

An Analysis of Classification of Imbalanced Datasets by Using Synthetic Minority Over-Sampling Technique

Authors : Ghada A. Alfattni

Abstract : Analysing unbalanced datasets is one of the challenges that practitioners in machine learning field face. However, many researches have been carried out to determine the effectiveness of the use of the synthetic minority over-sampling technique (SMOTE) to address this issue. The aim of this study was therefore to compare the effectiveness of the SMOTE over different models on unbalanced datasets. Three classification models (Logistic Regression, Support Vector Machine and Nearest Neighbour) were tested with multiple datasets, then the same datasets were oversampled by using SMOTE and applied again to the three models to compare the differences in the performances. Results of experiments show that the highest number of nearest neighbours gives lower values of error rates.

Keywords : imbalanced datasets, SMOTE, machine learning, logistic regression, support vector machine, nearest neighbour

Conference Title : ICSRD 2020 : International Conference on Scientific Research and Development

Conference Location : Chicago, United States

Conference Dates : December 12-13, 2020