

K-Means Clustering Algorithm to Predict the Badulla Tomato Price Based on Weather Factors

K.T. Dananjali, J.B. Ekanayake and A.S. Karunaratne

¹*Faculty of Graduate Studies, Sabaragamuwa University of Sri Lanka*

²*Department of Physical Sciences and Technology, Sabaragamuwa University of Sri Lanka*

³*Department of Export Agriculture, Sabaragamuwa University of Sri Lanka*

Tomato is one of the most important cash crops in Sri Lanka and tomato is cultivating in several areas of the country. Among them, tomato farming in Badulla significantly contributes to the total local tomato production. However, the producer price of Badulla tomato is subjected to the fluctuation within a short period of time. Hence, farmers face great difficulties when selling their products. This study was aimed to explore the influence of weather factors on Badulla tomato price fluctuation. The data was collected from the Meteorological Department of Sri Lanka and the Hector Kobbekaduwa Agrarian and Research Institute for the past 10 years (2005-2015). These are the considered Badulla district weather factors: rainfall (BR), minimum (MinTB)/maximum (MxTB) temperature, minimum (MinRH)/maximum (MxRH) relative humidity and the farm gate price of tomato at Badulla district (BTP). The Data set was consisted only quantitative data and there were 574 instances. Analysis and investigation were done using data mining techniques. After preprocessing of data, 66% percentage from the total number of instances were considered as training data. The K-Means algorithm was used to cluster the above data vectors. The Euclidean distance function was used to compare the data vectors. The strength of K-Means clustering was validated using Elbow cluster validation technique. Five clusters were formed as the best number of clusters. Within cluster sum of squared errors: 24.68 and 15 number of iterations were performed within the clustering model. Highest Badulla tomato price centroid value was: Rs.49, other cluster centroid values were BR: 27.6mm, MxTB: 27.6 °C, MinTB: 14.2 °C, MxRH: 96.06%, MinRH: 63.1% of that cluster. As results of this research, it is possible to predict best weather conditions which are giving highest Badulla Tomato Price. That will be helpful for farmers as well as the decision makers to take correct decisions related to tomato farming.

Keyword: Clustering, K-mean, Producer price, Tomato, Weather