

M5P Model Tree in Predicting Potato Price in the Intermediate Zone in Sri Lanka

K. T. Dananjali^a, J. B. Ekanayake^b and A.S. Karunaratne^c

^a Faculty of Graduate Studies, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka

^b Department of Physical Sciences and Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka

^c Department of Export Agriculture, Faculty of Agricultural Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka
ktdananjali@gmail.com

There are three major climatic zones in Sri Lanka named as wet, intermediate and dry zones according to the average annual rainfall. Potato has become an essential food in Sri Lanka and the potato varieties grown in Welimada are also grown in intermediate zones especially in Badulla district. The intermediate zone's potato highly affects the total local potato production. However, the intermediate potato price is subjected to a price fluctuation within a short period of time and on some occasions the price variation negatively affects the farmers as well as the customers and subsequently farmers are unable to cover up their production cost. This study investigates the factors affecting the variations of production prices of potato grown at the intermediate zone of Sri Lanka, basically in the Badulla area. To that end, M5P model tree algorithms were trained to find out the triggering forces behind this issue and develop a producer price prediction model. Following features are considered as independent variables for predicting the producer price of potato grown at the intermediate zone; imported potato wholesale prices (IPW), producer prices of Nuwara Eliya potato (NPP), rainfall (MxRB), minimum temperature (MinTB), maximum temperature (MxTB), minimum relative humidity (MinRH) and maximum relative humidity (MxRH) in Badulla district, extent (ExBP) and production (PWPMT) of Welimada potato. The data were collected from the Meteorological Department of Sri Lanka, the Hector Kobbekaduwa Agrarian and Research Institute and the Census and Statistics Department of Sri Lanka, for the period from 2005 to 2015. The dataset consists of 574 instances. The data were preprocessed as there were missing values. Then the M5P models were trained and tested using 10 fold cross-validation. The prediction quality of the models is evaluated using the following factors and M5P smoothed model tree gave the best prediction quality; Correlation coefficient: 0.744, Mean absolute error: 8.2511, Root mean squared error: 10.7904, Relative absolute error: 62.6481 %, Root relative squared error: 66.7029 %, out of all the M5P models. The model gave the eleven (11) rules and according to the rules, it shows that the factors: NPP, MxTB and MinRH positively affecting and ExBP, PWPMT, IPW, MxRB, MinTB and MxRH factors negatively affecting the producer price of intermediate zone's potato. The outcome of this study is helpful for the farmers to increase their profit of potato farming at the intermediate zone by considering these factors for their cultivation planning and it is able to predict the production price of potato in the intermediate zone in Sri Lanka.

Keywords: intermediate zone, m5p, potato, price, prediction