A Ratio-Weighted Decision Tree Algorithm for Imbalance Dataset Classification

Authors : Doyin Afolabi, Phillip Adewole, Oladipupo Sennaike

Abstract : Most well-known classifiers, including the decision tree algorithm, can make predictions on balanced datasets efficiently. However, the decision tree algorithm tends to be biased towards imbalanced datasets because of the skewness of the distribution of such datasets. To overcome this problem, this study proposes a weighted decision tree algorithm that aims to remove the bias toward the majority class and prevents the reduction of majority observations in imbalance datasets classification. The proposed weighted decision tree algorithm was tested on three imbalanced datasets- cancer dataset, german credit dataset, and banknote dataset. The specificity, sensitivity, and accuracy metrics were used to evaluate the performance of the proposed decision tree algorithm on the datasets. The evaluation results show that for some of the weights of our proposed decision tree, the specificity, sensitivity, and accuracy metrics gave better results compared to that of the ID3 decision tree and decision tree induced with minority entropy for all three datasets.

Keywords : data mining, decision tree, classification, imbalance dataset

Conference Title : ICSTAIML 2022 : International Conference on Systems Thinking, Artificial Intelligence and Machine Learning

Conference Location : Cape Town, South Africa

Conference Dates : November 03-04, 2022