Name: Mahad Ashraf

Roll No. 20P-0563

Name: Muhammad Sherjeel Akhtar

Roll No.: 20P-0101

Lab Task: Final Project

Subject: Artificial Intelligence Lab

Section: BCS-6C

Submitted to Respected Ma'am: Miss Hurmat Hidayat

**Report**

**Abstract:**

The aim of this project was to apply different classification and clustering algorithms to classify cyber-attacks in network traffic. The dataset provided contained 23 different attack types, which were converted into five classes. The data was preprocessed by handling missing values, outliers, and feature scaling. The most relevant features were identified using correlation analysis. The classification algorithms used were decision trees, K-nearest neighbours, and artificial neural networks. The performance of each algorithm was evaluated using appropriate metrics such as accuracy, precision, recall, and F1-score. In the last task, clustering was performed using the k-means algorithm, and the results were visualized using a scatter plot.

**Introduction:**

The increasing use of the internet and technology has led to an increase in cyberattacks. It is important to develop methods to detect and classify these attacks to ensure the security of network traffic. In this project, we aim to develop classifiers for the prediction of attack types given their attributes. This project is divided into different tasks, including data preprocessing, feature engineering, classification using different algorithms, and clustering.

**Data Preprocessing:**

The first step in our project was to explore the dataset to understand the characteristics and distributions of different features. We found that the dataset contained missing values and outliers, which were handled by filling the missing values with the mean and removing the outliers. The dataset was then converted to five classes by mapping the original 23 classes to the new five classes. Feature selection was performed using correlation analysis, which identified the most relevant features for classification.

**Classification using a Decision Tree Algorithm:**

A decision tree algorithm was used to develop a classification model. The model was trained on the preprocessed dataset, and its performance was evaluated using appropriate metrics such as accuracy, precision, recall, and F1-score. The results showed that the decision tree algorithm had an accuracy of 97.46%, which is a good performance.

**Classification using the K-Nearest Neighbours Algorithm:**

The KNN algorithm was implemented and trained using the preprocessed and selected features. The optimal value of k was determined to be 3, and the algorithm achieved an accuracy of 97.25%. Although the accuracy was slightly lower than the decision tree algorithm, the performance was still considered good.

**Classification using Artificial Neural Networks (ANN):**

An ANN model was chosen that suits the specific requirements of the task. The selected model was trained on the preprocessed dataset to learn from the provided data. The performance of the trained model was evaluated using appropriate evaluation metrics such as accuracy, precision, recall, and F1-score. The hyper parameters such as the learning rate and the number of hidden layers were optimized, which resulted in an accuracy of 97.76%, which is the best performance among the classification algorithms used in this project.

**Clustering using the k-means algorithm:**

The last task involved classification followed by clustering. In this task, the column containing labels was dropped from the dataset, and the dataset was labelled using the k-Means algorithm. The clustering returned a label for each record, and the results were visualized using a scatter plot. The scatter plot showed that the dataset was well-clustered, and the results were consistent with the classification results.

**Conclusion:** In conclusion, this project aimed to classify cyber-attacks in network traffic using different classification and clustering algorithms. The dataset was preprocessed by handling missing values, outliers, and feature scaling. Feature selection was performed using correlation analysis. The classification algorithms used were Decision Tree, K-Nearest Neighbours, and Artificial Neural Networks, and the results showed that all algorithms performed well with an accuracy of over 97%. The clustering results were consistent with the classification results. Overall, this project showed the effectiveness of different classification and clustering algorithms for the classification of cyber-attacks in network traffic.