Project Report: **INTRUSION DETECTION SYSTEM**

SUBMITTED BY : BITF20M018-KABEER, BITF20M031-HAMZA

A- Problem Statement:

The task is to build a network intrusion detector, a predictive model capable of distinguishing between bad connections, called intrusions or attacks, and good normal connections.

B- Solution to the problem:

So we developed a ML Model to Detect Intrusion in a System the description of the ML model are follows as:

The Attacks fall into four main categories:

- #DOS: denial-of-service, e.g. syn flood;
- #R2L: unauthorized access from a remote machine, e.g. guessing password;
- #U2R: unauthorized access to local superuser (root) privileges, e.g., various “buffer overflow” attacks;
- #probing: surveillance and another probing, e.g., port scanning.

Dataset Description: Data files:

- kddcup.names : A list of features.
- kddcup.data.gz : The full data set
- kddcup.data_10_percent.gz : A 10% subset.
- training_attack_types : A list of intrusion types.
- typo-correction.txt : A brief note on a typo in the data set that has been corrected

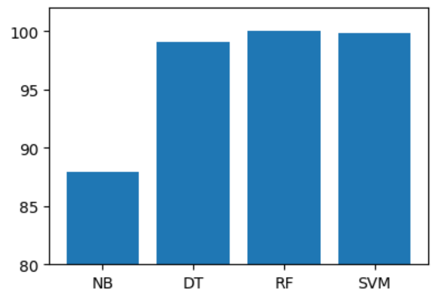
Our Solution Steps: Here are some steps we have followed to find the Solution

1. **Data Loading and Exploration**: The dataset is loaded and explored to understand its structure and features.
2. **Data Preprocessing:** Initial preprocessing steps are performed, including handling missing values and exploring categorical features.

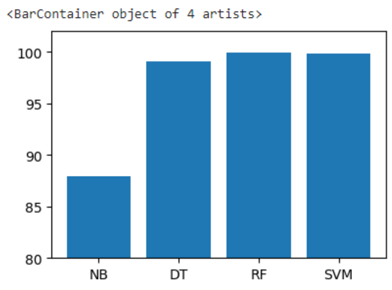
3. **Data Visualization:** Visualizations are created to understand the distribution of categorical features and the target variable.
4. **Data Correlation:** Correlation analysis is conducted, and highly correlated features are identified and removed.
5. **Feature Mapping:** Categorical features such as protocol_type and flag are mapped to numerical values.
6. **Modeling:** The notebook employs four different machine learning models:
    - Gaussian Naive Bayes
    - Decision Tree
    - Random Forest
    - Support Vector Machine (SVM)
7. **Model Evaluation:** The performance of each model is evaluated using training and testing accuracy. The notebook also includes visualizations of training time, testing time, training accuracy, and testing accuracy for each model.
8. **Predicted Attack Types:** The notebook provides insights into the predicted attack types for a subset of test data for each model.

C- Technical Details:

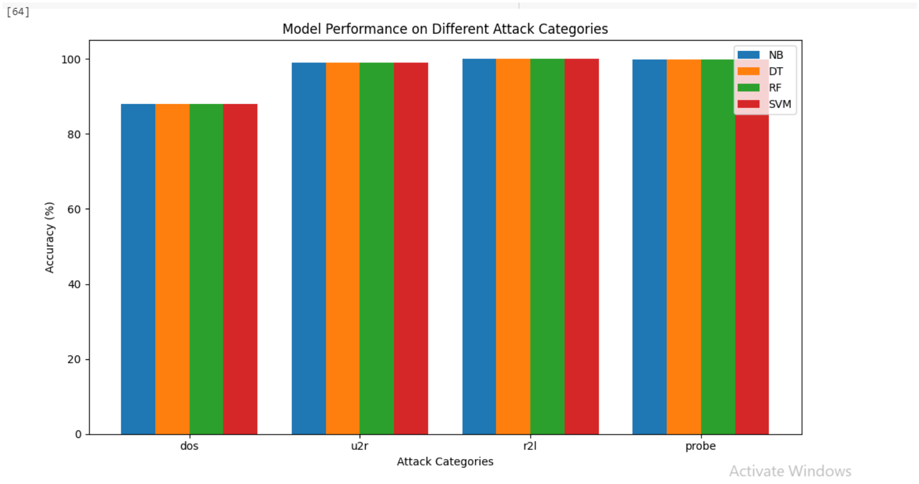
Training accuracy Measure:



Test Accuracy Measure:



```
<BarContainer object of 4 artists>
```



D- Future Directions or challenges

1. Future Directions:

The field of intrusion detection is constantly evolving. Some of the current trends include:

- Machine learning and artificial intelligence can be used to improve the accuracy of IDS systems and to detect new and unknown attacks.
- **The use of big data:** Big data can be used to store and analyze large amounts of data from IDS systems. This can help to identify trends and patterns that can be used to improve the effectiveness of IDS systems.
- **The use of cloud-based IDS:** Cloud-based IDS systems can provide a more scalable and flexible way to deploy IDS.

2. Challenges:

- Data collection and Preparation
- Algorithms selection and Training or Testing
- False Positives and Negatives:
  - IDS systems often generate false alarms (false positives) which can waste time and resources.
  - They can also miss real attacks (false negatives) which can have serious consequences.
- Evasion , deployment and maintenance