

PERFORMANCE OF TREE-BASED CLASSIFIERS WITH MINIMUM DESCRIPTION LENGTH DISCRETIZATION TECHNIQUE FOR EPILEPSY DATASET

¹Zam Zam Syafurah Samsudin & ²Muhamad Hasbullah Mohd Razali

¹zamzamsyafurah30@gmail.com

²hasbullah782@perlis.uitm.edu.my

ABSTRACT

A decision tree classifiers represented by a tree-like structure is a hierarchical model composed of internal nodes, leaf nodes and branches. There are several reasons why tree-based classifiers is quite popular among the other classifiers, this is because decision trees are easy to construct and the resulting trees are readily interpretable. However, different tree-based classifiers with different setting will give different results, hence the objective of this research are to identify the performance of tree-based classifiers with Minimum Description Length discretization technique for epilepsy dataset where specific objective are to compare the performance of tree-based classifiers based on the discretized epilepsy dataset with Principle Component Analysis, and to identify the optimal parameter for the best tree-based classifier. Minimum Description Length is the discretization method that was used to discretize an epilepsy dataset. Type of tree-based classifiers that were compared are Decision Stump, J48, Random Forest and Random Tree. From the result, it shows that Random Forest is the most appropriate tree-based classifier with setting number 1. Where several default settings were changed and the result was improved. The changed settings for Random Forest are Break Ties Randomly was changed to True and the Number of execution Slots is 1. The actual accuracy percentage of Random Forest has increased from 95.313 % to 95.3652 % after setting number 1 was applied. Besides that. The value of accuracy percentage has also shows improvement after Principle Component Analysis was applied.

Keywords: Discretization, Minimum Description Length, Tree-Based Classifier, Epilepsy