Name : Maatrika P
UGID : 21WU0102056

SMA Project Report

Intrusion Detection System

(INTERNAL PROJECT)

Submitted to : Dr. Punit Singh

Done By :
Name: Maatrika P
UGID: 21WU0102056
Specialization: DSAI

Name : Maatrika P
UGID : 21WU0102056

Problem Statement:

The challenge is to develop an effective intrusion detection system (IDS) capable of identifying and mitigating unauthorized and malicious activities within computer networks. With the increasing frequency and sophistication of cyberattacks, there is a pressing need for an IDS that can distinguish between normal network traffic and potential threats, safeguarding sensitive data and critical infrastructure.

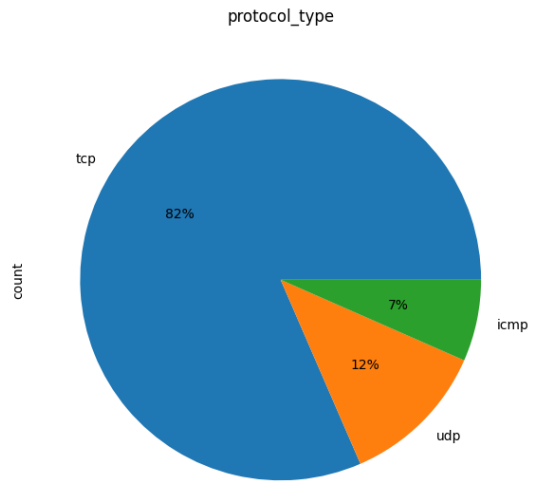
The problem encompasses the design, training, and deployment of a robust IDS that can operate in real-time, adapt to evolving threats, and minimize false positives while maximizing detection accuracy, ultimately enhancing the security and integrity of networked systems.

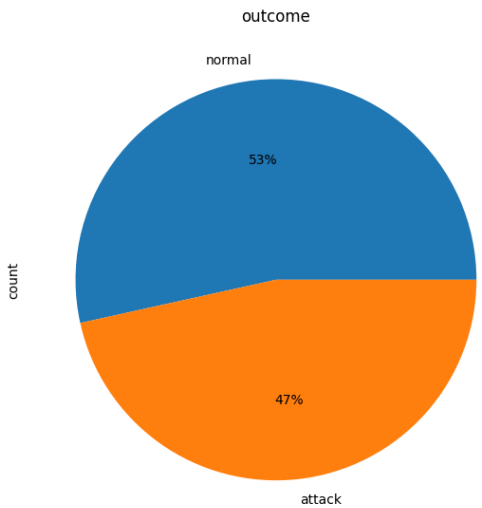
Dataset utilized :

NSL-KDD: https://www.kaggle.com/datasets/hassan06/nslkdd/data

The NSL-KDD dataset is an improved version of the KDD'99 dataset used for intrusion detection system evaluation. It offers several advantages over the original KDD dataset, including the removal of redundant and duplicate records, ensuring unbiased classifier training and evaluation.

Records are selected based on difficulty level and proportionality, allowing a wider range of classification rates for accurate assessment of learning techniques.





Name : Maatrika P
UGID : 21WU0102056

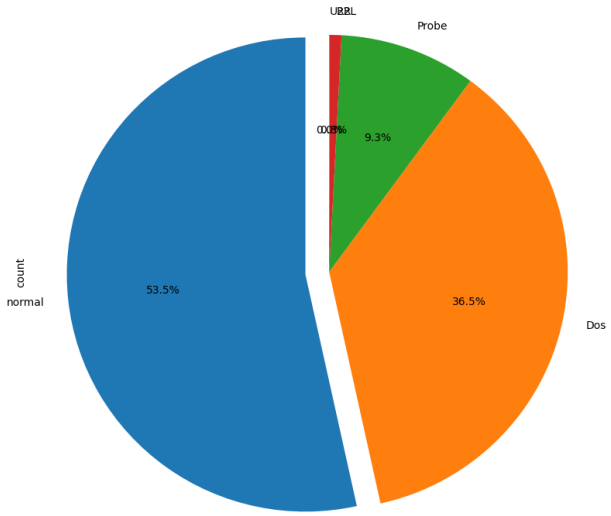
There are four kinds of attacks in the dataset:

-> DOS

-> Probe

-> R2L

-> U2R



Models Used :

CNN - The following is the model architecture and the accuracy scores

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv1d_2 (Conv1D)           (None, 122, 32)           128

 max_pooling1d_2 (MaxPoolin  (None, 30, 32)            0
 g1D)

 dropout_2 (Dropout)         (None, 30, 32)            0

 conv1d_3 (Conv1D)           (None, 30, 32)            3104

 max_pooling1d_3 (MaxPoolin  (None, 7, 32)             0
 g1D)

 dropout_3 (Dropout)         (None, 7, 32)             0

 flatten_1 (Flatten)         (None, 224)               0

 dense_2 (Dense)             (None, 50)                11250

 dense_3 (Dense)             (None, 5)                 255

=================================================================
Total params: 14737 (57.57 KB)
Trainable params: 14737 (57.57 KB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

```
Accuracy: 0.9888073030363167
Precision: 0.9885610142429992
Recall: 0.9888073030363167
```

Name : Maatrika P
UGID : 21WU0102056

Input Layer: It takes input data with the shape (X_train.shape[1], 1). This indicates that it expects one-dimensional data.

Convolutional Layer: model starts with a convolutional layer. It uses the ReLU activation function.

MaxPooling Layer: After each convolutional layer, a max-pooling layer with a pool size of 4 is applied to downsample the data.

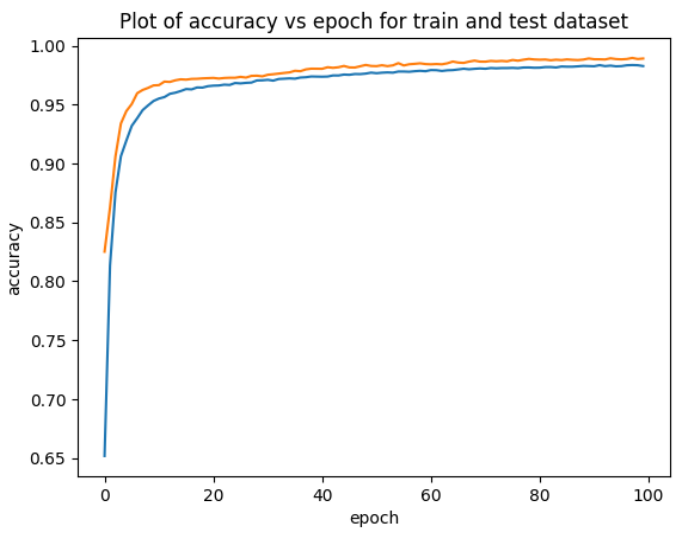
Dropout Layers: Dropout layers are added to prevent overfitting(randomly deactivate neurons)

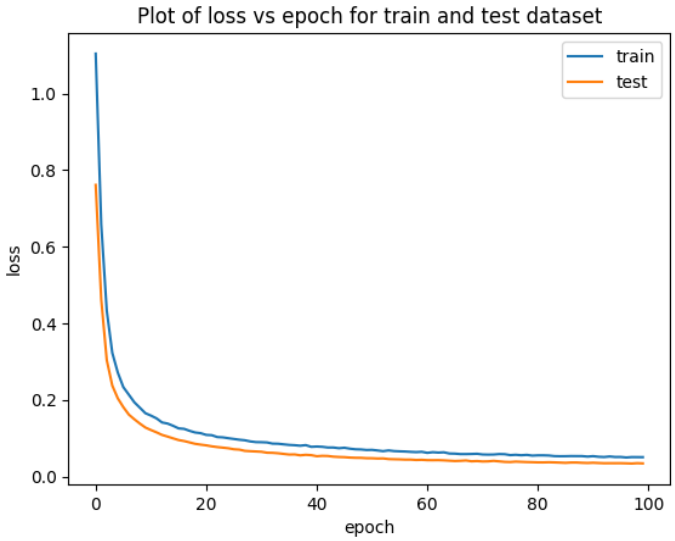
Flatten Layer: The flatten layer transforms the multidimensional data into a one-dimensional vector.

Dense Layer: This is a fully connected layer with 50 neurons and ReLU activation.

Output Layer: The output layer has 5 units, designed for multi-class classification. It uses the softmax activation function to provide class probabilities.

The model was trained for 100 epochs with a batch_size of 5000, and a validation split of 20%





Name : Maatrika P
UGID : 21WU0102056

LSTM - The following is the model architecture and the accuracy scores

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 20, 64)            16896

 dropout (Dropout)           (None, 20, 64)            0

 lstm_1 (LSTM)               (None, 20, 64)            33024

 lstm_2 (LSTM)               (None, 32)                12416

 dropout_1 (Dropout)         (None, 32)                0

 dense (Dense)               (None, 16)                528

 dense_1 (Dense)             (None, 1)                 17

=================================================================
Total params: 62881 (245.63 KB)
Trainable params: 62881 (245.63 KB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

```
Accuracy: 0.9932526294899782
Precision: 0.9932828841084942
Recall: 0.9922704493332201
F1 Score: 0.992776408600323
```

Input Layer: Receives one-dimensional sequential data.

LSTM Layers: Three layers are used to model sequential data, each with 64 units. These layers capture dependencies in the data over time.

Dropout Layers: Two dropout layers are added with a 20% dropout rate to prevent overfitting.

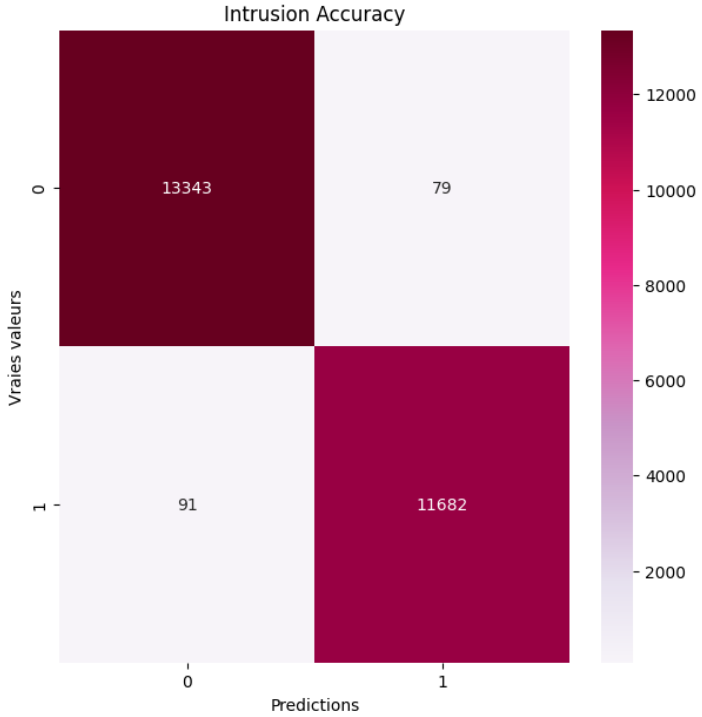
Flatten Layer: Transforms the data from the LSTM layers into a one-dimensional format.

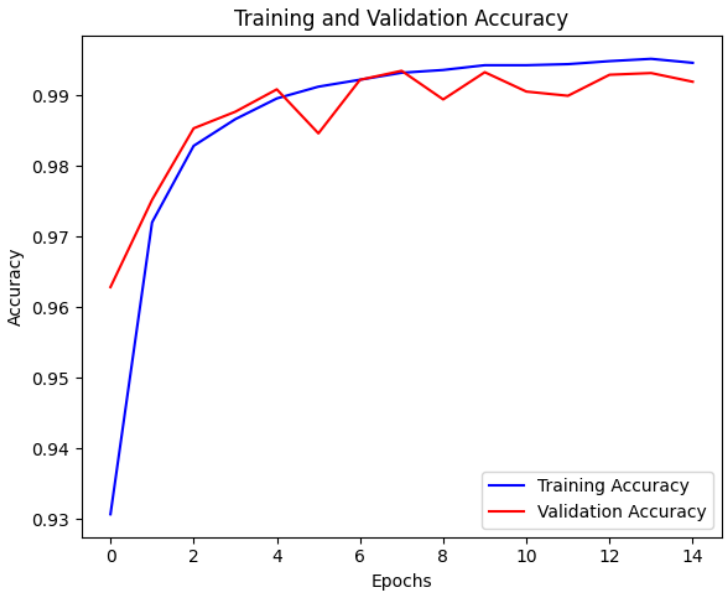
Dense Layer: The dense (fully connected) layers with 50 units and ReLU activation functions for further feature processing.

Output Layer: 5 units for multi-class classification and uses softmax activation to produce class probabilities.

The model was trained for 20 epochs with a batch_size of 32 with early stopping callback to prevent overfitting.

Name : Maatrika P
UGID : 21WU0102056





Name : Maatrika P
UGID : 21WU0102056

LSTM + CNN - The following is the model architecture and the accuracy scores

```
Model: "sequential_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv1d_1 (Conv1D)           (None, 18, 64)            256

 max_pooling1d_1 (MaxPoolin  (None, 9, 64)             0
 g1D)

 lstm_6 (LSTM)               (None, 9, 64)             33024

 dropout_4 (Dropout)         (None, 9, 64)             0

 lstm_7 (LSTM)               (None, 9, 64)             33024

 lstm_8 (LSTM)               (None, 32)                12416

 dense_4 (Dense)             (None, 16)                528

 dense_5 (Dense)             (None, 1)                 17

=================================================================
Total params: 79265 (309.63 KB)
Trainable params: 79265 (309.63 KB)
Non-trainable params: 0 (0.00 Byte)
```

```
Accuracy: 0.9944433419329232
Precision: 0.995907579503794
Recall: 0.9921855092160027
F1 Score: 0.9940430601650923
```

Input Layer: Receives one-dimensional data.

Convolutional Layer: Utilizes 64 filters with a kernel size of 3 and ReLU activation to extract spatial features.

MaxPooling Layer: Max-pooling layer with a pool size of 2 downsampled the data.

LSTM Layers: Three Long Short-Term Memory (LSTM) layers are used to model sequential data, each with 64 units and returning sequences.

Dropout Layers: Dropout layers with a 20% dropout rate are employed to prevent overfitting.

Dense Layer: A fully connected layer with 16 units and ReLU activation to process features.

Output Layer: The output layer with 1 unit and sigmoid activation for binary classification.

The model was trained for 20 epochs with a batch_size of 32 with early stopping callback to prevent overfitting.

Name : Maatrika P
UGID : 21WU0102056