Paper ID: 300

Analyzing the Influence of Various Factors for Vegetable Price using Data Mining

I.M.G.L. Illankoon, B.T.G.S. Kumara

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

gayanilakshi101@gmail.com

Abstract: The price fluctuation of vegetables is one of the economic problems faced by every country, including Sri Lanka. Many factors such as environmental conditions as well as supply, demand, social, cultural, and political situations of the country cause the price of vegetables to fluctuate. In this study, the Waikato Environment for Knowledge Analysis (WEKA) tool and association based Apriori algorithms used to identify the most influential factors that affect price fluctuation. Results show that the low supply from cultivation areas is caused by an increased vegetable price and favourable supply from cultivation areas is caused by decrease vegetable price. Prices of vegetable varieties demonstrated mixed movements because of supply variability from respective areas. The findings of this study can be used by farmers to make their production plans, customers to plan their budget, and sellers to make their marketing plans.

Key Words: Data mining, Association rule, Apriori algorithm, Vegetable price patterns

Introduction

Sri Lanka has over 2500 years of rich agricultural history and agriculture is one of the most important industries of the Sri Lankan economy. In Sri Lanka, rice production ranks first in terms of agriculture. In addition to rice, vegetable production also plays a major role in national economic and social stability. In Sri Lanka, more than forty different varieties of vegetables are grown in various agro-climatic areas and 602,000 metric tons of vegetables are produced annually (Sri Lanka Export Development Board, 2013). Upcountry vegetables and low country vegetables are the two types of vegetables grown in Sri Lanka. Some of them are cultivated as commercial crops. Low-land and dry-wet areas are most suitable for a variety of tropical vegetables such as green chili, ladies' fingers, eggplant, cucumber, pumpkin, red onion, bitter gourd, etc. In the upcountry, the cool and stable weather conditions are perfect for temperate crops such as carrot, cabbage, leek, salad leaves, cauliflower, beet, tomato, bell pepper, and salad cucumber. But, low country varieties are also grown in these areas. These vegetables remain an important source of income. Most of the vegetables grown in Sri Lanka are consumed locally and less than one percent is exported (Esham and Usami, 2006).

Vegetable prices in Sri Lanka are highly volatile. There are many reasons for the vegetable price fluctuation. Environmental conditions such as rainfall patterns, temperature, wind, soil, and humidity impact on vegetable price fluctuation. In addition to the above reasons, some of the other factors such as the social, cultural, economic, political situation of the country, supply, and demand can be affected by vegetable price variation.

Many researchers conduct research works based on analyze factors' impact on agricultural production and analyze price behaviour using various methods. Most of the researches has been conducted using surveys, case studies, and statistical approaches. But today, using techniques from data science to increase efficiency, especially in agriculture is very important. Data mining may help to convert raw data into meaningful information for improving agricultural uses because data mining represents a set of specific methods and algorithms aimed at extracting patterns from row data. Data mining techniques can be divided into predictive and descriptive. Descriptive data mining is used to analyze data and provide information about past and recent events. The predictive data mining provides answers to possible questions using historical data. Classification, regression, time series, and prediction can be considered as the predictive type and clustering, association rules, summarizing and sequence discovery can take as the descriptive methods (Kaur, Gulati and Kundra, 2014). Among them, association rules mining can be used to identify the association between factors such as weather conditions, supply, demand, seasonality affecting vegetable price, and price behaviours. (Rashid, Nohuddin and Zainol, 2017).

In this research, association rule-based Apriori algorithms are used to identify the patterns and association between most influential factors that affect price fluctuation and price indicates. The findings of this research will be useful for farmers to prepare and decide on production plans and improve profit, sellers to plan about the market for these vegetables, customers to plan their budget, and the government to create laws and regulations regarding cultivation and export-import concerns.

The rest of this paper is organized as follows. Section II describes the related work. Section III provides a brief overview of the research methodology and experimental design. Section IV describes the results and finally, we conclude the paper in Section V.

Related Works

Many researchers around the world have conducted many studies from different perspectives to find out the main reasons for the fluctuations in prices for different crops, price behaviours, and the factors affecting agricultural production. Research work (Armstrong and Gandhi, 2016) presented some interest rules related to the influence of distributed seasonal rainfall on rice production in Rajasthan, India. Apriori algorithm was applied on data sets and results showed that normal or good rainfall is required to get a good rice harvest. Another research (Tanna and Ghodasara, 2015) work has demonstrated how pattern mining is done using association rules for agricultural datasets. Research work (Geetha, M.C.S, 2015) discusses the role of data mining techniques in the agriculture field and their related work by implementing association rule mining for different soil types in the context of the agriculture domain. Another study by (Wankhede, Armstrong, and Gandhi, 2018) investigated the spatial and temporal variation in rainfall and jowar crop production in the Maharashtra state of India. Authors used association rule mining techniques to observe the relationship between rainfall patterns and crop productivity across districts and years.

Research Methodology

Figure 1 shows the proposed approach and it contains several steps including data collecting and creation of data sets, computerized data, and analysis, processes the data, and feature extraction.

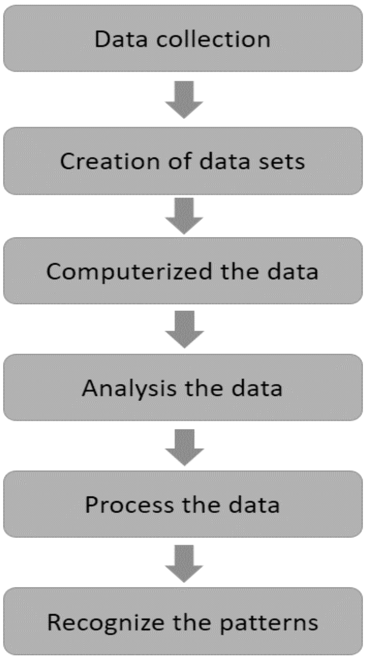

Figure 22: Research methodology

A. Collecting and Creation of the Data sets

We conducted our research based on wholesale vegetable prices in Sri Lanka and the most influential factors for the change of vegetable prices. Daily wholesale prices and average weekly wholesale vegetable prices of three vegetables namely beans, carrot, and brinjal and, the reason for the influence of vegetable price in Pettah market, Sri Lanka from June 2017 to December 2019 which were published by the Department of Census and Statistics in Sri Lanka, Central Bank of Sri Lanka and Hector Kobbakaduwa Agrarian Research and Training Institute have been utilized for this research.

To get an idea about price variation patterns, we created graphs using average weekly wholesale prices. To discover frequent associations, we used daily wholesale vegetable prices and details about influential factors.

The dataset contains four attributes and they are total wholesale vegetable price, month, influential factor, and fluctuation type.

1) Total wholesale vegetable price: The total wholesale vegetable price of beans, carrot, and eggplant were classified as very low (Rs 0 to Rs 199), low (Rs 200 to Rs 399), high (Rs 400 to Rs 699), and very high (Rs 700 to Rs 1000).

2) Month: The attribute named 'Month' specifies the name of the month.

3) Influential factor: The attribute named 'Influential factor' is categorized into weather, low supply, high supply, festival, seasonality, mix supply, and other.

4) Fluctuation type: Fluctuation type is categorized into increase, decrease, and mix movements.

B. Computerized data and analysis

Before the analysis, the collected data was prepared and computerized. The data was converted to .csv format for further applying preprocessing techniques using WEKA (Waikato Environment for Knowledge Analysis) filters. Here the dataset was checked for removing missing values and outliers.

C. Process the data and feature extraction

The technology used for extract patterns and analysis is WEKA (Waikato Environment for Knowledge Analysis) tool. WEKA contains a group of algorithms and visualization tools for data mining and predictive modeling, and also it contains graphical user interfaces (GUI) for access to those functions. WEKA supports many standard data mining tasks like data pre-processing, clustering, classification, regression, visualization, and feature selection (Frank et al., 2017). In this research, the Apriori algorithm was used to extract association rules from the dataset. Apriori algorithm is the best and most commonly used algorithm for mining frequent patterns for association rules (Hashim, Hamoud, and Awadh, 2018).

D. Association rule mining

Association rule mining is a widely used approach in data mining. Association rules are capable of finding frequent patterns, associations, and important relationships in large databases and it provides an effective scientific base for decision making. Association rule mining uses the measuring criteria support and confidence to determine the most important relationships. Support means how many times the items appear in the dataset and the confidence means the number of events if-then observations are found true.

There are various algorithms for finding association rules such as Apriori Tid, Apriori, Eclat, AIS, SETM, Apriori hybrid, and FP-growth (Mishra, Pani, and Ratha, 2019). Among them, the Apriori algorithm is the best and most commonly used algorithm for mining frequent patterns for association rules.

Results

A. Results of the Data Creation and graphical analysis

According to graphs that are created using the average weekly wholesale prices, the following figures show the price behaviours for beans, carrot, and eggplant in the year 2017, 2018, 2019, and 2020.

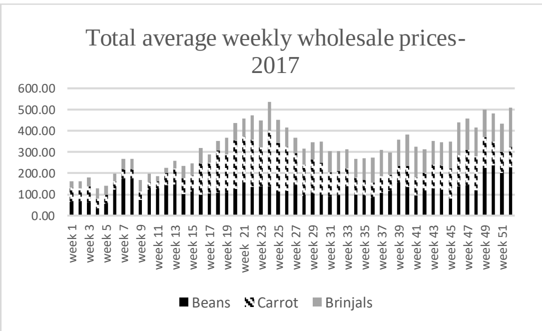

Figure 23: Price behaviours for 2017

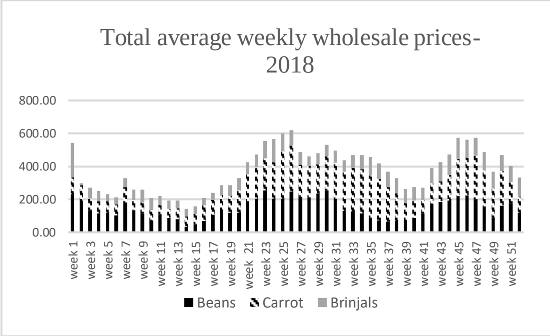

Figure 24: Price behaviours for 2018

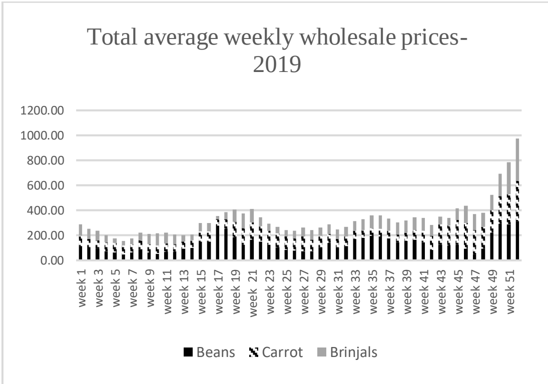

Figure 4: Price behaviours for 2019

According to the graphical analysis, from November to February and from May to July selected vegetables indicate high prices. Furthermore, prices dropped from August to October and from March to April.

And the collected data about vegetable price and influence factors consist of detailed descriptions that data cannot analyze using data mining algorithms. Under this step, data were summarized using abbreviations. Table 1 shows those abbreviations.

Table 2: Abbreviation

Attribute	Description	Abbreviation
Total price	Very low (Rs 0 to Rs 199)	VL
	Low (Rs 200 to Rs 399)	L
	High (Rs 400 to Rs 699)	H
	Very high (Rs 700 to Rs 1000)	VH
Influential factors	Weather condition	w
	High supply	hs
	Low supply	ls
	Festivals	f
	Seasonality	se
	Supply variability	s
	Other reasons	otr
Fluctuation type	Increase price	i
	Decrease price	d
	Mixed movements	m

The following figure shows some parts of the created dataset before the analysis.

No.	1: price Nominal	2: month Nominal	3: factor Nominal	4: fluctuation Nominal
1	H	may	otr	m
2	H	may	otr	i
3	H	may	otr	i
4	H	may	otr	i
5	H	jun	w	i
6	H	jun	w	i
7	H	jun	w	i
8	H	jun	w	i
9	H	jun	w	i
10	H	jun	w	i
11	H	jun	w	i
12	H	jun	w	i
13	H	jun	s	m
14	H	jun	s	m
15	H	jun	s	m

Figure 5: Created dataset

B. Results of the data analyze

The created dataset has analyzed using WEKA. Before analyzing, WEKA pre-processing used to remove missing values from the data set. The following figures visualized all the data after analyzing.

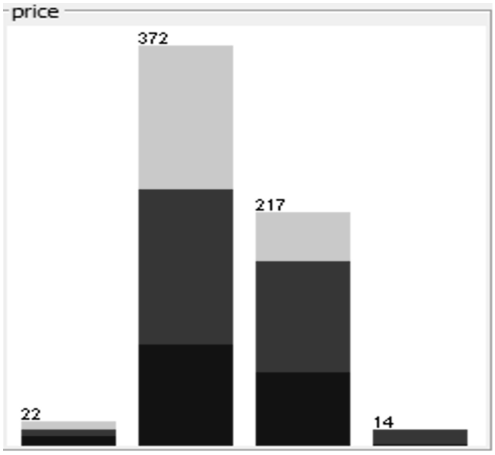

Figure 6: Data visualizing (price)

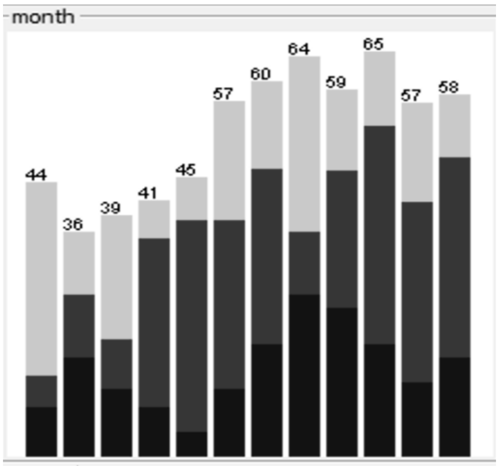

Figure 7: Data visualizing (month)

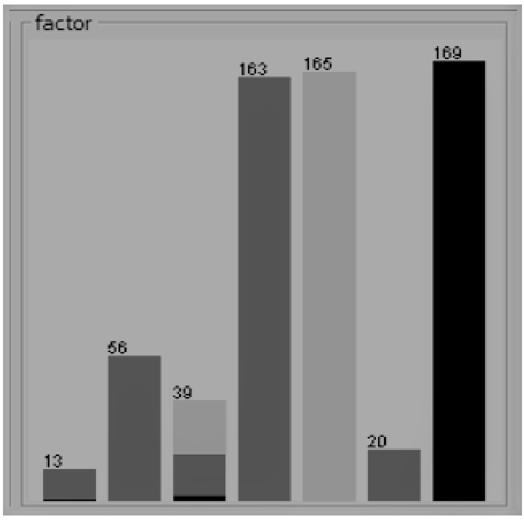

Figure 8: Data visualizing (factor)

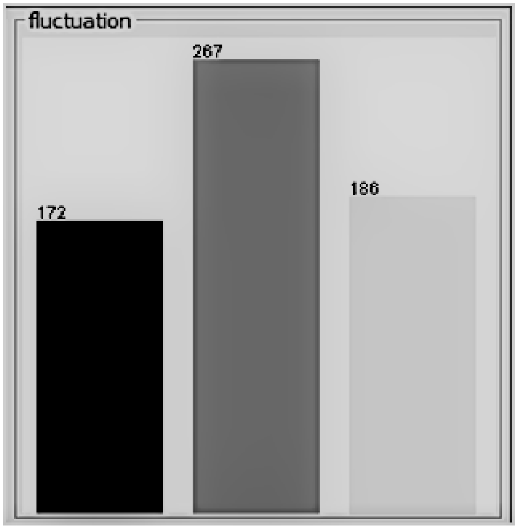

Figure 9: Data visualizing (fluctuation)

C. Association Mining

The best rules were extracted from the dataset using the WEKA tool. Table 3 shows the best rules found in the Apriori algorithm. The results depend on the comparison of confidence, leverage, and convince. The measure of minimum support performed over Apriori was 0.1 (70 instances), and minimum confidence was 0.91.

Table 3: Association model for Apriori algorithm

Minimum support	Minimum metric <confidence>	Number of cycles performed
0.1 (62 instances)	0.9	18

After the Apriori algorithm is executed, we can obtain many results. According to Rule 1 and Rule 9, wholesale prices of vegetable varieties demonstrated mixed movements (m) because of supply variability (s) from respective areas. Rule 2 describes that prices of vegetable varieties decreased (d) due to favorable supply (hs) from respective areas. Rule 3 explains that the price of vegetable varieties increased (i) due to low supply (ls) from respective areas. According to Rule 4, prices of vegetable varieties decreased (d) due to favorable supply (hs) from respective areas, and the total price of carrot, beans, and eggplant can fluctuate between Rs.200 to Rs.399 in each day. Rule 5 displays the price of vegetable varieties increased (i) due to low supply (ls) from respective areas and the total price of carrot, beans, and eggplant can fluctuate between Rs.200 to Rs.399 in each day. Total wholesale prices of carrot, beans, and eggplant can fluctuate between Rs.200 to Rs.399 and it demonstrated mixed movements (m) because of supply variability (s) from respective areas according to Rule 6 and Rule 8. Finally, Rule 7 and Rule 10 display that, wholesale prices of vegetable varieties demonstrated mixed movements (m) because of supply variability (s) from respective areas, and the total price of carrot, beans, and eggplant can fluctuate between Rs.400 to Rs.699 in these days.

Table 3: Best rules obtained after applying Association Rule

No.	Rule
1	factor=s 169 ==> fluctuation=m 169 <conf:(1)> lift:(3.65) lev:(0.2) [122] conv:(122.69)
2	factor=hs 165 ==> fluctuation=d 165 <conf:(1)> lift:(3.35) lev:(0.19) [115] conv:(115.82)
3	factor=ls 163 ==> fluctuation=i 163 <conf:(1)> lift:(2.34) lev:(0.15) [93] conv:(93.25)
4	price=L factor=hs 117 ==> fluctuation=d 117 <conf:(1)> lift:(3.35) lev:(0.13) [82] conv:(82.13)
5	price=L factor=ls 100 ==> fluctuation=i 100 <conf:(1)> lift:(2.34) lev:(0.09) [57] conv:(57.21)
6	price=L factor=s 93 ==> fluctuation=m 93 <conf:(1)> lift:(3.65) lev:(0.11) [67] conv:(67.51)
7	price=H factor=s 67 ==> fluctuation=m 67 <conf:(1)> lift:(3.65) lev:(0.08) [48] conv:(48.64)
8	price=L fluctuation=m 94 ==> factor=s 93 <conf:(0.99)> lift:(3.65) lev:(0.11) [67] conv:(34.27)
9	fluctuation=m 171 ==> factor=s 169 <conf:(0.99)> lift:(3.65) lev:(0.2) [122] conv:(41.56)
10	price=H fluctuation=m 68 ==> factor=s 67 <conf:(0.99)> lift:(3.64) lev:(0.08) [48] conv:(24.79)

Discusion and Conclusion

In this paper, we aimed to find patterns related to the most influential factors that affect price fluctuation and price indicates. In this work, we do our research based on the average weekly and daily wholesale prices of three vegetables namely beans, carrot, and eggplant, and reason for influence vegetable price from 2017 to 2019. We have taken data that contains attributes such as total price, month, influential factor, and fluctuation type.

According to the graphical analysis of average weekly wholesale vegetable prices, it implies that the price of vegetables can fluctuate from a higher value to a lower value on any given day and that is not possible to provide an exact price for vegetables.

The results of applying the association rule-based Apriori algorithm through WEKA is showing that low supply from cultivation areas is caused by increase vegetable price and favourable supply from cultivation areas is caused by decrease vegetable price. Wholesale prices of vegetable varieties demonstrated mixed movements because of supply variability from respective areas.

In future work, we planned to evaluate our approach with classification methods. We also plan to implement the vegetable price prediction approach by considering the identified factors.

Acknowledgment

We pay our sincere gratitude to the staff of the Department of Computing and Information Systems, Sabaragamuwa University of Sri Lanka.

Author Biographies

I.M.G.L.Illankoon is a degree candidate for the Department of Computing and Information Systems, Sabaragamuwa University of Sri Lanka. She applies her knowledge about Data Science for this research.

Banage T.G.S. Kumara received a Bachelor's degree in 2006 from Sabaragamuwa University of Sri Lanka. He received a master's degree in 2010 from the University of Peradeniya and a Ph.D. degree in 2015 from the University of Aizu, Japan. His research interests include semantic web, web data mining, web service discovery, and composition.

References

Armstrong, L. J. and Gandhi, N. (2016) 'Assessing impact of seasonal rainfall on rice crop yield of Rajasthan, India using Association Rule Mining', pp. 1021–1024.

Esham, M. and Usami, K. (2006) 'Procurement behavior of the fruit and vegetable industry in Sri Lanka', Journal of Agricultural Sciences, 2(3), p. 36. doi: 10.4038/jas.v2i3.8135.

Frank, E. et al. (2017) 'The WEKA Data Mining Software: An Update', 15(4), pp. 452–465. doi: 10.1080/14733285.2016.1271943.

Geetha, M.C.S (2015) 'Implementation of Association Rule Mining for different soil types in Agriculture', Ijarcce, 4(4), pp. 520–522. doi: 10.17148/ijarcce.2015.44119.

Hashim, A. S., Hamoud, A. K. and Awadh, W. A. (2018) 'Analyzing Students' Answers Using Association Rule Mining Based on Feature Selection', Journal of Southwest Jiaotong University, 53(5), pp. 1–16.

Kaur, M., Gulati, H. and Kundra, H. (2014) 'Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications', International Journal of Computer Applications, 99(12), pp. 1–3. doi: 10.5120/17422-8273.

Mishra, A. K., Pani, S. K. and Ratha, B. K. (2019) 'Association Rule Mining with Apriori and FPGrowth Using WEKA', International Journal of Advanced Technology in Engineering and Science, (September 2015).

Rashid, R. A. A., Nohuddin, P. N. E. and Zainol, Z. (2017) 'Association Rule Mining Using Time Series Data for Malaysia Climate Variability

Prediction', (October). doi: 10.1007/978-3-319-70010-6.

Sri Lanka Export Development Board (2013) 'Industry Capability Report Sri Lankan Fresh Fruit & Vegetable', (January), p. 10.

Tanna, P. and Ghodasara, Y. (2015) 'Using Apriori with WEKA for Frequent Pattern Mining', International Journal of Engineering Trends and Technology (IJETT), Volume 12(June 2014). doi: 10.14445/22315381/IJETT-V12P223.

Wankhede, S. S., Armstrong, L. J. and Gandhi, N. (2018) 'Characterising the influence of drought on Jowar crop production in India using association rule mining', Proceedings - 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development, TIAR 2017, 2018-January(TIAR), pp. 57–63. doi: 10.1109/TIAR.2017.8273686.