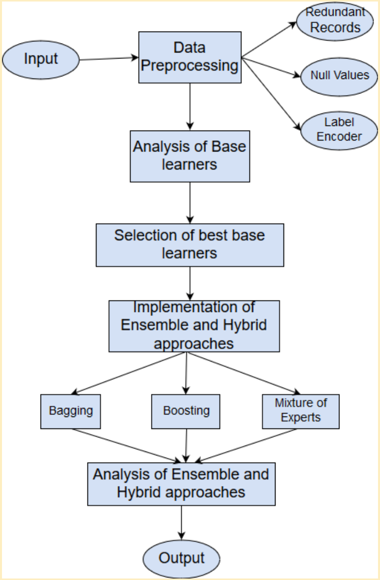

Input
Data Preprocessing
Redundant Records
Null Values
Label Encoder
Analysis of Base learners
Selection of best base learners
Implementation of Ensemble and Hybrid approaches
Bagging
Boosting
Mixture of Experts
Analysis of Ensemble and Hybrid approaches
Output

Figure: 1 Work Flow

Experimental setup

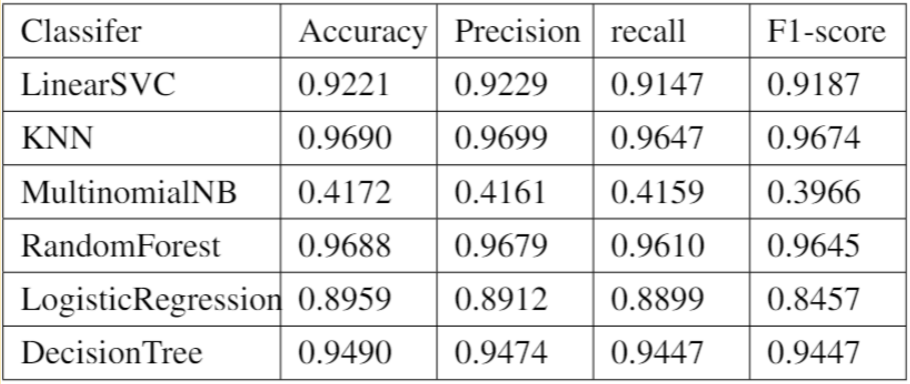
- Scikit-learn
- Training on NVIDIA Quadro P2000 GPU
- Implemented on jupyter notebook platform

Activities Completed

- Literature Survey
  - We have reviewed many research papers to get the knowledge of various work done in the field of intrusion detection and what are the research gaps which has to be covered.
- Data Cleaning and Preprocessing
  - Removing the data redundancy
  - Filling the missing values
  - Encoding the data
  - Analysis of different machine learning algorithms

Intermediate Results

- Intermediate results



Classifer	Accuracy	Precision	recall	F1-score
LinearSVC	0.9221	0.9229	0.9147	0.9187
KNN	0.9690	0.9699	0.9647	0.9674
MultinomialNB	0.4172	0.4161	0.4159	0.3966
RandomForest	0.9688	0.9679	0.9610	0.9645
LogisticRegression	0.8959	0.8912	0.8899	0.8457
DecisionTree	0.9490	0.9474	0.9447	0.9447

Figure: 2 Intermediate Results

Contd...

- Classifier for future Work
  - K-nearest Neighbor, Random Forest Classfier and Decision Tree Classifier got the highest accuracy .
  - These three will be used as the base learners in the model .

Future Activities

- Dimension Reduction
  - We will reduce the dimensionality of the dataset with feature selection approaches like PCA(Principal Component Analysis).
- Implementation of selected base learners with hybrid approaches
  - Implementation of the best selected base learners with the proposed hybrid approach mixture of experts.
- Analysis of final results
  - We will analyse the final results and finding the best method for the intrusion detection on the basis of accuracy score, precise, recall and F1-Score.

Contd...

- Web Application
  - Detect the type of attack generated dynamically on the system using the Intrusion Tool and web application.

Thank You