Practical Questions

For the practical questions, assume that you have a dataset called "heart_data.csv" containing various features related to heart disease. You are tasked with predicting whether a patient has heart disease (target variable: "heart_disease") using logistic regression, KNN, and decision trees. The dataset has already been preprocessed and split into training and testing sets.

1. Load the dataset and split it into training and testing sets (80% training, 20% testing).

2. Apply logistic regression to the training set and evaluate its accuracy on the testing set. Report the accuracy score.

3. Implement KNN with k=5 on the training set and compute its accuracy on the testing set. Report the accuracy score.

4. Build a decision tree classifier on the training set and evaluate its accuracy on the testing set. Report the accuracy score.

5. Compare the accuracy scores of logistic regression, KNN, and decision trees. Which algorithm performed the best in this case?

6. Choose one of the algorithms that performed well and briefly discuss why it might be suitable for this specific problem.

7. What are some potential ways to improve the performance of the chosen algorithm?

8. Evaluate the importance of each feature in the decision tree classifier. Plot a bar chart showing the feature importances.

9. Explain one practical application for each of the three algorithms: logistic regression, KNN, and decision trees.

10. In what scenarios would you prefer to use logistic regression over KNN and decision trees, and vice versa?

Note: For practical questions, you are expected to write code and provide explanations/results.

Dataset Link :
https://drive.google.com/file/d/14yejlC3AAQggFNovo0q6AgXAe9elHBd4/view?usp=sharing