Improving Random Forest with Pre-pruning technique for Binary classification

IlhemTarchoune[*], Akila Djebbar[2] and Hayet Farida Merouani[3]

Computer Science Department, LRI Laboratory, SRF equip,

Badji Mokhtar University, BP 12, 23000, Annaba, Algeria

[*]*(tarchouneilhem@gmail.com)*

Abstract –Random Forest (RF) is a popular machine learning algorithm. It is based on the concept of ensemble learning, which is a process of combining several classifiers to solve a complex problem and improve model performance. The random forest allows extending the notions of decision trees (DT) in order to build more stable models. In this work we propose to further improve the predictions of the trees in the forest by a pre-pruning technique, which aims to optimize the performance of the nodes and to minimize the size of the trees. Two experiments are performed to evaluate the performance of the proposed method; in the first experiment we applied the Classical Random Forest algorithm (CRF) with several different trees. While in the second one, a pre-pruning technique is established on the trees in order to define the optimal size of the forest. Finally, we compared the results obtained. The main objective is to produce accurate decision trees with high precision. The effectiveness of the proposed method is validated on five medical databases; the prediction precision will be improved with 83%, 94%, 95%, 97%, and 81% for Diabetes, Hepatitis, SaHeart, EEG-Eye-State, Prostate-cancer databases respectively. The performance results confirm that the proposed method performs better than the classical random forest algorithm.

Keywords –Random Forest (RF), Decision tree (DT), Pre-pruning, Binary Classification, Medical Data.