Report: Week 1 - Disease Prediction

Steps Performed

The project involved multiple stages starting with preprocessing of the dataset, splitting it into training and testing sets, and applying different machine learning models. The models evaluated included Logistic Regression and Random Forest Classifier. Evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix were calculated to compare performance.

Results Obtained

The Logistic Regression model achieved an accuracy of around 86.9%, with reasonable precision and recall. However, the Random Forest Classifier outperformed Logistic Regression by achieving a higher accuracy of approximately 88.5%, a recall of 87.5%, and an F1-score of 0.889. The confusion matrix analysis showed Random Forest made fewer false negatives, making it a better choice for disease prediction. Hence, Random Forest was selected as the final model.