# Performance Analysis of Supervised Classification Models on Heart Disease Prediction

**Ezekiel Adebayo Ogundepo1; Waheed Babatunde Yahya2; Olubayo Adekanmbi1**

1Data Science Nigeria (DSN), Lagos, Nigeria

2Department of Statistics, University of Ilorin, Ilorin, Nigeria.

e-mail: ezekiel@datasciencenigeria.ai, ogundepoezekiel@gmail.com1; wbyahya@unilorin.edu.ng2; olubayo@datasciencenigeria.ai1

**Abstract**

This paper presents a predictive analysis of data on heart disease patients to determine the possible risk factors that are associated with their heart disease status. Two independent (but similar) published heart disease datasets, the Cleveland data (used to build classification models) and the Statlog data (used for results’ validation) were considered for analysis. Detail exploratory analysis using the Chi-square test of independence was performed on the Cleveland data after which ten standard classification models were constructed for class prediction. The classification models were constructed by partitioning the Cleveland data randomly into 208 (70%) training samples and 89 (30%) test samples over 200 replications. Preliminary results showed that some of the bio-clinical categorical variables are strongly associated with the heart disease conditions of the patients (p < 0.001). The classification results from the test samples indicated that the Support Vector Machine yielded the best predictive performances with 85% Accuracy, 82% Sensitivity, 88% Specificity, 87% Precision, 91% Area under the ROC curve (AUC), and 38% Log Loss value. These results were validated on the Statlog data in 10-fold cross-validation which were all consistent with those obtained from the Cleveland dataset.

**Keywords**

Machine learning, Classifiers, Feature Selection, Heart Disease Classification, Exploratory Data Analysis (EDA), Performance measures.