MIE1628 Assignment 3 Answers

Part A

Q1: What is an Intrusion Detection System? Is it possible to implement an

Intrusion Detection System on this dataset? Explain the workflow as described in the

paper for implementing Intrusion Detection System.

Intrusion detection system (IDS) is a system that monitors and analyzes data to detect any intrusion in the system or network.

10% subset of KDD Cup 99 data has 41 different features and 494021 samples. Intrusion Detection System can be implemented on this dataset, but it need to use Big Data tools(Spark) and techniques(such as SVM) to analyze and store data to reduce computation and training time.

The workflow in this paper can be split into three parts. First, loading and preprocessing the data by change the type of data, encoding, and standardize the features. Second, select the features which are highly relevant to the target and get rid of the irrelevant features by Chi Squared test. This can help reduce the processing time of model training. Third, they split the data into a training set and a test set. They use the training set to train the SVM model and check the performance of the model by the prediction error of the model on the test set.