Name: Pranay Rokade

Classification Algorithms Performance Report

Problem Statement
Compare the performance of various classification algorithms.

Dataset
Heart Attack Data Set

Link of the Dataset
https://www.kaggle.com/datasets/pritsheta/heart-attack

Methodology
The project begins with Exploratory Data Analysis (EDA) to understand the Heart Attack dataset obtained from Kaggle. EDA involves visualizing data trends and patterns, which is crucial for informing modeling decisions. During EDA, histograms were plotted to assess the distribution of each feature, and a correlation matrix was used to identify relationships between features.

After the initial analysis, the data was prepared for modeling. This included splitting the dataset into training and testing sets to evaluate model performance effectively. The training data was scaled using StandardScaler to standardize the feature values, ensuring that all features contribute equally to the model's performance, particularly for distance-based algorithms like KNN.

Several classification algorithms were implemented sequentially, including Logistic Regression, Naive Bayes, K-Nearest Neighbors, Decision Trees, Random Forest, and KMeans. For each model, the classification report, which includes precision, recall, and F1-score metrics, was printed. Additionally, the confusion matrix was visualized to better understand each model's performance in terms of true positives, true negatives, false positives, and false negatives.

Finally, the results of all models were compiled into a comparison table, presenting their respective accuracy, precision, recall, and F1-score, providing a clear overview of the effectiveness of each classification technique on the Heart Attack dataset.

Selected Techniques
Logistic Regression, Naive Bayes, K-Nearest Neighbors, Decision Tree, Random Forest, KMeans

Results

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.868852	0.852941	0.90625	0.878788
Naive Bayes	0.885246	0.903226	0.87500	0.888889
K-Nearest Neighbors	0.852459	0.828571	0.90625	0.865672
Decision Tree (DT)	0.786885	0.827586	0.75000	0.786885
Random Forest	0.868852	0.900000	0.84375	0.870968
KMeans	0.605960	0.723728	0.60596	0.656556

Conclusion

Naive Bayes outperforms the other algorithms in terms of accuracy and F1-score, while KMeans shows lower performance due to its clustering nature. Logistic Regression and Random Forest also demonstrate strong performance.