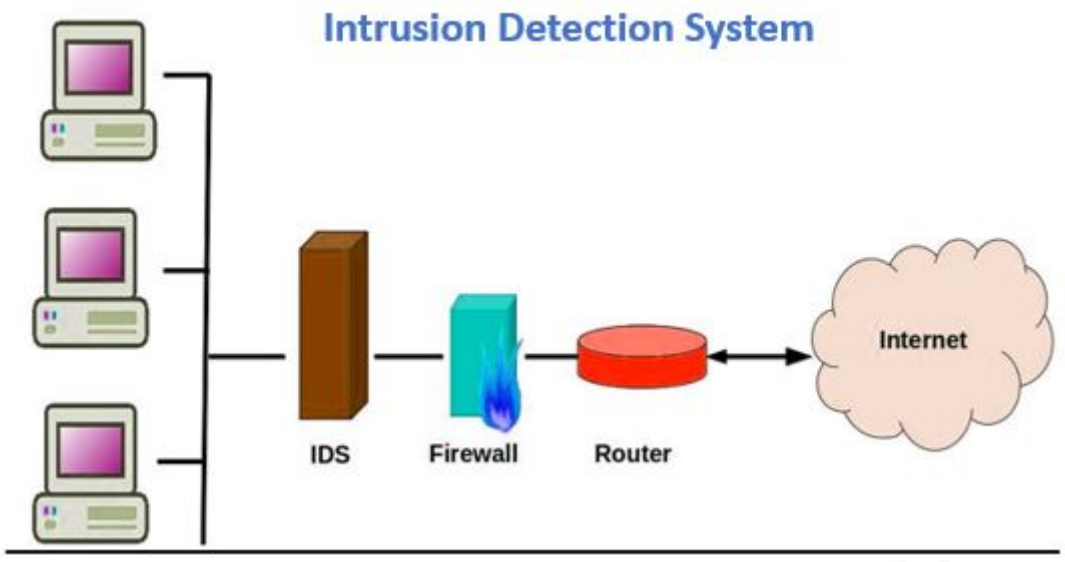
Name - Manish Arora Mis - 111803036 BTech Comp Div1 B1

Implementation of Intrusion Detection System

An ***Intrusion Detection System (IDS)*** is a system that monitors network traffic for suspicious activity and issues alerts when such activity is discovered. It is a software application that scans a network or a system for harmful activity or policy breaching. Intrusion prevention systems also monitor network packets inbound the system to check the malicious activities involved in it and at once send the warning notifications.



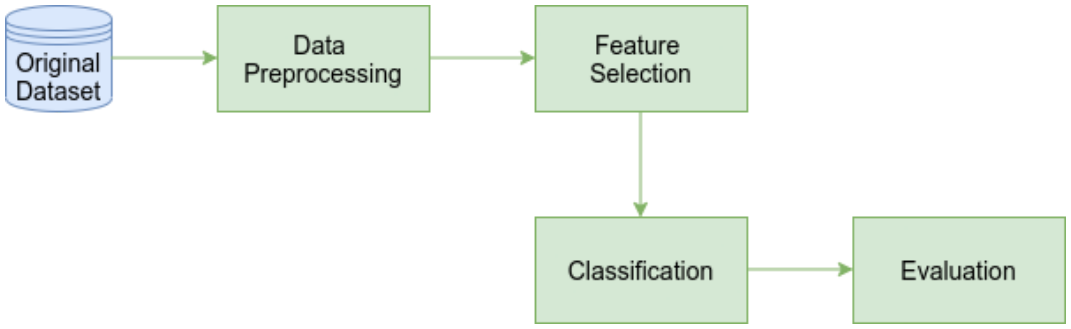
5 Types of Intrusion Detection System →

1. Network Intrusion Detection System (NIDS) - NIDS are set up at a planned point within the network to examine traffic from all devices on the network. It performs an observation of passing traffic on the entire subnet and matches the traffic that is passed on the subnets to the collection of known attacks.
2. Host Intrusion Detection System (HIDS) - HIDS run on independent hosts or devices on the network. A HIDS monitors the incoming and outgoing packets from the device only and will alert the administrator if suspicious or malicious activity is detected. It takes a snapshot of existing system files and compares it with the previous snapshot.
3. Protocol-based Intrusion Detection System (PIDS) - PIDS comprises a system or agent that would consistently reside at the front end of a server, controlling and interpreting the protocol between a user/device and the server. It is trying to secure the web server by regularly monitoring the HTTPS protocol stream and accepting the related HTTP protocol.

4. Application Protocol-based Intrusion Detection System (APIDS) - APIDS is a system or agent that generally resides within a group of servers. It identifies the intrusions by monitoring and interpreting the communication on application-specific protocols. For example, this would monitor the SQL protocol explicit to the middleware as it transacts with the database in the web server.
5. Hybrid Intrusion Detection System - It is made by the combination of two or more approaches of the intrusion detection system. In the hybrid intrusion detection system, host agent or system data is combined with network information to develop a complete view of the network system.

2 Detection Methods of IDS -

1. **Signature-based Method** - Signature-based IDS detects the attacks on the basis of the specific patterns such as number of bytes or number of 1's or number of 0's in the network traffic. The detected patterns in the IDS are known as signatures. Signature-based IDS can easily detect the attacks whose pattern (signature) already exists in the system but it is quite difficult to detect the new malware attacks as their pattern (signature) is not known.
2. **Anomaly-based Method** - Anomaly-based IDS was introduced to detect unknown malware attacks as new malware are developed rapidly. In anomaly-based IDS there is use of machine learning to create a trustful activity model and anything coming is compared with that model and it is declared suspicious if it is not found in the model. Machine learning-based methods have a better-generalized property in comparison to signature-based IDS as these models can be trained according to the applications and hardware configurations.

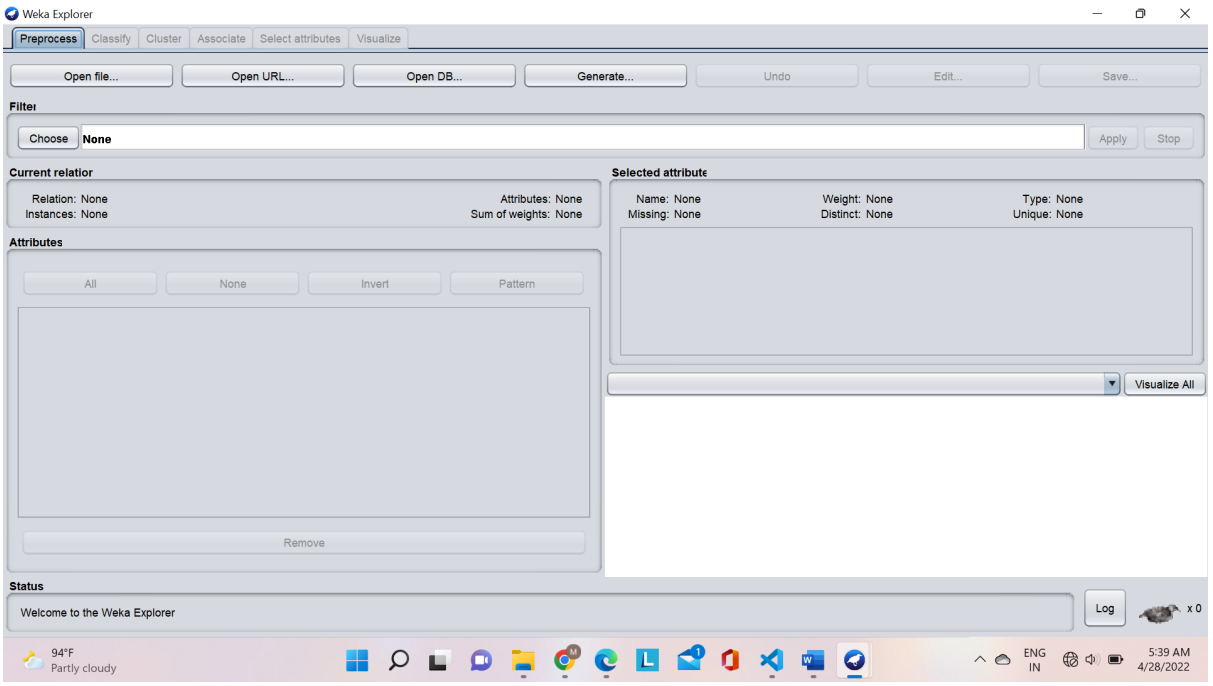


Implementation Details →

1. **Dataset Used - Portmap** - Reflection Based DDoS Attack from CiCDDoS2019 dataset.
   Site - https://www.unb.ca/cic/datasets/ddos-2019.html

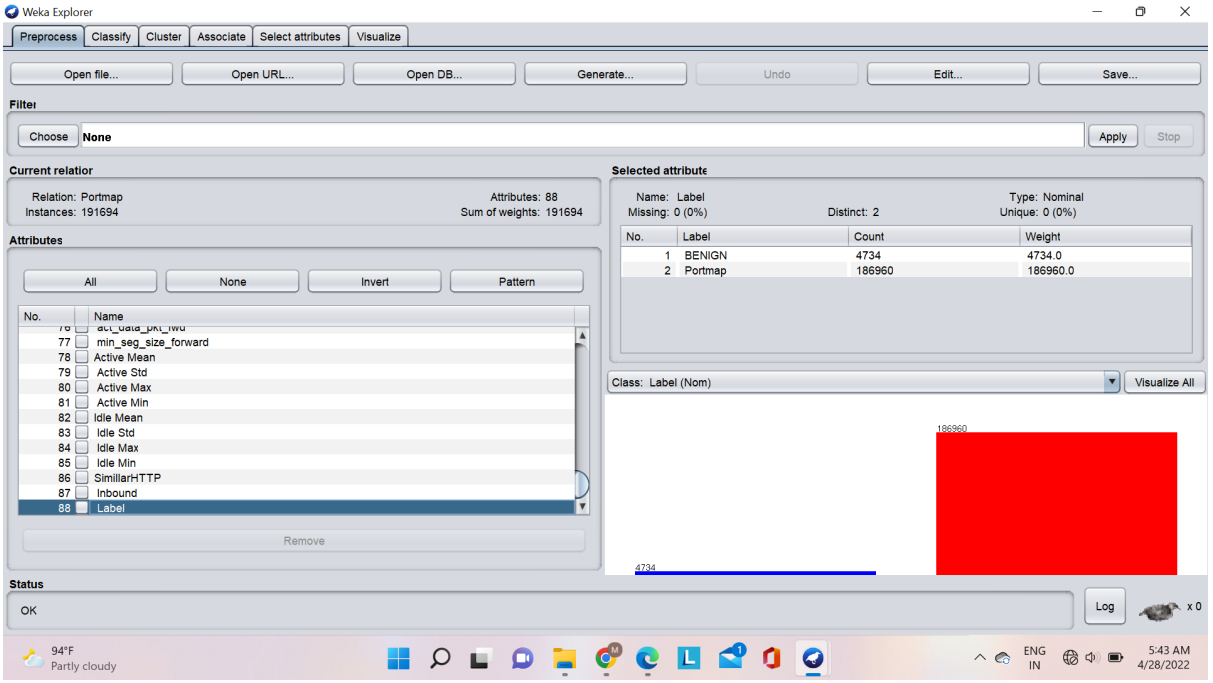
Portmapper (also referred to as rpcbind, portmap or RPC Portmapper) is a mechanism to which Remote Procedure Call (RPC) services register in order to allow for calls to be made to the Internet. Think of it as a directory service for RPC. When a client is looking to find the appropriate service, the Portmapper is queried to assist.

2. **Tool Used - Weka** - Weka is a collection of machine learning algorithms for data mining tasks. It contains tools for data preparation, classification, regression, clustering, association rules mining, and visualization.

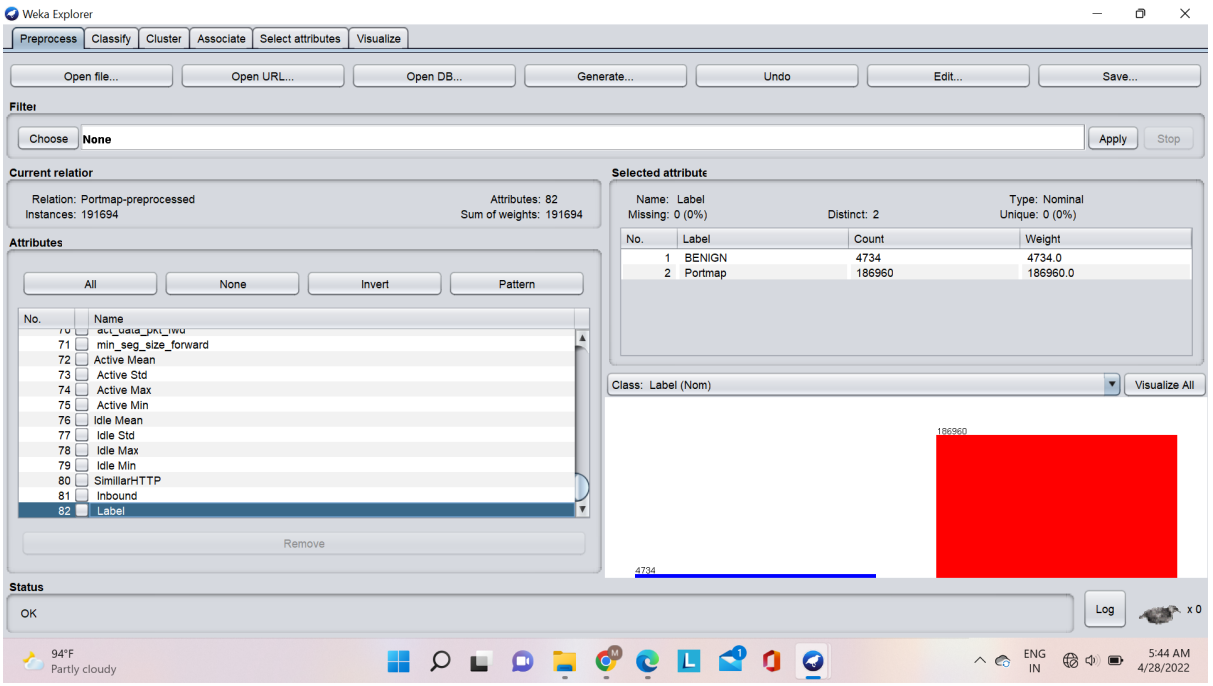


3. **Data preprocessing -** Data preprocessing is a step in the data mining and data analysis process that takes raw data and transforms it into a format that can be understood and analyzed by computers and machine learning.
   Steps followed are:-

1. Removed the columns Unnamed, FlowID, SourceIP, DestinationIP, Timestamp
2. Removed the duplicate column of Fwd Packet Length.1
3. Replaced Infiniti values with 0
4. Applying StringToNominal filter to the Column SimilarHTTP using Weka to convert it into a nominal attribute
   Dataset before preprocessing - 87 Attributes, 191694 instances, 1 Label



Dataset After preprocessing - 82 Attributes, 191694 instances, 1 Label

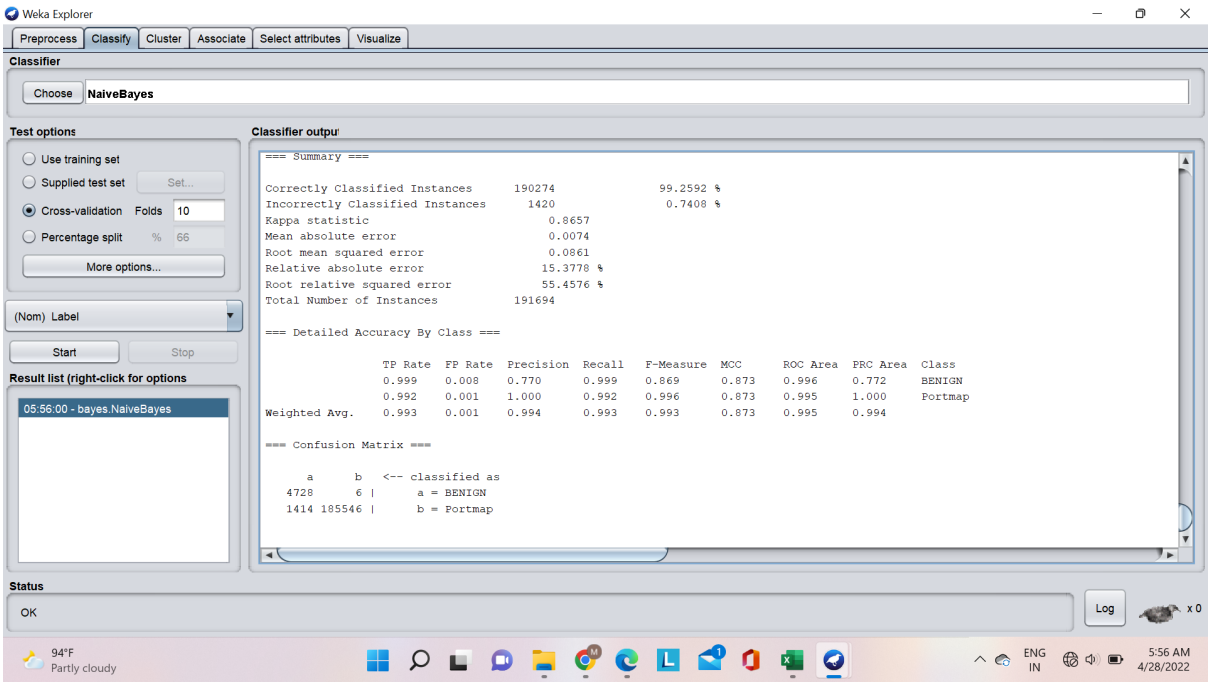


4. Classification without feature Selection - The dataset is given to different classification algorithms which includes Naive Bayes, JRip and PART. These classifiers take the preprocessed dataset and classify the label into Portmap and Benign traffic. Model Build Time, accuracy, precision are the parameters used to check which classifier performs best and give better results for Portmap dataset.

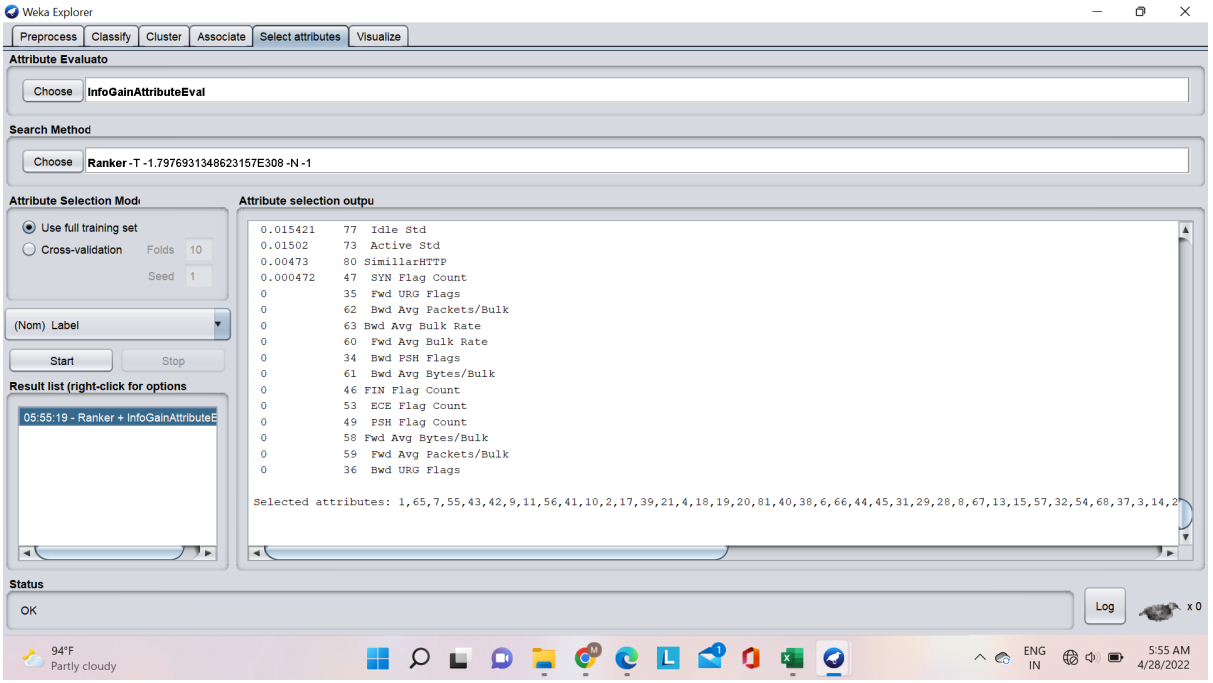
- **PART** - PART is an indirect method for rule generation. PART generates a pruned decision tree using the C4. 5 statistical classifier [16] in each iteration. From the best tree, the leaves are translated into rules.
- **Naive Bayes -** Naive Bayes is a probabilistic classifier inspired by the Bayes theorem under a simple assumption which is the attributes are conditionally

independent. It can be easily scalable to larger datasets since it takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers.

- **JRip -** JRip It implements a propositional rule learner called as “Repeated Incremental Pruning to Produce Error Reduction (RIPPER)” and uses sequential covering algorithms for creating ordered rule lists. The algorithm goes through 4 stages: Growing a rule, Pruning, Optimization and Selection

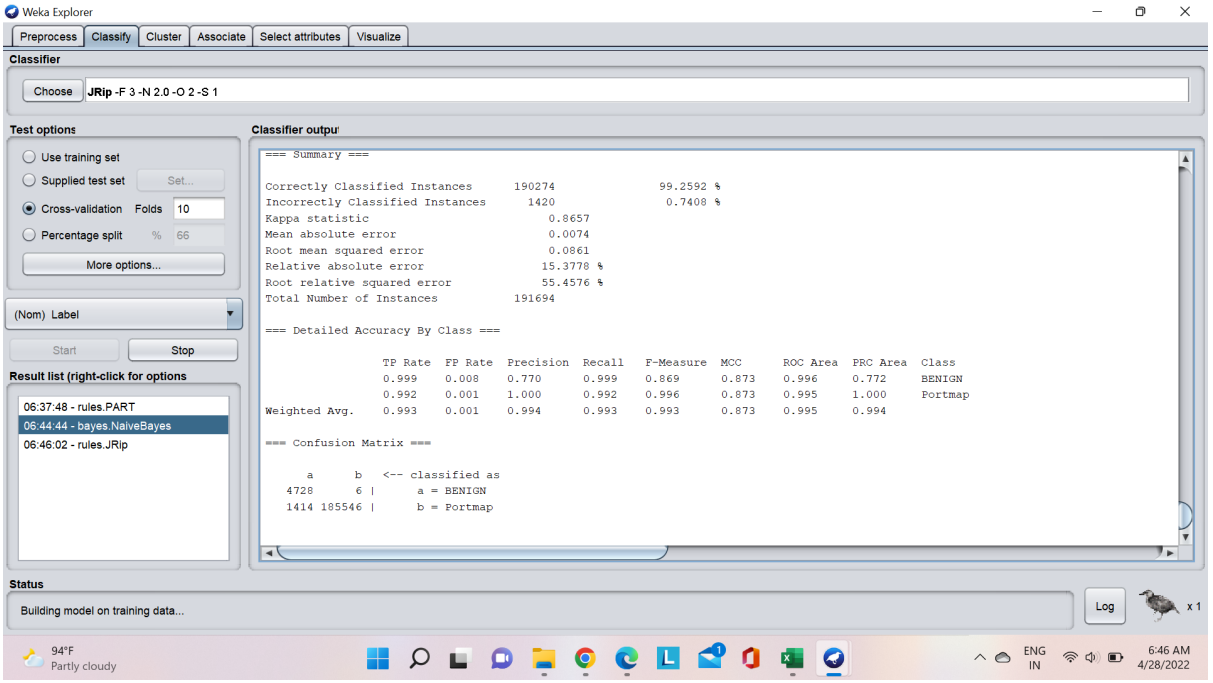


5. **Feature Reduction -** Feature selection or reduction is a significant process for intrusion detection system (IDS) in finding optimal features. Irrelevant features present in the dataset increase load on computing resources and affect the performance of the system. There are various types of feature reduction techniques which includes filter based, wrapper based and evolutionary feature selection techniques. Information Gain is one of the filter based techniques which uses statistical methods to calculate the entropy. **Infogain -** The InfoGain class is an implementation of a feature selection method by information gain. Information gain is a measure of the reduction in entropy of the class variable after the value for the feature is observed.



6. Classification After Feature Selection -

After selecting the correct set of attributes using InfoGain attribute selector, the resultant dataset contains 69 most important attributes out of 81 total attributes. The new dataset is then generated using the python script and again tested on the three classifiers PART, Jrip and Naive Bayes.



7. Results and Comparsion -

Classification without Feature Reduction -

Classifier	Accuracy(%)	Model Build Time(sec)	Precision	Recall
PART	99.9901	13.35	1	1
JRiP	99.9937	63.36	1	1
Naïve Bayes	99.2592	2.54	0.994	0.993

Classification with Feature Reduction using InfoGain -

Classifier	Accuracy(%)	Model Build Time(sec)	Precision	Recall
PART	99.9901	12.73	1	1
JRiP	99.9937	101.52	1	1
Naïve Bayes	99.2592	2.12	0.994	0.993

8. **Conclusion -** In this assignment, we first studied the basics of intrusion and its detection, and the various methods to detect intrusion. Then for the implementation purpose of IDS, we used the Portmap dataset. We used the Weka tool for performing data cleaning and performing classification as either Portmap or Traffic using three classifiers - PART, Naive Bayes and JRip - both before feature reduction and after. Finally, we show and compare both the results.