Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 9, No. 3, 2021, 341-349.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

Investigating Changes in Household Consumable Market Using Data Mining Techniques

Atefeh Hasan-Zadeh*, Faezeh Asadi, Najmeh Garbazkar

Fouman Faculty of Engineering, College of Engineering, University of Tehran, Fouman, Iran.

Article Info

Article History:
Received 01 September 2020
Revised 04 November 2020
Accepted 05 June 2021

DOI:10.22044/jadm.2021.10024.2139

Keywords:
Data Mining, Bayesian Rule, Decision Tree, Associative Rule, Households' Consumer Goods.

**Corresponding author:*
hasanzadeh.a@ut.ac.ir (A. Hasan-Zadeh).

Abstract

For an economic review of the food prices in May 2019 in order to determine the trend of rising or decreasing prices compared to the previous periods, we consider the price of food items at that time. The types of items consumed during specific periods in the urban areas and the whole country are selected for our statistical analysis. Among the various methods of modelling and statistical prediction, and in a new approach, we model the data using the data mining techniques consisting of the decision tree methods, associative rules, and Bayesian law. Then the prediction, validation, and standardization of the accuracy of the validation are performed on them. The results of data validation in the urban and national area and the results of standardization of the accuracy of validation in the urban and national areas are presented with the desired accuracy.

1. Introduction

Investigating the economic conditions of a statistical society and providing measures to curb inflation, reduce or eliminate inflation, if possible, achieve the optimal and ideal conditions, and subsequently, increase the economic growth is of particular importance [1-3].

In economics and industrial engineering, new methods are used in order to address this issue including new data mining techniques. More specifically, data mining is used in the areas of customer relationship management [4, 5], preventive maintenance, [6, 7], supply chain management [8, 9], production planning, [10, 11], quality control [12], project management [13], safety, health, and environment [14-18]. Such studies and explorations can be regarded as the continuation of the ancient and ubiquitous knowledge of statistics. The major differences are in the scale, breadth, and a variety of contexts, applications, dimensions, and sizes of the today's data that demand machine learning, modelling, and training.

In this work, we conducted an economic study of food prices in May 2019 in order to determine the trend of rising or decreasing prices relative to the definite periods, which might or might not lead to inflation.

The data mining methodologies and the RapidMiner software were utilized in order to identify future behaviors to predict the presence or absence of inflation and other factors based on the past behaviors. This technique refers to the extraction of the hidden information or patterns and relationships in large volumes of data in one or more large databases. It analyzes the databases and large datasets as they are discovered and extracted. The methodology is analyzed step by step from the associative rules, Bayesian rule, and decision tree. The validation results are also presented in the form of figures and tables at each stage.

2. Statement of Problem

Initially, the basic food price information for the mentioned period was obtained from the Iranian Statistical Site (https://www.amar.org.ir/). The problem under consideration is modelling, forecasting, validation, and standardization of the accuracy of the validation of market changes related to food items in May 2019 in both the urban and national areas. Our goal is to predict the changes in the food market in the urban and national areas in order to determine whether or not the statistical population is economically stable.

3. Inputs of Problem

The input data are comprised of the price of food items and their variations over specific periods, as illustrated in Tables 1 and 2.

Table 1. Index of national households' consumer goods and services price index by main sections and some groups and commodity classes in May 2019.

Description	Coefficient of significance	Index	Percentage of change of index compared to the previous month	Percentage of change of index compared to the same month last year	Percentage of change in the twelve-month index ending this month compared to the same period last year
Total index	100	173.5	1.5	52.1	34.2
1- Foods and beverages	26.64	216.7	-0.8	81.7	49.7
Foods	25.55	217.3	-1	82.3	49.7
Bread and cereals	6.67	159.8	7.4	36.7	23.7
Red and white meat and their products	5.91	253.8	-5.3	101.5	63.9
Red meat and poultry meat	5.14	257.1	-6.1	100.7	63.7
Fish and shellfish	0.77	231.5	0.9	107.1	65.7
Milk, cheese, and eggs	2.8	174.9	0.7	47	39.6
Oils and fats	1.29	178.8	3.4	56.1	39.2
Fruit and nuts	3.43	250.7	0.4	112.6	74.6
Vegetables (vegetables and beans)	3.02	288.4	-10.7	136.2	59
Sugar, jam, honey, chocolate, and pastries (sugar loaf, sugar, and sweets)	1.44	203.2	14.5	85.8	36.6
Food products not classified elsewhere	0.98	243.3	4.3	120.7	78.4
Tea, coffee, cocoa, soft drinks, and juices (non-alcoholic drinks)	1.1	202.8	4.2	69	50
2- Tobacco	0.59	277.3	1.3	119.6	118.8
3- Clothing and shoes	4.78	179.8	3.5	60.8	38.4
4- Housing, water, electricity, gas, and other fuels	35.5	140.2	1	25.4	19.6
Housing	31.12	141.2	0.6	24.9	20.9
Rent	30.72	141.1	0.6	24.8	20.9
Residential unit maintenance (service)	0.41	149.6	3.4	34.6	23.4
Water, electricity, and fuel	4.38	132.9	4	28.8	10.5
5- Furniture, home appliances, and their usual maintenance	3.93	206.5	4.4	83.2	55.4
6- Health and treatment	7.14	144.1	1.8	28.6	20.5
7- Transportation	9.41	177.7	7.8	62.4	35.8
8- Communication	2.87	146.4	1.1	36.7	28.7
9- Recreation and culture	1.65	205.8	4.9	75.7	54.1
10- Education	1.86	144.4	0.1	21.3	18.9
11- Hotel and restaurant	1.44	168.8	2.5	50.3	28.7
12- Miscellaneous goods and services	4.18	189.1	3.1	58.9	42.5

Table 2. Index of urban households' consumer goods and services price index by main sections and some groups and commodity classes in May 2019.

Description	coefficient of significance	Index	Percentage of change of index compared to the previous month	Percentage of change of index compared to the same month last year	Percentage of change in the twelve-month index ending this month compared to the same period last year
Total index	100	172	1.6	50.7	33.7
1- Foods and beverages	24.53	217.4	-0.6	81.8	50.1
Foods	23.57	218.1	-0.8	82.4	50.2
Bread and cereals	5.98	161.6	8.3	38.4	23.7
Red and white meat and their products	5.54	255	-5.4	102.3	64.6
Red meat and poultry meat	4.79	258.6	-6.2	101.6	64.4
Fish and shellfish	0.76	231.9	1	107.4	65.6
Milk, cheese, and eggs	2.7	176.7	0.7	48.3	40.6
Oils and fats	1.15	176.1	3.1	52.5	37.3
Fruit and nuts	3.31	252.2	1	111	74.9
Vegetables (vegetables and beans)	2.7	284.1	-11	133.5	58.1
Sugar, jam, honey, chocolate, and pastries (sugar loaf, sugar, and sweets)	1.3	200.7	13.8	82.2	36.3
Food products not classified elsewhere	0.89	246.4	4.4	120.7	78.5
Tea, coffee, cocoa, soft drinks and juices (non-alcoholic drinks)	0.96	200.5	4.2	66.4	48.9
2- Tobacco	0.5	270.5	1.4	114	114.1
3- Clothing and shoes	4.52	179.2	3.5	60.4	38
4- Housing, water, electricity, gas, and other fuels	38.07	140.8	0.9	25.7	20/3
Housing	34.1	141.8	0.6	25.4	21.4
Rent	33.73	141.6	0.6	25.2	21.3
Residential unit maintenance (service)	0.37	154.2	3.3	37.9	25.8
Water, electricity, and fuel	3.97	132.3	4	28.5	11
5- Furniture, home appliances, and their usual maintenance	3.64	203.6	4.3	80.8	53.8
6- Health andt	7.13	144.3	1.8	28.8	20.7
7- Transportation	9.44	180.5	8.2	64.8	37.1
8- Communication	2.85	146.8	1.2	37.2	28.9
9- Recreation and culture	1.64	205.7	4.9	75.2	54.2
10- Education	2.02	144.3	0.1	21.5	19.1
11- Hotel and restaurant	1.54	168.3	2.3	50	28/6
12- Miscellaneous goods and services	4.13	189.6	3.1	58.8	42.7

4. Problem-solving Steps

4.1. Step 1: Predictor Models

All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified. The predictor models are a set of machine learning techniques that discover patterns in large datasets to perform accurate predictions in new situations, and are summarized in Figure. 1.

As shown in this figure, the Bayesian rule, decision tree, and associative rules are the categorical models. In each data mining method, the training dataset is assigned to one of these classification algorithms in order to build the model in question.

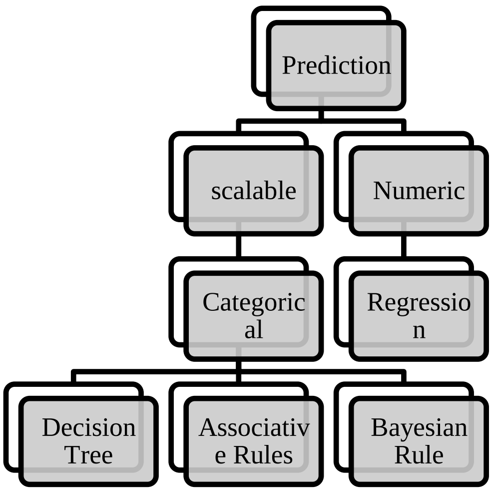


Figure 1. Types of predictor models.

The Bayesian rule is a method of classifying phenomena based on the probability of occurrence or non-occurrence of a phenomenon, which is applicable in the probability theory. If we can partition a given sample space so that its members can determine which part of the sample space is affected by an event, a significant portion of the uncertainties will be reduced. This rule is useful because it can calculate the probability of an event occurring on the condition of the occurrence or non-occurrence of another event. In many cases, it is difficult to directly calculate the probability of an event. By applying this law and conditioning one event to another, one can calculate the probability. The decision tree presents the output model as a tree of different states of attribute values. The decision tree categories are fully interpretable. The benefits of the decision tree are the ability to work with the discrete and continuous data, ease of describing conditions, absence of necessity for distribution estimation performance, and unexpected or unknown relationships.

The disadvantage of the decision tree is that it is costly to produce if the nodes overlapping the number of end nodes are increased and the wrong relationships are possible. The associative rules represent their output knowledge as a set of "if-then" rules. Each rule has a condition section (LHS: Left Hand Side) and a result section (RHS: Right Hand Side). If all the requirements of the first section of a rule about a particular record are correctly interpreted, the rule covers the record. The two criteria of accuracy and coverage can be calculated for each rule; the higher the two criteria for each rule, the more valuable the rule is. The coverage section of the rule is equal to the percentage of records to which the condition part of the rule in question applies. Thus the higher the value, the more general is the rule. The accuracy part of the rule states that among the records to which the terms of the rule apply, a certain percentage of both parts of the rule in question is true.

4.1.1. Pre-processing of Input Data

Pre-processing of the data should be done first so that the data with missing values is deleted, moved or completely erased. In addition, if the data type is a real number, it changes to integers.

The point is that the data must be labelled before modelling. In Tables 1 and 2, the percentage of the index change compared to the previous month, which has positive and negative values, is considered as the problem labels.

As a result of the calculations done in RapidMiner, the following results were obtained in the urban and national areas. The processes performed on the urban and national data are shown in Figure 2.

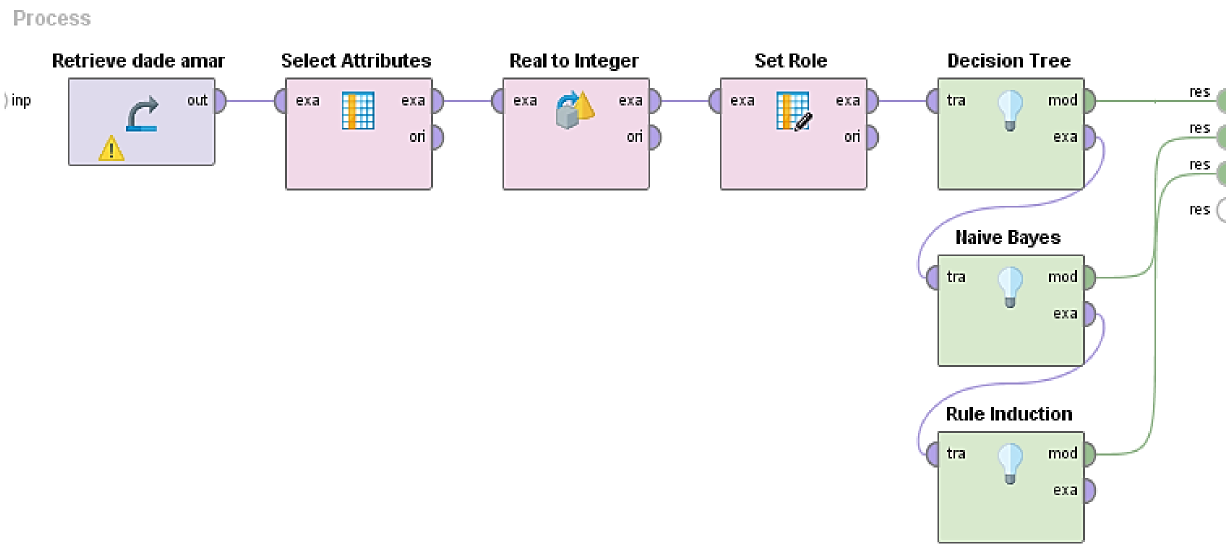


Figure 2. Processes for data modelling in urban and national areas in RapidMiner.

4.1.2. Results Obtained from Categorization Models in Urban Areas

The results in the metropolis are as follow:

4.1.2.1. Associative Rules

Analysis of the associative rules used in the RapidMiner software in the metropolitan areas indicates that if the index value is less than or equal to 211, all samples will experience a boom in price (inflation).

Also if the percentage change of index is 64.5% smaller than the preceding 2 years, none of the samples will have inflation.

In other cases (three more cases), we also experience inflation.

4.1.2.2. Bayesian Rule
By analyzing the Bayesian rule results, we concluded that the probability of inflation was 0.5% and the probability of non-inflation was 0.7%, and so the situation is not economically desirable.

4.1.2.3. Decision Tree
Analysis of the decision tree results shows that the inflation is 100% when the percentage change of index is less than 49 years compared to the preceding two years. Also in cases where the percentage change in the index is more than 49, depending on the coefficient of significance, if the coefficient is less than 3.5, there is very little inflation, and if the coefficient is more than 3.5, there is inflation.

4.1.3. Results Obtained from Categorization Models in National Areas
The results in the national domain are as follow:

4.1.3.1. Associative Rules
Analyzing the results of associative rules shows that if the index value is less than or equal to 211, all samples will experience an unexpected rise in price (inflation). Also if the percentage change of the twelve-month index ended to the current month is 64% smaller than the same period in the last year, none of the samples will have inflation.
In other cases (three more cases), we also experience inflation.

4.1.3.2. Bayesian Rule
An analysis of the Bayesian rule results shows that the probability of inflation is 0.5% and the probability of non-inflation is 0.5%, so the situation is not economically desirable.

4.1.2.3. Decision Tree
Analysis of the decision tree results shows that the inflation is 100% when the percentage change is less than or equal to 45.5 compared to the last two years. Also in cases where the percentage change in the index is more than 45.5 two years ago, given the significance coefficient, if the coefficient is less than 4, we have a very low inflation, and if the coefficient is more than 4, the inflation is present. Finally, in order to check the accuracy of the results, the charts can be checked in the RapidMiner program, and based on different criteria, the results can be correctly applied.

4.2. Step 2: Prediction of New Data
Unlabeled training data are used to predict the data. The process of predicting the new data is called scoring. With automatic scoring enabled, the operating model is built, and then with the help of the operational model, decisions can be made better and faster. The Bayesian law is used to predict the unlabeled data. The processes used to predict the urban and national data in the RapidMiner program, as shown in Figure 3, are as follow:

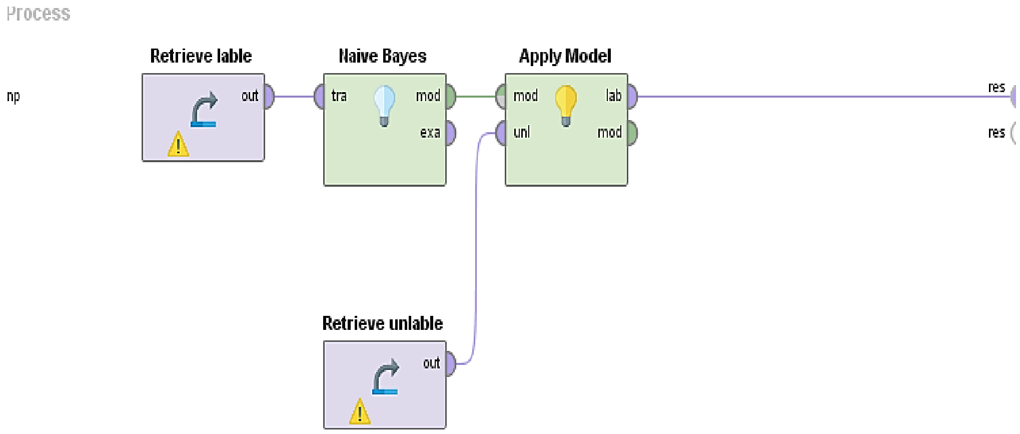


Figure 3. Processes used to predict the urban and national data in RapidMiner.

The urban forecasting results include predicting the actual rate of inflation (percentage change of index compared to the previous month) and obtaining a confidence percentage for predicting yes or no. The comparative diagram obtained from the urban and national forecasting process in RapidMiner is illustrated in Figure 4.

4.3. Step 3: Data Separation and Validation
Data separation is used in order to measure the accuracy of the prediction model. For testing, part of the data whose results are known is used to measure the model's future performance. As such, we hold a piece of the labelled data for testing, which has not been tested by the model. Finally, it is clear how well the model has predicted. The prediction model response is compared with the

actual response in order to determine the correct prediction percentage of the model.

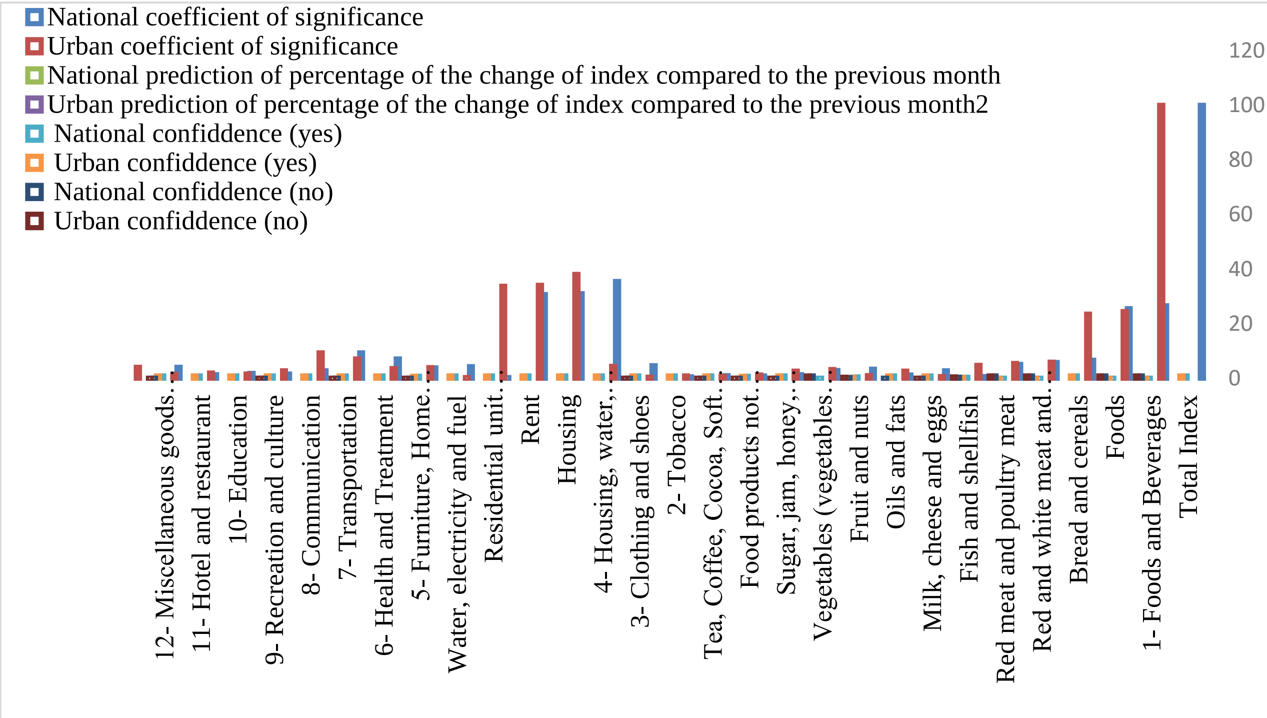


Figure 4. Comparative graph obtained from the urban and national prediction processes in RapidMiner

Notably, the training data is not employed in the model testing because it overestimates the accuracy of the model to 100%. In the calculations, the ratio of 70:30 is used, which means that 70% of the data is used to learn the algorithm and 30% for testing. The processes used for data separation and validation in the urban and national areas in the RapidMiner program are shown in Figure 5.

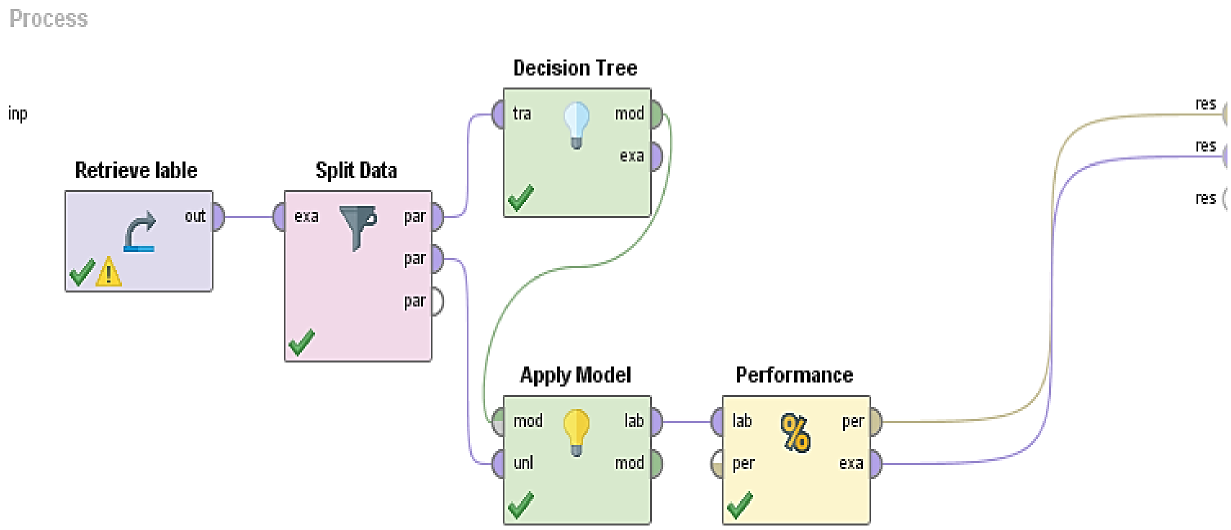


Figure 5. Processes used to separate and validate the data in urban and national areas in RapidMiner.

The comparative diagram obtained from the urban and national validation processes in RapidMiner is shown in details in Figure 6.

4.3.1. Results of Data Validation in Urban Area

The results in the metropolitan area are summarized in Table 3. The analysis of the results in Table 3 shows that in the first prediction, the probability of “yes” is 100% and the probability of “no” is 0%, and the other cases are analyzed afterwards. The predictive model also performed well at 88.9%. More precisely, in the seven models tested, the true value is yes and the prediction model is the same for the price increase. Also the prediction accuracy of the class “yes” is 100%, meaning that all the predictions are correct.

In one case, the tested model had a true value of "yes" but predicted "no", and in the other, the true value was "yes" and correctly predicted, i.e. the accuracy of the prediction of the "no" class is 50%.

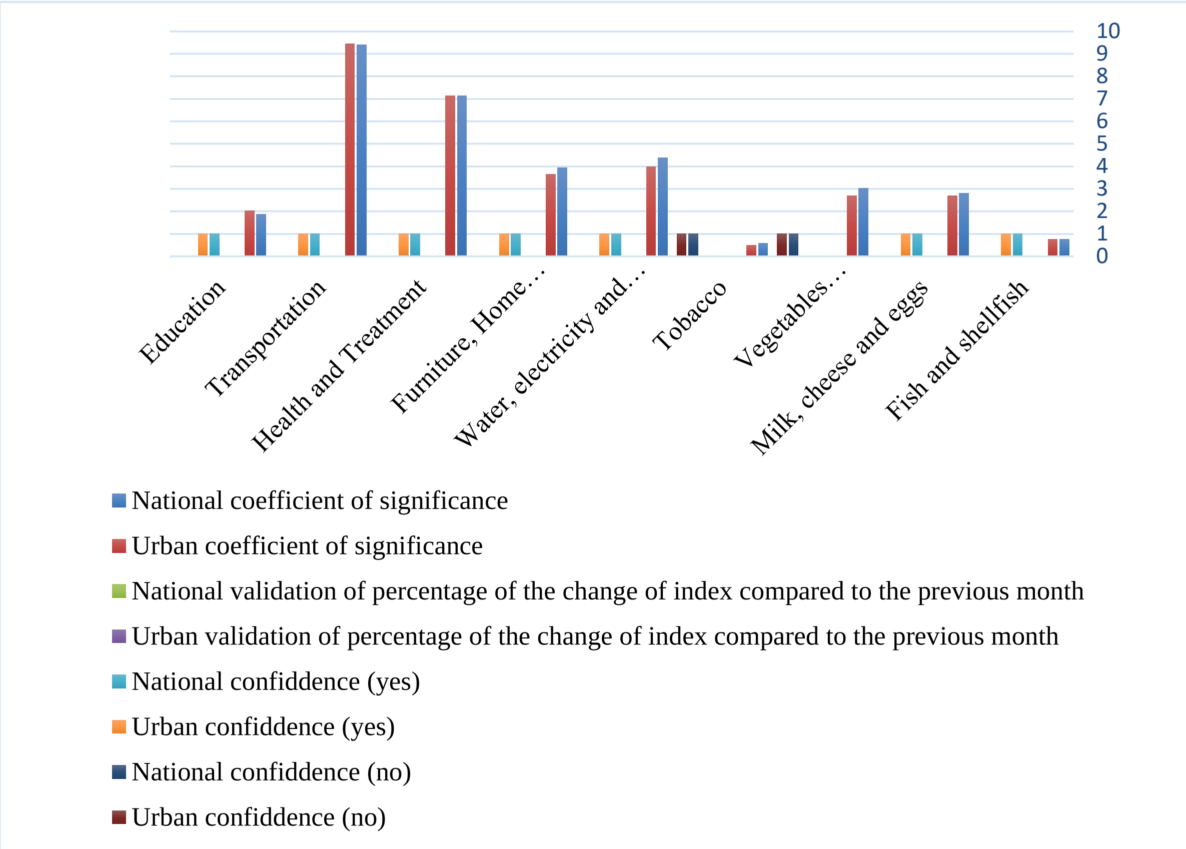


Figure 5. Comparative graph obtained from the urban and national prediction processes in RapidMiner.

Table 3. Results of data validation in urban area.

			Accuracy: 88.89%
	True yes	True no	Class precision
Pred. yes	7	0	100.00%
Pred. no	1	1	50.00%
Class recall	87.50 %	100. 00%	

Concerning the accuracy of the prediction of the result, the prediction accuracy of "yes" is 87.50%, which predicted seven cases correctly and one case was predicted incorrectly. Also the accuracy of the prediction of "no" is 100%, which means that the model correctly predicted the class of "no". One out of nine cases was wrongly predicted and the other cases were correctly predicted, meaning that the model had 11.11% error.

4.3.2. Results of Data Validation in National Area

The results in the metropolitan area are summarized in Table 4. Analysis of the results from Table 4 shows that for example in the first prediction, the probability of being "yes" is 100%, and the probability of "no" is 0%; then the other cases are analyzed subsequently.

The prediction model performed well at 88.89%. Analysis of the results is quite similar to Section 4.3.1.

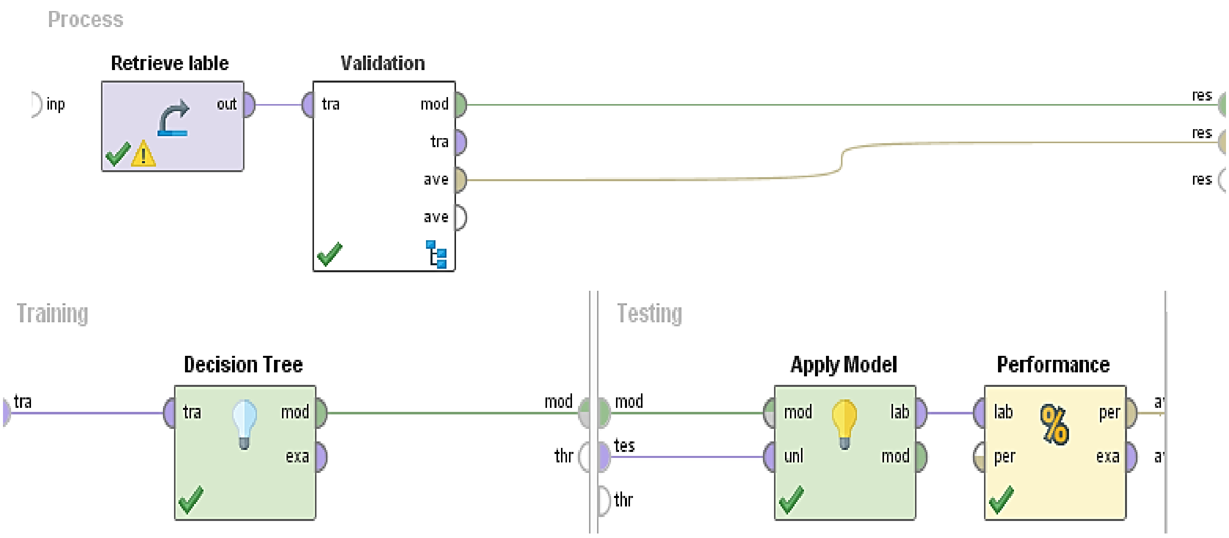
Table 4. Results of data validation in national area.

			Accuracy: 88.89%
	True yes	True no	Class precision
Pred. yes	7	0	100.00%
Pred. no	1	1	50.00%
Class recall	87.50 %	100. 00%	

4.4. Step 4: Standardization of Accuracy of Validation

Standardization of the accuracy of validation is concerned with the extent to which the data tested is similar to the training data, and with enhancing accreditation to a more reliable estimate.

The processes followed to standardize the accuracy of the urban and national levels in the RapidMiner program are illustrated in Figures 6a and 6b.



Figures 6a and 6b. Processes of standardizing accuracy of validation in urban and national areas in RapidMiner.

4.4.1. Results of Standardization of Accuracy of Validation in Urban Area

The results in the metropolitan area are summarized in Table 5.

Table 5. Results of standardization of accuracy of validation in urban area.

			Accuracy: 88.89%
	True yes	True no	Class precision
Pred. yes	21	1	95.45%
Pred. no	3	4	57.14%
Class recall	87.50 %	80.0 0%	

Analysis of the results from Table 5 shows that the data can be divided into 10 equal parts each time, one part of which is used for testing and the remaining 9 parts for training. We calculate the performance percentage each time and take it as the modelling accuracy.

As shown in Table 5, a factor is added as a standard deviation (33 16.33%) to the mean value (86.67%) as a standard deviation is generated each time the above steps are taken.

In 21 items of the cases tested, the true value was “yes”, and the prediction has come to the same conclusion: rising prices and inflation. In 1 item, the actual value was “no” but the prediction result is “yes”. In other words, the prediction accuracy of class “yes” is 95.45%.

Also in 3 items of the tested cases, the true value was “yes” but the prediction result was “no”, and in 4 cases the true value was “yes” and it was correctly predicted, so the prediction accuracy of class “yes” is 57.14% (Table 5).

Consequently, the prediction accuracy of “yes” is 87.50% (21 correct predictions and 3 false predictions) and the accuracy of the prediction of “no” is 80% (4 correct and 1 wrong predicted).

4.4.2. Results of Standardization of Accuracy of Validation in the National Area

The results obtained are shown in Table 6.

Table 6. Results of standardization of accuracy of validation in national area.

			Accuracy: 88.89%
	True yes	True no	Class precision
Pred. yes	21	1	95.45%
Pred. no	3	4	57.14%
Class recall	87.50 %	80.0 0%	

In analyzing the results of Table 6, the data is divided into 10 equal parts each time, one part of which is used for testing and the remaining 9 parts for training. We calculate the performance percentage each time and take it as the modelling accuracy. Other interpretations are exactly similar to Section 4.4.1.

5. Conclusions

The studies on measuring the market changes in the urban and national markets in May 2019 use the data mining techniques in RapidMiner. For this purpose, the three techniques of associative rules, Bayesian rule, and decision tree were used, which represented the same result in different ways.

By examining the urban and national results, we found that the market for the urban and national foodstuffs was affected by inflation over a set period, and the situation was not suitable from an economic viewpoint w because the percentages assigned to the price increase were higher than the percentages devoted to non-changes of prices.

The comparative graph obtained from the urban and national prediction processes in RapidMiner through the associative rule, Bayesian rule, and decision tree show the accuracy and efficiency of this methodology.

In the tables of input data obtained from the Iran Statistics site, it is also clear that the percentage values of change in the index of goods compared to the previous month, the year before, and two years earlier were significant. The result of the calculations also confirms its accuracy. The results of data validation in the urban and national areas and the results of the standardization of the accuracy of validation in the urban and national areas were presented with a desired accuracy.

References

[1] A. R. De Carvalho, R. S. M. Ribeiro, and A. M. Marques, "Economic development and inflation: a theoretical and empirical analysis," *International Review of Applied Economics*, vol. 32, no. 4, pp. 546-565, 2018.

[2] L. Katusiime, "Private Sector Credit and Inflation Volatility," *Economics*, vol. 6, no. 2, pp. 1-13, 2017.

[3] L. Zhao, J. Mbachu, and Z. Liu, "Identifying Significant Cost-Influencing Factors for Sustainable Development in Construction Industry using Structural Equation Modelling," *Mathematical Problems in Engineering*, vol. 2020, 4810136, 16 pages, 2020.

[4] E. W. T. Ngai, L. Xiu, and D. C. K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2592-2602, 2009.

[5] A. Zarei, M. Maleki, D. Feiz, and M. A. Siahsarani kojuri, "Competitive Intelligence Text Mining: Words Speak," *Journal of AI and Data Mining*, vol. 16, no. 1, pp. 79-92, 2018.

[6] C. J. Romanowski, and R. Nagi, *Analyzing Maintenance Data using Data Mining Methods, Part of the Massive Computing book series (MACO, volume 3): Data Mining for Design and Manufacturing*, Kluwer Academic Publishers, pp. 235-254, 2001.

[7] B. Grabot, "Rule mining in maintenance: Analyzing large knowledge bases," *Computers and Industrial Engineering*, vol. 139, 15 pages, 2020.

[8] R. Y. Zhong, S. T. Newman, G. O. Huang, and S. Lan, "Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives," *Computers and Industrial Engineering*, vol. 1021, pp. 572-591, 2016.

[9] M. E. Kara, S. Ü. O. Fırat, and A. Ghadge, "A data mining-based framework for supply chain risk management," *Computers and Industrial Engineering*, vol. 139, 12 pages, 2018.

[10] P. Vazan, D. Janikova, P. Tanuska, M. Kebisek, and Z. Cervenanska, "Using data mining methods for manufacturing process control," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 6178-6183, 2017.

[11] C. Yu, W. Zhang, X. Xu, Y. Ji, and S. Yu, "Data mining based multi-level aggregate service planning for cloud manufacturing," *Journal of Intelligent Manufacturing*, vol. 29, no. 6, pp. 1351–1361, 2018.

[12] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data Mining and Analytics in the Process Industry: The Role of Machine Learning," Book: Data-Driven Monitoring, Fault Diagnosis, and Control of Cyber-Physical Systems, IEEE Access, vol. 5, pp. 20590-20616, 2017.

[13] B. T. Hazen, J. B. Skipper, C. A. Boone, and R. R. Hill, "Back in business: operations research in support of big data analytics for operations and supply chain management," *Annals of Operations Research*, vol. 270, no. 1-2, pp. 201-211, 2018.

[14] R. Ghousi, "Applying a decision support system for accident analysis by using data mining approach: A case study on one of the Iranian manufactures," *Journal of Industrial and Systems Engineering*, vol. 8, no. 3, pp. 59-76, 2015.

[15] S. Shoorabi Sani, "A case study for application of fuzzy inference and data mining in structural health monitoring," *Journal of Artificial Intelligence and Data Mining*, vol. 6, no. 1, pp. 105-120, 2018.

[16] T. Ahmad, and H. Chen, "Short and medium-term forecasting of cooling and heating load demand in building environment with data-mining based approaches", *Energy and Buildings*, vol. 166, no. 1, pp. 460-476, 2018.

[17] R. Torkaman, and R. Safabakhsh, "Robust Opponent Modeling in Real-Time Strategy Games using Bayesian Networks," *Journal of Artificial Intelligence and Data Mining*, vol. 7, no. 1, 149-159, 2019.

[18] M. Gul, F. Guneri, F. Yilmaz, and O. Celebi, "Analysis of the relation between the characteristics of workers and occupational accidents using data mining," *The Turkish Journal of Occupational/Environmental Medicine and Safety*, vol. 1, no. 4, pp. 102-118, 2016.

بررسی تغییرات در بازار مصرفی خانگی با استفاده از تکنیک‌های داده کاوی

عاطفه حسن زاده*، فائزه اسدی و نجمه گربازکار

دانشگاه تهران، پردیس دانشکده‌های فنی، دانشکده فنی فومن، فومن، ایران.

ارسال ۲۰۲۰/۰۹/۰۱؛ بازنگری ۲۰۲۰/۱۱/۰۴؛ پذیرش ۲۰۲۱/۰۶/۰۵

چکیده:

برای بررسی اقتصادی قیمت مواد غذایی در ماه مه ۲۰۱۹ (اردیبهشت ۱۳۹۸) و به منظور تعیین روند افزایش یا کاهش قیمت‌ها نسبت به دوره‌های قبلی، قیمت اقلام غذایی را در آن زمان در نظر می‌گیریم. انواع اقلام مصرف شده در دوره‌های خاص در مناطق شهری و کل کشور برای تجزیه و تحلیل آماری ما انتخاب می‌شوند. در میان روش‌های مختلف مدل‌سازی و پیش بینی آماری و در یک رویکرد جدید، ما داده‌ها را با استفاده از تکنیک‌های داده-کاوی متشکل از روش‌های درخت تصمیم، قوانین انجمنی و قانون بیزی مدل‌سازی می‌کنیم. سپس پیش بینی، اعتبارسنجی و استانداردسازی صحت اعتبارسنجی بر روی آنها انجام می‌شود. نتایج اعتبارسنجی داده‌ها در منطقه شهری و روستایی و نتایج استانداردسازی صحت اعتبار سنجی در مناطق شهری و روستایی با دقت مطلوب ارائه شده است.

کلمات کلیدی: داده کاوی، قانون بیزی، درخت تصمیم، قانون انجمنی، اقلام مصرفی خانگی.