*Team members- ANANYA THAKUR (01501042024) Indira Gandh Delhi AARUSHI JAIN (00601042024) Technical University For Women*

*(MAE-1 ) (IGDTUW)*

***SYNOPSIS***

***Topic -*** *Intrusion Detection System Using Machine Learning Algorithms.*

***Theory -******Intrusion Detection System****is a software application that detects network intrusion using various machine learning algorithms. IDS monitors a network or system for malicious activity and protects a computer network from unauthorized access by users, including perhaps insiders. The intrusion detector learning task is to build a predictive model (i.e., a classifier) capable of distinguishing between 'bad connections' (intrusion/attacks) and 'good (normal) connections'. 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 other probing, e.g., port scanning.*

***Objective -*** *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.*

***Dataset source –*** *We have download the dataset used in this project from Kaggle (the name of the dataset is Intrusion Detection System Using Machine Learning).*

*Dataset Description: Data files:*

*kddcup.names: A list of features.*

*kddcup.data\_10\_percent: A 10% subset of the dataset.*

*training\_attack\_types: A list of intrusion types.*

***Algorithms and Libraries used –*** *Numpy, Panda, Matplotlib, Seaborn, Scikit-learn, OS.*

*Algorithms: Gaussian Naive Bayes, Decision Tree, Random Forest, Support Vector Machine, Logistic Regression.*

***Approach Used -*** *We have applied various classification algorithms that are mentioned above on the KDD dataset and compare their results to build a predictive model.*

***Methodology:***

***Step 1: Importing and Setting Up the Data:*** *Importing libraries and reading features list.**Appending columns to the dataset and adding a new column name 'target' to the dataset. Shape of dataframe and getting data type of each feature. Finding missing values of all features.*

***Step 2: Data Exploration:*** *Finding Categorical Features. Visualizing Categorical Features using various graphs.*

***Step 3: Data Preprocessing***

***Step 4: Splitting the Dataset***

***Step 5: Feature Encoding***

***Step 6: Correlation Analysis and Removing Highly Correlated Features:*** *Dropping Columns that don't provide high value.*

***Step 7: Scaling the Data***

***Step 8: Model Training and Test Accuracy***

***Step 9: Model Selection***

***Conclusion -*** *To build a machine learning-based Network Intrusion Detection System for classifying network traffic as normal or malicious. This project will evaluate key performance metrics such as accuracy, precision, and classification reports across different classification models. The goal is to identify the best-performing model that can accurately detect intrusions, thereby enhancing the security of network infrastructure*.