Case Study: Heart Disease Prediction Using Machine Learning

Problem Statement

Heart disease is one of the leading causes of death globally. Traditional medical diagnosis often depends on manual interpretation of numerous clinical parameters, which can be time-consuming and prone to human error. The goal of this project is to build a machine learning model that can accurately predict the presence of heart disease based on patient data such as age, cholesterol, blood pressure, and ECG results.

Objectives

• To analyze clinical data and identify key predictors of heart disease.
• To implement and evaluate different ML algorithms for heart disease prediction.
• To propose an improved, data-driven model for early detection and prevention.

Data Preprocessing

Dataset Used: Heart Disease dataset from UCI repository (processed CSV version).

Steps Taken:
• Handled missing values and formatted columns properly.
• Scaled numerical features using StandardScaler.
• Split dataset into 80% training and 20% testing sets.
• Encoded categorical features as required.

Model Selection and Development

Tested Models: Logistic Regression, Decision Tree, and KNN.
Final Model: Logistic Regression was selected for its higher accuracy and stability.
Cross-validation: 5-fold cross-validation was applied to ensure generalization.

Visualizations and Insights

• Correlation heatmap showed strong relationships between features like cp (chest pain type) and thalach (maximum heart rate achieved) with target outcome.
• The model achieved ~85% accuracy, suggesting reliable predictive capability.
• Visualization of predictions confirmed a balanced distribution of predicted classes (disease vs no disease).

Recommendations

• Integrate this model into healthcare dashboards for real-time patient risk evaluation.
• Extend the model using ensemble techniques or deep learning for even better accuracy.
• Collect more recent and diverse patient data to improve model generalization.

Conclusion

This case study demonstrates that machine learning models, particularly Logistic Regression, can effectively support medical professionals in predicting heart disease risk. By providing early warnings, such systems can help reduce diagnostic time and improve patient outcomes.