DATA MINING APPROACH FOR CLUSTERING, CLASSIFICATION AND VISUALIZING BASED ON ANTHROPOMETRIC MEASUREMENT

Amarasena P.T.[1*], Kumara B.T.G.S.[1] and Jonition S.[2]

[1]Department of Sport Sciences and Physical Education, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Sri Lanka

[2]Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Sri Lanka

*pasindutharushika@gmail.com

Anthropometric measurements are generally used to determine and predict achievement in different sports. An athlete's anthropometric and physical characteristics may perform important precondition for successful participation in any given sport. Further, anthropometric profiles indicate whether the player would be suitable for the competition at the highest level in a specific sport. Recently, more researches has been carried out on Sport Data mining. This study was conducted as three parts. In first part it proposes a visualization approach to identify most suitable sport for beginners using data mining and anthropometric profiles. Here it proposes a clustering based approach and applies a spatial clustering technique called the Spherical Associated Keyword Space, which projects clustering result from a three-dimensional sphere to a two-dimensional (2D) spherical surface for 2D visualization. Empirical study of our approach has proved the effectiveness of clustering results. In the second part of the research, researcher describes and analyses accuracy by comparing K-means and Expectation-Maximization algorithms. Here the most affected attributes were selected using WEKA ranking method. Various sizes of Anthropometric measurements data sets were used to evaluate the algorithms and end of this section K-mean achieve highest accuracy rate. In the final part, where its aim is to measure accuracy of the given classification algorithms, prediction accuracy details of each and every data mining algorithm recorded separately. This study integrates four data mining algorithms (Naïve Bayes, Decision Trees, Random Forest, and Support Vector Machine) and an Ensemble approach (bagging, boosting, and stacking). In this part Anthropometric measurements of the tennis players are used and classification performance of these models has been evaluated using Accuracy, Precision, Recall, F-Measure, MCC, ROC Area, PRC Area, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Here for this, SVM is proposed as the most suitable algorithm to classify Anthropometric measurements of tennis players. Considering the entire research, researcher propose three different approaches to analyse Anthropometric measurements. Here it suggest to develop a sport selection algorithms using Anthropometric measurements as a future work.

Keywords: *Classification, Clustering, Data mining, Anthropometric measurement*