

Binary and Multi-Class Classification Using Supervised Machine Learning Algorithms and Ensemble Model

H Asela

Department of Electrical and Information Engineering, University of Ruhuna, Sri Lanka

asela.h@eie.ruh.ac.lk

Classification is a vital aspect in data mining, where vast quantities of data are segregated into discrete classes. Models based on different statistical and machine learning approaches are used for this task. However, the classification performance depends on multiple factors like selected algorithm, domain and features of the dataset. The objective of this study is to evaluate the classification performance of widely used supervised machine learning algorithms; Decision Tree (DT), Naïve Bayes (NB) algorithm, Support Vector Classifier (SVC), K-Nearest Neighbour (KNN) algorithm and the Ensemble Model (EM) based on soft voting technique. These algorithms are tested on 6 datasets in different domains, and the datasets contain both multi-class and binary class data as well as balanced and imbalanced data. Accuracy, Precision and Recall are used as evaluation metrics to evaluate the classification performance in balanced datasets, where F1-measure is used in imbalanced dataset for the same task. The evaluation results indicate that EM outperformed single algorithms at most instances. When comparing single algorithms, KNN performed best with multi class classification, where SVC performed best in binary classification in balanced datasets. Also, KNN showed the best classification performance when it comes to imbalanced dataset. All the algorithms performed well when the data set is balanced. However, the classification performance in all models including EM is below expectation, when the data distribution is highly imbalanced.

Keywords: *classification, machine learning, supervised algorithms, ensemble model, soft voting classifier*