Deep Science Publishing
https://doi.org/10.70593/978-93-49307-78-0_4

DeepScience
Open Access Books

Chapter 4

Performance metrics analysis: Evaluating machine learning models in detecting DDoS attacks

4.0 Performance Analysis Result

4.1 Overview of Performance Analysis Results

Tables 10 and 11 show the performance metric result of this experiment. The metrics used to evaluate the machine learning models are accuracy, precision, recall, f1-score, and computation time on new datasets and open-source datasets as shown in Tables 10 and 11. An 80:20 split of the overall dataset was used for the Model building where 80% was used for training and 20% was used for validation and testing. The objective of the evaluation is to assess the effectiveness of DDoS datasets in terms of their ability to detect DDoS attacks in the cloud system. The results demonstrate that the new dataset performed very well, achieving a 100% accuracy rate with Random Forest, SVM, Decision tree, and KNN, and then 98% with Naïve Bayes. An F1 score (measure) of 100% on all models except NB with 98%. This indicates that a model trained with a newly generated dataset performs very well, as it correctly predicts threats (precision) and captures all relevant cases of malicious traffic (recall) at a 100% rate in RF, DT, and KNN, 99% on SVM and NB with a 97% rate.

On the other hand, the CSE-CIC-IDS2018 dataset was used for training a larger sample size since time constrain could not allow the generation of such a sample size in the new dataset. The RF, DT and KNN achieved 100% accuracy followed by NB at 99% and the SVM performs the lowest with an accuracy of 95%. as shown in table 11. It depicts that the CSE-CIC-IDS2018 dataset is a very good data source for the detection of DDoS on a cloud network.

Additionally, the results achieved from the two datasets mean that the model performs well in the detection of threats as well as just missed detection of very few instances. The confusion matrix will be used use visualize the exact instances. All the models selected performed very well, though some performed better than others.

Furthermore, with the new dataset, all models took a shorter time like a minimum of a second and a maximum of 10 seconds to train in terms of computing time. This is because the dataset has a small data volume (28972 rows). On the other hand, The CSE-CIC-IDS2018 dataset had the longest computation times, with some models like SVM having the longest computation time 3751 seconds Followed by k-NN with 199 seconds The fact that this set had 300967 records is significant. On the new dataset, k-NN took the longest time for both training and validation, while on the CSE-CIC-IDS2018 dataset SVM took the longest time as shown in Table 11. The results of the new dataset on the performance metrics are shown in Table 10.

This chapter was concluded with some reflections to measure if they were carried out and achieved as stated in the investigation report.

4.2 New Dataset Performance metrics

The main performance metrics for this dataset are accuracy, precision-recall and F1 score, computation time was added to note how long it takes each model to produce a result.

Table 7: Performance metrics for the new dataset

Models	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**Computation time (Seconds)**
Random Forest	100.0	100.0	100.0	100.0	1.4
SVM	100.0	100.0	99.0	100.0	1.1
Decision Tree	100.0	100.0	100.0	100.0	0.1
Naive Bayes	98.1	97.0	99.0	98.0	0.1

K-NN	100.0	100.0	100.0	100.0	1.9

As shown in Table 10, the new dataset performs remarkably well, achieving a 98% and above accuracy rate for all machine learning models used in this research. This can be attributed to the 31 flow features extracted from the datasets and selected features used to train the models. All models took a shorter time to train due to the small record volume (28972 rows). However, KNN took the longest time 1.9 seconds. Figure 31 shows the accuracy results of the new dataset illustrating that NB performs the least with 98%. Figure 30 shows the performance metrics result of the new dataset.

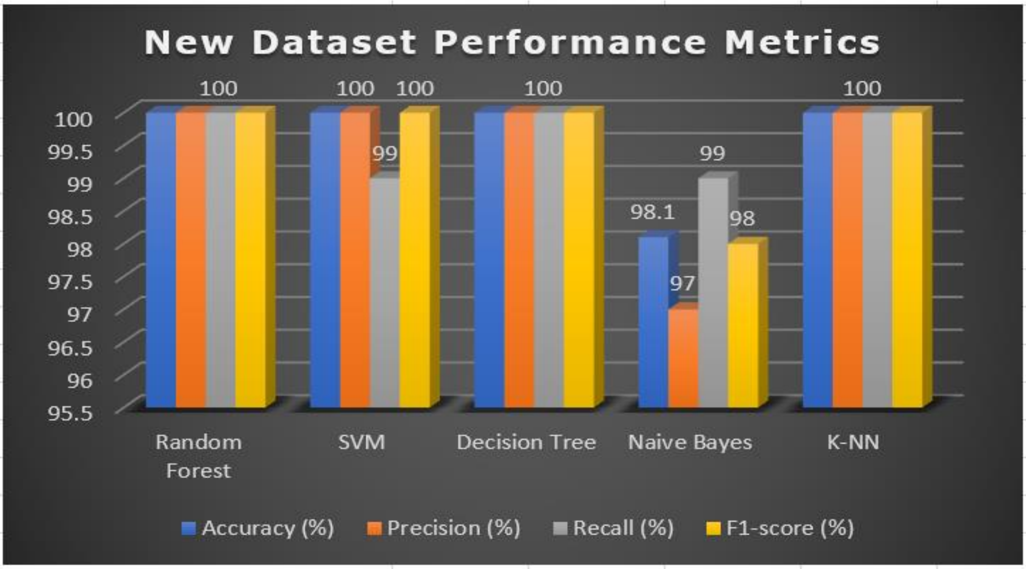

Figure 30: New dataset performance metrics result.

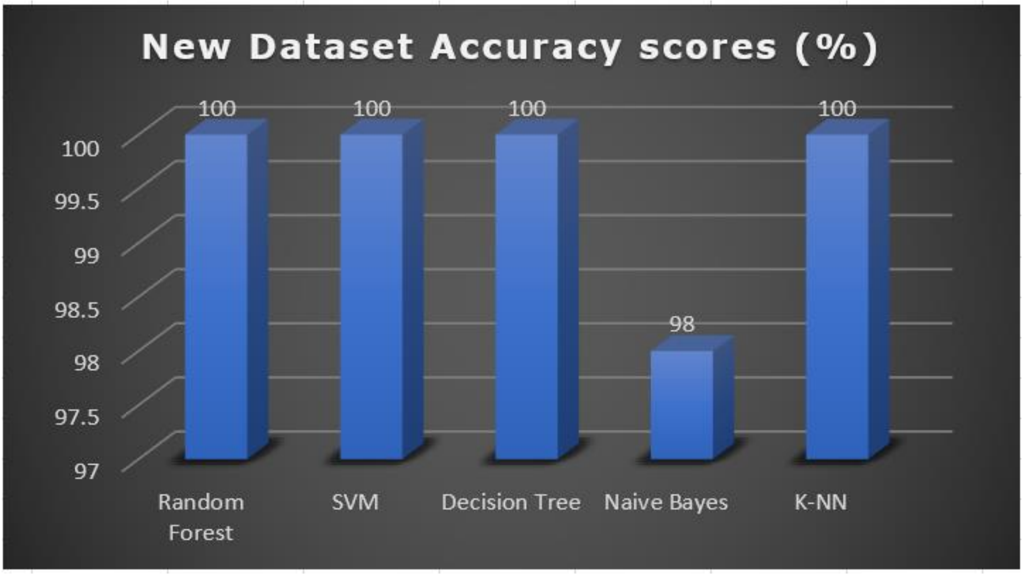

Figure 31: Model accuracy result of the new dataset.

In conclusion, the most relevant performed models with the new dataset as shown in Figure 31 are RF, SVM DT and K-NN. Though NB is relevant but not the most because of the accuracy score.

4.3 Open-Source Dataset Performance Metrics

Table 8: Performance metrics for the CSE-CIC-IDS2018 dataset

Models	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**Computation time (seconds)**
Random Forest	100	100	100	100	9.6
SVM	95	100	90	94	3751.4
Decision Tree	100	100	100	100	0.8
Naive Bayes	99	99	99	99	0.6
K-NN	100	100	100	100	199.8

As shown in table 11, SVM performs the lowest in accuracy with 95% as also shown in Figure 33 with the highest computation time (3751 seconds). This can be attributed to the dataset's larger record count (300967rows) and features selected which may not be compatible with the SVM but works well with other algorithms. However, K-NN still performs almost accurately at 100% though the training time was long 199.8 seconds. Figure 32 displays the performance metrics results on the accuracy, precision, recall, and F1 score.

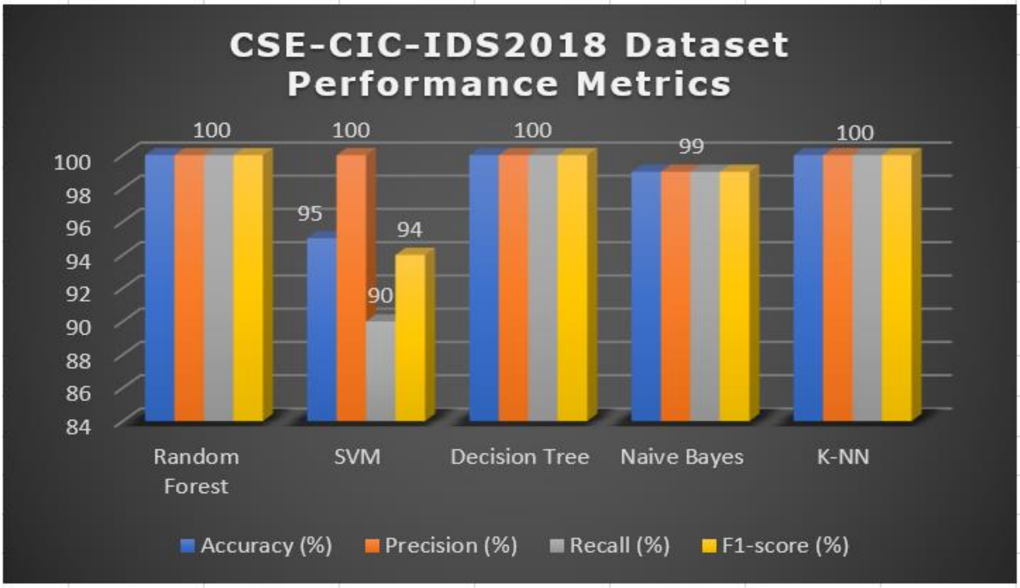

Figure 30: Model Performance metrics for the CSE-CIC-IDS2018 dataset.

In conclusion, the most relevant performed models with the open-source dataset as shown in Figure 33 are RF, DT and K-NN. Though NB is relevant but not the most.

4.4 Confusion matrix for the datasets

With features selected to train the models, the results display true positive rate, true negative rate, false positive rate, and false positive rate. The confusion matrix outcome is displayed in Table 12 to depict the miss classification rate for all algorithms. Improved accuracy is reflected by low miss classification.

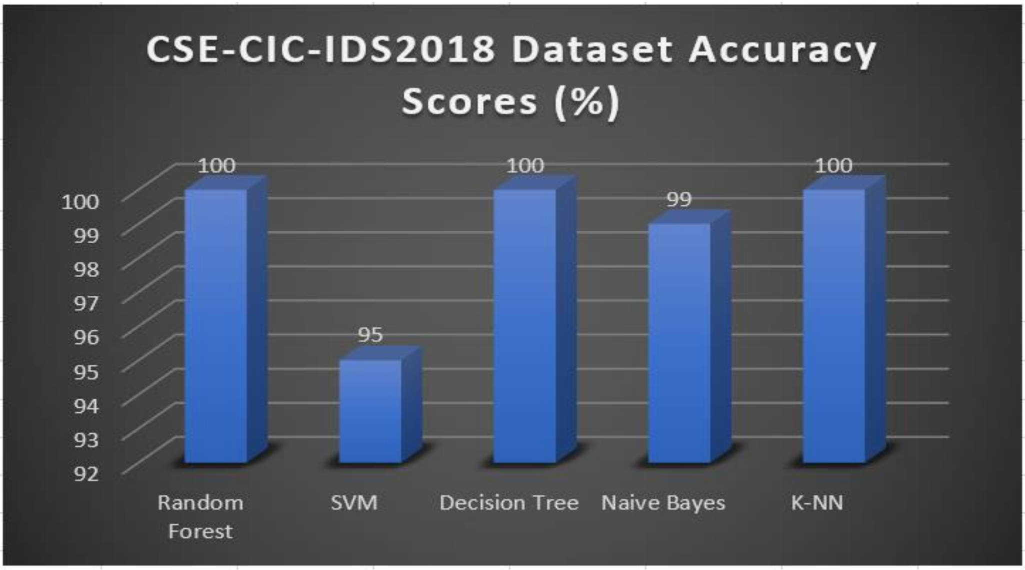

Figure 31: CSE-CSE-IDS2018 dataset accuracy result.

Table 9: Confusion matrix of the new dataset

Model	True Positive	False Positive	True Negative	False Negative
Random Forest	2873	0	2922	0
SVM	2867	6	2907	15
Decision Tree	2873	0	2918	4
Naïve Bayes	2791	82	2905	17
K-NN	2868	5	2916	6

As shown in Table 12 Naïve Bayes has the highest misclassification and the decision tree has the lowest misclassification. The RF has a perfect classification.

Table 10: Confusion matrix for the Open-source dataset

Model	TP	FP	TN	FN
RF	30046	0	30148	0
SVM	30036	10	26995	3153
DT	30045	1	30148	0
NB	29800	246	29705	443
K-NN	30042	4	30148	0

As seen in Table 13 SVM has the highest misclassification and DT has the lowest misclassification. The RF has a perfect classification with no FP and FN.

4.5 Validation Results

New Dataset Validation Set

Using k fold 5, the cross-validation result shows that all the models selected for model building in this study are good. As seen in Table 14 shows that RF, SVM, DT, and KNN are the best fit with an average score of 100% while Naive Bayes performs the lowest score of 98%.

Table 11: Cross-validation set for the new dataset.

Models	Average cross-validation scores (%)	Computation time (seconds)
Random Forest	100	6.2
SVM	100	4.8
Decision Tree	100	0.6
Naive Bayes	98	0.3
K-NN	100	9.2

Figure 34 shows that out of all the models, NB has very few overfitting errors while others have no overfitting errors. Overall, from Table 14, the result shows that the models learned well with the dataset in a short time.

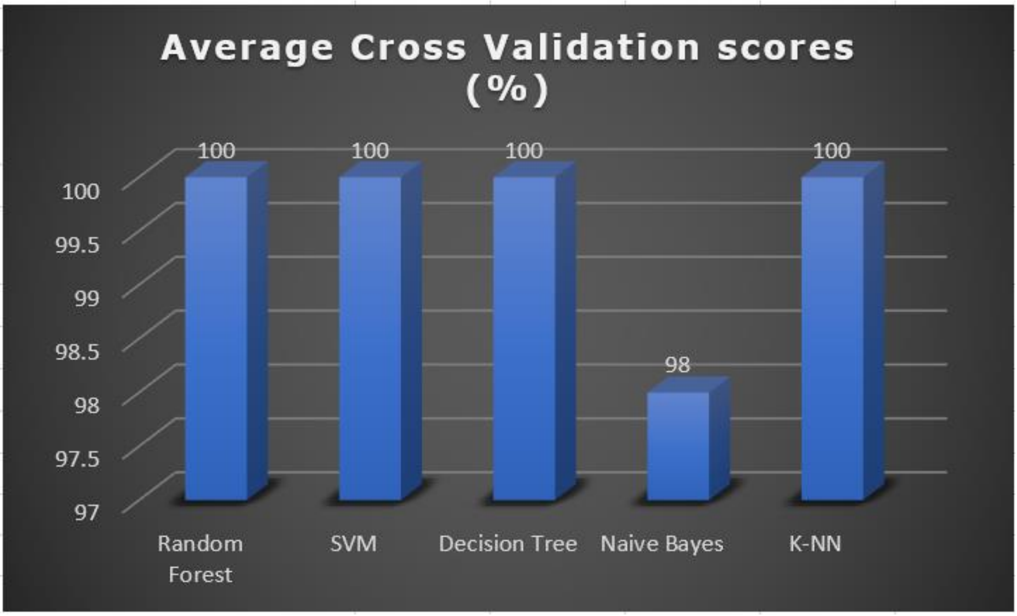

Figure 32: New dataset cross-validation scores.

Validation Set for CSE-CIC-IDS2018 Dataset

All the models performed very well in the validation set seen in Table 15 this shows that the models are fit for training the dataset through SVM and KNN took a long time for validation. However, SVM performed the lowest with the highest computation time. Figure 35 shows a visual performance of cross-validation.

Table 12: Cross-validation result for the CSE-CIC-IDS2018 dataset

Models	Average cross-validation scores (%)	Computation time (seconds)
Random Forest	97	37.3
SVM	92	10227.8
Decision Tree	97	3.9
Naive Bayes	97	3.0
K-NN	97	1115.4

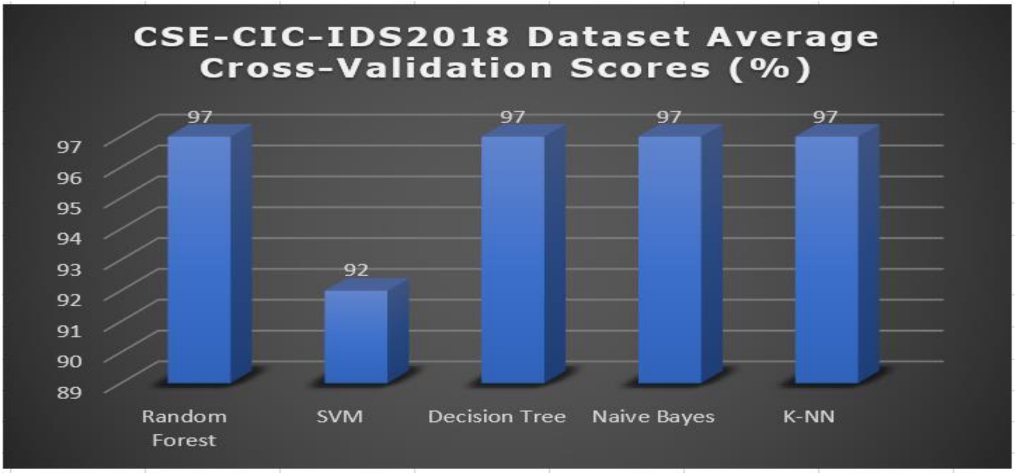

Figure 33: CSE-CIC-IDS2018 dataset cross-validation results.

Figure 36 shows the overall dataset accuracy performance illustrating that the new dataset approximately performed 100% in accuracy while the open-source dataset performed 99%.

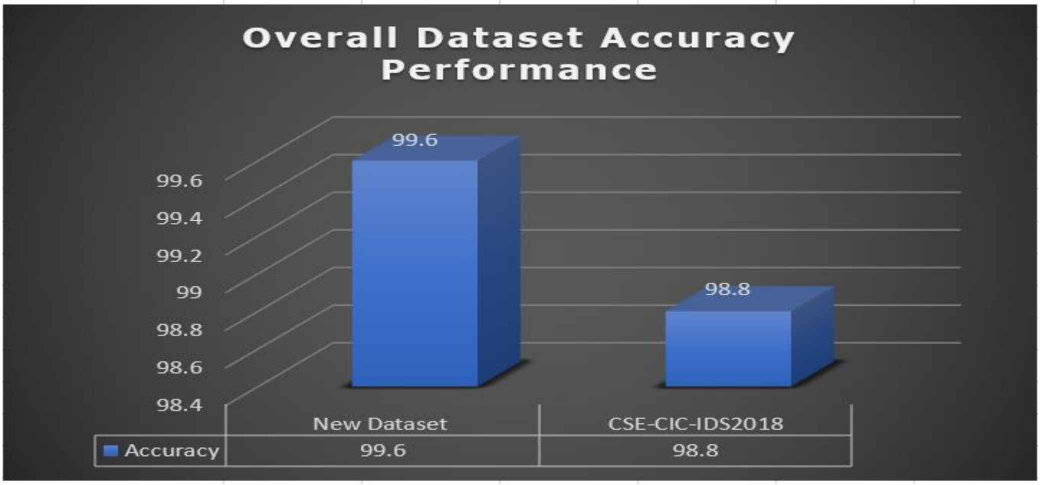

Figure 34: Overall dataset accuracy performance.

4.6 Reflections

This section reviews some measures stated in the investigation report to check whether they were properly carried out in this research. This includes the research questions, objectives of this study, Requirements and LESP issues.

4.6.1 Research Questions Reflections

This section focuses on addressing the research questions, as stated in the investigation report by examining the questions with the successes of the experimental sections in this report.

To perform this type of examination, the research questions in this study will be listed from the investigation report and presented, reflecting on whether it has been solved or not.

Q1 What type of dataset should be used to solve this problem?

Finding 1.1: A new set of data generated from a cloud platform (Owncloud) and an open-source dataset was used to support the evaluation of large data set. The datasets consist of two label classes namely benign and DDoS attack.

Q2 How can multi–supervised machine learning techniques be used to analyse the dataset (DDoS and benign traffic) to achieve better accuracy of DDoS attack detection in the cloud?

Finding 2.1: the new DDoS dataset achieved better accuracy (see Table 10 and figure 30) with different chosen supervised learning models using 31 features as shown in Figure 20 which were selected out of 84 features generated from the CICFlowMeter 4.0 tool.

Findings 2.2: Similarly, the supported open-source dataset performed well with better accuracy (See Table 11 and figure 33) with the selected ML models using 32 (see Figure 23) features selected from 80 features extracted from CICFlowMeter 3.0 tool.

Q3: What is the most efficient supervised ML algorithm to detect DDoS attacks in a cloud environment?

Finding 3.1: Out of the five models deployed for the analysis of both the new datasets, the RF, SVM, DT and KNN were the most efficient to detect DDoS attacks with a score of 100% across performance metrics namely accuracy, precision, recall and F1-score or measure as shown in table 10. Though, NB had a very high accuracy of 98%.

Finding 3.2: In the open-source dataset RF, DT and KNN were the most efficient with 100% accuracy, followed by NB with 99%.

In overall performance, the models of the new dataset performed better on all performance metrics (See Figure 36) than the open-source dataset. Most likely because the version of the CICFlowMeter 4.0 improved the flow features extracted which were selected for the Model building. Also, the validation result of the new dataset shows that the ML models learned very well.

4.6 2 Objective Comparison Reflections

this section focuses on addressing the objectives and aim of this study, as mentioned in the investigation report by evaluating the comparison with the successes of the experimental sections in this report.

To perform this type of evaluation, the objectives of this study will be listed from the investigation report and presented, reflecting on whether it has been attained or not.

1. Setup a cloud base system for conducting benign and malicious traffic.

The objective was met (see section 4.2 and Figure 9) in this report. An own cloud VirtualBox was set up to perform benign and DDoS traffic.

2. Examine the existing intrusion detection system and datasets.

The objective was met (see Chapter 3.0 of the investigation report) Existing intrusion detection systems and datasets were analysed.

3. Employ the most relevant machine learning methods to detect DDoS from a dataset (malicious and benign network traffic).

   The objective was met (see section 4.3) of this report. Supervised ML was selected with five distinct models RF, SVM, DT, NB, and KNN.

4. Create outcomes for each algorithm by training, validating, and testing the dataset against machine learning algorithms.

The objective was met (section 6.0) of this report. All the outcomes of ML building on training, testing and validation are documented (see Tables 10, 11, 14 and 15) in this report.

5. Analyse the models' performance and select the most efficient Machine learning algorithm.

The objective was met (see sections 6.2 and 6.3) conclusions of this report. Performance with the new dataset the most relevant models are RF, SVM, DT and KNN. With the open-source dataset, the most relevant models are RF, DT and KNN.

In addition, the project will be extended using part of the CSE-CIC-IDS2018 DDoS dataset.

4.6.3 Requirement Achievement Reflection

This section's main goal is to review the project's functional and non-functional requirements as specified in the investigation report and to assess the extent to which they have been met.

Some requirements listed in the investigation report were observed not to be relevant. However, some relevant ones were listed to examine the fulfilment.

Functional Requirement

The determining functional requirements for the system are listed beneath.

Hardware Requirement.

Software Requirements

1. **Must have** an operating system such as windows or ubuntu OS with a browser installed.

Owncloud has a Debian Operating system.

2. **Must have** a cloud environment

Owncloud VirtualBox was used for the experiment.

3. **Should have** an attacking platform like kali Linux virtual machine

(See section 4.2).

4. **Must have** a programming language and codes such as Python or R for analysis.

The machine learning model building was performed using Python language. The models were able to detect DDoS and benign traffic.

5. **Should have** Jupiter notebook installed through anaconda.

The system used the Jupyter notebook platform for the ML process.

Non-Functional Requirement

The system's non-functional requirements are listed beneath

1. **Must have** a dataset for training and testing using ML models

A dataset was produced, and an open-source dataset was used as well. Requirement met (See section 4.2).

2. **Must have** a tool to collect network packets and convert them to CSV format.

CICFlowMeter tool was used to achieve this requirement (See Section 4).

3. Irrelevant features **should** be removed from the dataset to achieve a fair result.
4. The accuracy and performance of each algorithm **should have** an 80% result and above.

The accuracy and performance of all the models supersede 80% least accuracy performance for the new dataset was 98% and the open-source dataset was 95% (See Tables 10 and 11).

4.6.4 Reflection on Legal, Ethical, Social and professional, Risk and safety

Legal Issues: This report affirms that there are no legal issues to be addressed. To comply with intellectual property laws and best practices for currently accessible material, all sources of information, including images, books, papers, datasets, and so forth, have been cited with the appropriate credit provided to the owners.

Ethical Issues: This report affirms that no living person or animal has been actively involved in the creation of this project from an ethical point of view. No Fraudulent acts were involved in any manner. This project was conducted professionally and closed environment by only the researcher.

Socially and professionally: This report affirms that no commercial companies were involved in the project's creation and that it was only performed out of personal interest and passion in this field of research. As a result, there are no plans to use the findings of this research for profit-making.

Risk and safety: The is affirm this report was conducted without risk involved. The attack was performed in an isolated environment with all the contents gathered safely. In all everything was conducted in a safe and risk-free environment.

4.6.5 Reflection on observation in this work

The proposed topic for this work "Analysing Cloud DDoS Attacks Using Supervised Machine Learning" did not exactly match this project. While starting this implementation report, it was discovered that the right topic would have been "Detection of DDoS attacks in the cloud using Supervised Machine Learning".