Lab Report: Applying Machine Learning Algorithms

Numan Shafi

Task 1: Data Preprocessing

Question 1: Load and Explore the Datasets

- Load the Iris dataset for the classification task. - Load the California Housing dataset for the regression task. - Explore the datasets, check for missing values, and identify feature and target columns.

Question 2: Feature Scaling and Splitting Data

- Apply feature scaling using `StandardScaler` for both datasets. - Split the datasets into training and testing sets (80% training, 20% testing).

Task 2: Classification Algorithms

Question 3: Apply k-Nearest Neighbors (k-NN) for Classification

- Train a k-NN model on the Iris dataset. - Test different values of k (e.g., 1 to 15) and observe the impact on accuracy.

Question 4: Apply Support Vector Machine (SVM) for Classification

- Train an SVM model on the Iris dataset using different kernels (*linear*, *poly*, *rbf*). - Compare the performance of different kernels based on accuracy.

Question 5: Apply Random Forest Classifier

- Train a Random Forest Classifier with varying numbers of estimators (e.g., 10, 50, 100). - Compare the performance based on accuracy and discuss the impact of the number of estimators.

Task 3: Regression Algorithms

Question 6: Apply Linear Regression for Regression

- Train a Linear Regression model on the California Housing dataset. - Evaluate the performance using Mean Squared Error (MSE) and R^2 score.

Question 7: Apply Decision Tree Regression

- Train a Decision Tree Regressor on the California Housing dataset. - Compare the model's performance with the Linear Regression model using MSE and R^2 score.

Task 4: Model Evaluation and Comparison

Question 8: Evaluate Classification Models Using Classification Metrics

- Use accuracy, precision, recall, and F1-score to evaluate the k-NN, SVM, and Random Forest models.

Question 9: Evaluate Regression Models Using Regression Metrics

- Compare the performance of Linear Regression and Decision Tree models using MSE and R^2 score.

Task 5: Conclusion

Question 10: Compare and Summarize the Findings

- Summarize the key differences in performance between the classification models (k-NN, SVM, Random Forest). - Summarize the differences between the regression models (Linear Regression, Decision Tree).