

Evaluating Models Through Feature Selection Methods Using Data Driven Approach

Authors : Shital Patil, Surendra Bhosale

Abstract : Cardiac diseases are the leading causes of mortality and morbidity in the world, from recent few decades accounting for a large number of deaths have emerged as the most life-threatening disorder globally. Machine learning and Artificial intelligence have been playing key role in predicting the heart diseases. A relevant set of feature can be very helpful in predicting the disease accurately. In this study, we proposed a comparative analysis of 4 different features selection methods and evaluated their performance with both raw (Unbalanced dataset) and sampled (Balanced) dataset. The publicly available Z-Alizadeh Sani dataset have been used for this study. Four feature selection methods: Data Analysis, minimum Redundancy maximum Relevance (mRMR), Recursive Feature Elimination (RFE), Chi-squared are used in this study. These methods are tested with 8 different classification models to get the best accuracy possible. Using balanced and unbalanced dataset, the study shows promising results in terms of various performance metrics in accurately predicting heart disease. Experimental results obtained by the proposed method with the raw data obtains maximum AUC of 100%, maximum F1 score of 94%, maximum Recall of 98%, maximum Precision of 93%. While with the balanced dataset obtained results are, maximum AUC of 100%, F1-score 95%, maximum Recall of 95%, maximum Precision of 97%.

Keywords : cardio vascular diseases, machine learning, feature selection, SMOTE

Conference Title : ICHISFD 2022 : International Conference on Healthcare Information Systems and Future Directions

Conference Location : Toronto, Canada

Conference Dates : September 20-21, 2022