MACHINE LEARNING CYBERSECURITY

INTRUSION DETECTION

WRITING A CLASSIFIER FOR KDD99 DATASET

Description: Implement a binary classifier to distinguish normal connections from attacks. You are required to read the data from the training set (2,799 records) and test set (1199 records).

You are required to implement two parts:

- Writing a python script with the use of the package sklearn
- Writing a python script with the use of the package tensorflow and deep learning techniques.

Environment:

- Linux system or VM
- Python is required as well as some packages such as numpy, tensorflow and sklearn.

Files that are Needed:

- For this project you will need 2 files
- train_kdd_small.csv and test_kdd_small.csv for python script.

PART 1

- You need to implement several classifiers with the use of sklearn.
- Import the required libraries
- Read the features and class values from training set and test set with proper method
- Ensure that all features are in a datatype that your classification algorithm can operate on.

- You may need to create labels for each of kdd classes. When you finish the preprocess step, you can write the python script with the use of sklearn package to build your architecture of classifier.
- Print the statistics metrics such as accuracy, recall, precision and f1 score.
- Implement the classifiers based on Logistic Regression, Support Vector Machine and Random Forest

PART 2

- Use the same data you use in Part 1.
- In this part, you will implement an artificial neural network classifier based on Tensorflow
- Import the required libraries
- Repeat the same steps to preprocess the data as Part 1. Read the data, standard scale the feature and encode the labels.
- Define the learning rate, number of epochs and batch size for artificial neural network
- An extra step in preprocess is to perform the one-hot encoding for the labels.
- Define the parameters to store the shape of placeholder.
- Define the function to draw the plot of performance
- Define your own architecture of neural network
- Print the statistics metrics such as accuracy, recall, precision and f1 score.
- Initialize the variables and placeholders. Then perform the training and testing on subset of kdd dataset.

WHAT TO SUBMIT

Submit a project report which includes:

- See the project rubric on Canvas