Intrusion Detection System on Network

By

Arun POOVELIKUNNEL VARGHESE

Introduction

The intrusion detection system design based on the data of DDoS attack. The origin of the data from the University of New Brunswick server logs used to analyze DDoS. The python notebook will give how the model is developed, trained and evaluated. And for easy understanding of this IDS the python notebook divided into different sections. Loading of Data, Data Processing, Scaling of Attributes, Feature Selection, Data Splitting, Fitting Model, Validating Model and Evaluating Model these are different steps of the intrusion detection system. In this data the Label column which will differentiate the data which comes under what type of DDoS.
Labels:
0 - **Benign**
1- **Bot**

Intrusion Detection System

Loading of Data:
This stage the data in csv format has been loaded to the notebook with the help of the read_csv function of pandas library. Now from the data we need to identify the number of types of labels there are in the DDoS data, so it can help in data processing.

Data Processing:
In this stage the data is cleaned and segregated based on the different types of DDoS so the one attack won't have more influence than the other in the model. Cleaning is mainly removing NaN and infinite values from the dataset.

Scaling of Attributes:
Here scaling is done as two sets, one is from the numerical attributes and scale it to have zero mean and unit variance and the categorical attributes encoded and separated to do analysis with the sample set.

Feature Selection:
Feature Selection is done with the random classifier by giving a score to each feature based on importance of the feature and from that list which most influencing 20 features has been selected. Then the new sample data is extracted based on the selected features from scaled data.

Data Splitting:
Now the splitting of data is done as taking the ratio of 70:30, this means that 70 percent of data will be used for training models and cross validating the trained model and 30 percent of data in testing of the designed model (evaluating of model).

Fitting Model:
Here the train data is fitted to the classifier to create the model, just creating the model is not only the part needed to cross validate and see how the model is good against the DDoS attacks.

Validating Model:
This stage we validate the created model with the same train data used for the creation of the model.The cross-validation mean will give the score of how good the model, the recall is better over the fist model than second model then choose the first model. If it is same then check precision score

Evaluating Model:
This stage the validated model is tested and evaluated based on the test data by confirming the accuracy of the model towards the DDoS attack. This is also evaluated based on the recall is better over the fist model than second model then choose the first model. If it is the same then check the precision score.

Conclusion
In this project the LogisticRegression Model is better over the Naive Baye Model because the recall score is better for the lgr_model than bnb_model. And even the precision is also better for the lgr_model than bnb_model.