Executive Summary

Overview

This report provides an evaluation of the implemented neural network model that detects and classifies intrusion connections. The analysis is conducted using charts generated in the testing process.

Problem

The intrusion detection problem was the main objective of KDD Cup 1999. A subset of the original dataset containing ten percent of the original data is the pre-requisite for the IMAT5235 Module Coursework. It contains 42 columns and 494021 rows. The column names are the names of the connection features with the last column being the output indicating either a standard, safe connection or an intrusion of various kinds.

Solution

The intrusion detection problem is solved using an artificial neural network with 3 hidden layers and various parameters. The parameters and the number of hidden layers as well as the net's model type are selected in an experimental way. The process of implementation and testing can be found in the presentation that is submitted along with this paper. Results of the experiments indicate that the network's architecture is suitable for the problem and can operate on the dataset efficiently. The preprocessing of data using Principal Component Analysis and MinMax Scaler help transform the input data and make it adequate for the network, while encoding the output enables proper classification. Further changes in batch size and number of components in PCA result in further improvement of the model's accuracy and make it more time efficient. The addition of a monitoring function (which stops the training process whenever no improvement on the validation error is detected) resolves the issue of uncertainty about the number of epochs.

Conclusions

The proposed solution provides satisfactory results. The implemented neural network is able to solve the problem formerly mentioned in this paper. Based on the tests, the parameters and their values are optimal however, the experimental approach to the implementation makes it impossible to determine whether they are generally optimal.

Future work recommendations

Given the experimental approach to solving the problem however, certain fields for improvements are identified and these include:

- changing the type of output encoding so the net not only recognizes the intrusion but is also able to determine its type,
- reducing the dataset by removing all continuous data from it.