An Early Warning Method for Basic Commodities Price Based on Artificial Neural Networks

Jesús Silva[1], Noel Varela[2], Hugo Martínez Caraballo[3], Jesús García Guiliany[4], Luis Cabas Vásquez[5], Jorge Navarro Beltrán[6] and Nadia León Castro[7]

[1]Universidad Peruana de Ciencias Aplicadas, Lima, Perú.
jesussilvaUPC@gmail.com
[2] Universidad de la Costa, St. 58 #66, Barranquilla, Atlántico, Colombia
nvarela2@cuc.edu.co
[3,4] Universidad Simón Bolívar, Barranquilla, Colombia
{hugo.martinez, jesus.garcia}@unisimonbolivar.edu.co
[5,6,7] Corporación Universitaria Latinoamericana, Barranquilla, Colombia.
lcabas@ul.edu.co, jorgeelbacan05@gmail.com, cinpro@ul.edu.co

Abstract. The prices of products belonging to the basic family basket are an important component in the income of producers and consumer spending; its excessive variations constitute a source of uncertainty and risk that affects producers, since it prevents the realization of long-term investment plans, and can refuse lenders to grant them credit. His study to identify these variations, as well as to detect their sources, is then of great importance. The analysis of the variations of the prices of the basic products over time, include seasonal patterns, annual fluctuations, trends, cycles and volatility. Because of the advance in technology, applications have been developed based on Artificial Neural Networks (ANN) which have helped the development of massive sales forecast on consumer products, improving the accuracy of traditional forecasting systems. This research uses the RNA to develop an early warning system for facing the increase in basic agricultural products, considering seasonal factors.

Keywords: support vector machines, cyclic variation, predictive model, multilayer Perceptron, Multiple Input Multiple Output, forecast.

1. Introduction

In the markets of agricultural products, the quantities offered and demanded in each period (month or week) are disparate, which causes variations in prices. During the harvest periods, a large quantity of the product is offered in the markets, greater than the amount that is usually demanded for consumption. In these cases, the prices determined by the market, through supply and demand, are relatively low, lower than the average price of the year. Conversely, during periods when there is no harvest and therefore supply is low, less than the quantity normally demanded, market prices are high, reflecting the relative scarcity of the product [1], [2], [3].

To measure the price fluctuations in the different months of the year, it is very useful to build seasonality indexes [4]. If you take the average annual price of a product, it is obvious that the prices of some months would be higher than this average, while prices in other months will be lower. Percentage, the price average in relation to itself will be

equal to the unit; prices above the average, in relation to this average, will result in coefficients greater than 1; and prices below the annual average will result in coefficients less than 1 [5].

On the other hand in terms of predictions, there are probabilistic, deterministic, and hybrids, such as [4], [5]: Simple Moving Average, Weighted Moving Average, Exponential Smoothing, Regression Analysis, Box-Jenkins method (ARIMA), trend projections, etc., which have been used to generate forecasts, providing certain advantages and disadvantages compared to the others. However, these models are still unable to offer good results in an environment of high uncertainty and constant changes. To this end, new paradigms based on numeric modeling of nonlinear systems are necessary, such as the Artificial Neural Networks (ANN), and the Support Vector Regression (SVR) [7].

The present study proposes to Multi-Layer Perceptron with Multiple-Input Multiple-Output (MLP and MIMO) as a model for the prediction of prices of the products that belong to the family basket from Colombian State as the warning level criterion. Due to the nature of the products, the seasonal factor is integrated.

2. Theoretical Review

2.1 Artificial Neural Networks

Artificial Neural Networks (ANNs) can learn from data and can be used to construct reasonable input-output mapping, with no prior assumptions are made on the statistical model of the input data [6]. ANNs have nonlinear modeling capability with a data-driven approach so that the model is adaptively formed based on the features presented from the data [7].

An introduction to ANNs model specifications and implementation and their approximation properties has been provided from an econometric perspective [8]. Numerous studies have shown that ANNs can solve a variety of challenging computational problems, such as pattern classification, clustering or categorization, function approximation, prediction or forecasting, optimization (travelling salesman problem), retrieval by content, and control [9].

Some studies of ANN application related to financial early warning models have been conducted by [10] as well as [11] who used ANN as a classifier with a categorical output. Other authors used ANNs as financial forecasting models with continuous value. Some of them are [12] as well as [13], who implemented ANNs with a single-step prediction output. A previous study on ANNs forecasting model was also proposed by [14] for a multi-step prediction with a direct strategy, so the number of models is equal to the number of the prediction horizon. In the context of basic commodities price, the need for prediction is not limited to one-step forward but could be extended to include multi-step ahead predictions. Three strategies to tackle the multi-step forecasting problem can be considered, namely recursive, direct, and multiple output strategies [15]. The Multiple Input Multiple Output (MIMO) techniques train a single prediction model f that produces vector outputs of future prediction values [16].

The present study proposes to Multi-Layer Perceptron with Multiple Input and Multiple Output (MLP-MIMO) as to agricultural products price prediction model coupled with the coefficient of variation from the Colombian state price reference to the criteria of warning level.

2.2 Garson's algorithm to determine the level of importance.

Garson's algorithm was developed to determine the degree or level of importance of an entry indicator in an ANN. In many cases related to the measurement of the variables, the weights in the hidden layer and their interactions in the output network are considered. A measure proposed by Garson [17] consists of dividing the weights of the hidden layer into components associated with each input node and then assigning each of them a percentage of the total weights.

Several studies show the effectiveness of the Garson algorithm to evaluate the importance of an entry in the RNA [18], [19], [20]. The certainty of the algorithm of Garson was experimentally determined, concluding that the measure is applied successfully under a wide variety of conditions. As a result of this analysis, the Garson´s algorithm, on a scale from 0 to 1, determines a unique value for each explanatory variable that describes the relationship with the response variable in the model.

3. Material and methods

3.1 Data

In this study, the data were obtained from the National Administrative Department of Statistics of Colombia (DANE - National Administrative Department of Statistics), which provided a sales database of 1054 distributors of products of the basic basket from the main regions of Colombia in the time period from 2016 to 2018 [21]. The macroeconomic variables considered in this study range from food inflation, GDP, employment rate, minimum wage to commercial balance and capital flow of the nation [22]. Internal factors such as demand, and substitute and complementary products were also analyzed. Seasonal factors were incorporated to adjust the predictions obtained [23].

3.2 Methods

The early warning model consists of three main components, namely preprocessing, predictive model, and post-processing, as depicted in Figure 1 [24].

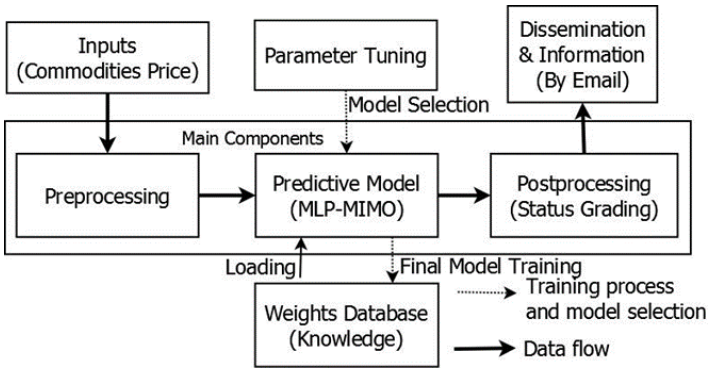

Fig. 1. Model of the early warning system [24].

Preprocessing: Before all raw data about commodity prices are presented to the predictive model, the preprocessing operations are applied on the data. The price surveys were conducted by local government at working days, so the commodity price data represent a daily basis with missing values in weekends and holidays. The data are therefore added to weekly data with the mean function to reduce the volume of the data for computational efficiency [24].

Predictive models: The predictive models are built from the trained final MLP-MIMO models obtained from the parameter tuning. Every single commodity in every city has its own model parameter structure. The weights after training are stored in a weight database so the model can be reloaded at any time [7].

Post-processing: The output of the predictive model is a normalized price prediction for eight weeks ahead. The post-processing is responsible for denormalizing the predicted price and determining early warning status based on the maximum predicted price according to Section 'Price Spike and Early Warning Status Leveling'. An alert will be sent to the stakeholders when the price is above a given threshold (on status 'watch' or 'monitoring') by an email service [10].

Finally, the Garson's algorithm for determining the level of importance was developed to determine the degree or level of importance of an input indicator in ANN. In many cases related to the measurement of the variables, the weights in the hidden layer and their interactions in the output network are considered. A measure proposed by [25] consists of dividing the weights of the hidden layer into components linked to each input node and then assigning each of them a percentage of the total weights.

3.3 Seasonality and unitary seasonal roots

When working with time series, one of the most important questions that the researcher should ask about is: what is the data generating process (DGP, in English, Data Generating Process) from which the studied sample comes? The conventional approach is to try to detect the different components of the DGP. Typically, 4 components are considered: the trend (stochastic or not), the cyclical part, the purely random component

and the seasonal component. Precisely, the seasonal component can be of a different nature: deterministic or stochastic. The most common ways of modeling seasonality involve: using dummy variables, seasonal autoregressive models (ARMA) (SARMA) or seasonal integration, and later modeling with a SARMA or ARMA model [10], [13] [20].

If the DGP implies that the stationary behavior of the series is purely deterministic, then it can be expressed as follows (Ec. 1) [26]:

$$y_t = \mu + \sum_{i=1}^{11} \alpha_i D_{i,t} + e_t \tag{1}$$

Where e_t is a Gaussian white noise error term, and Di, t is a dummy variable that takes the value of 1 if the observation corresponds to month i, and 0 otherwise.

Now, if the stationary behavior is stochastic, it is possible that it is stationary or not. In other words, the behavior can be such that in the event of disturbances in the series, the system tends to return to its seasonal but non-deterministic behavior (stationary stochastic seasonality) or that such disturbances, on the contrary, imply a permanent change in seasonal behavior (non-stationary stochastic seasonality) [10].

The case of stationary seasonal behavior can be represented with SARMA (p, q) × (P, Q) s models that take the following structure (Ec. 2) [27]:

$$\Phi(L^s)\phi(L)\, y_t = \Theta(L^s)\theta(L)e_t \tag{2}$$

where L represents the lag operator and Φ (·), φ (·), Θ (·) and θ (·) represent polynomials in the lag operator.

4. Results and Discussion

4.1 Product selection.

According to the United Nations Organization for Agriculture and Food (FAO) the basic products are divided into [2]:

- Food and non-alcoholic beverages
- Alcohol and tobacco
- Restaurants and hotels
- Dress and shoes
- Rental housing
- Housing services
- Furniture, home equipment
- Health
- Transport
- Communications

- Recreation and culture
- Personal care
- Educational services
- Financial services
- Others

Taking into account these categories, it is easy to identify each month how much the value of products and services increases and if inflation remains stable. For the purposes of the following investigation, the group of foods belonging to the perishable category will be assumed [10].

Perishable foods are those that are likely to spoil, break down or become unsafe to consume. They should be stored refrigerated at 40 degrees F (4.4 degrees C) or less to remain safe or prolong the time they will remain healthy because refrigeration slows bacterial growth. There are two completely different families of bacteria that can be found in food: "pathogenic bacteria", the class that causes food poisoning disease, and "spoilage bacteria", the class of bacteria that causes food spoilage and develops odors, unpleasant flavors and textures. Examples of foods that should be kept refrigerated for safety include meats, poultry, fish, dairy products, soft cheeses, cheesecake, most cakes, all cooked leftovers and any foods purchased refrigerated or labeled "keep refrigerated" ("keep refrigerated"). Very few fresh fruits and vegetables will remain safe at room temperature for a long time, so most should be stored in the refrigerator to prevent spoilage or mold growth. Some condiments that are safe at room temperature (such as ketchup, mustard, and soy sauce) can be kept chilled to preserve texture or flavor, but it is not necessary [7], [9], [20].

To select the products on which the forecasts were made, the f1-score criterion is used [24]. The ordering by this factor considered both the quantity and the value of sales for the selection of the most important products. Table 1 presents the values of the f1-score factor for each selected product.

Table 1. Prioritization and selection of products

Code	Product	Quantity	F1-SCORE
P1	Dairy products	815	145879.957
P2	Vegetables	1025	68264.5132
P3	Fruits	1254	67848.756
P4	Condiments	626	81443.7233
P5	Red meats	458	58743.3935
P6	White meats	1478	58489.3872
P7	Fish	325	32722.393

4.2 Model of the early warning system.

According to [16], the increase in the price is considered normal when it is below a certain threshold. The threshold is derived from the government reference price

established by the Ministry of Commerce and the variation coefficient noted (CVtarget). There are four degrees of warning status: normal, advisory, monitoring, and warning, whose criteria are presented in Table 2.

Table 2. Levels of warning status and their criteria.

Level	Status	Interval Price Increase Related to the Reference Price
I	Normal	$\leq$ 1.85CVtarget
II	Advisory	(1.85CVtarget, 2CVtarget]
III	Monitoring	(2CVtarget, 3.14CVtarget]
IV	Warning	> 3.14CVtarget

Table 3 shows the descriptive statistics for this market in each of the months. It can be deduced from this table that the distributions of the prices in all the months present an asymmetry towards the right, besides that the distribution in all the months is leptocurtic. Moreover, the descriptive statistics show an apparently different behavior in both the average and volatility in each of the months.

Table 3. Descriptive statistics for the price of the products under study

Month	Min	Max	Average	Variance	Standard error	Coefficient asymmetry	Kurtosis
January	111,35	443,78	210,2227	4983,6393	70,5949	1,6387	6,6491
February	112,6	577,95	225,2414	9455,707	97,2405	2,2091	8,9857
March	95,45	451,52	206,8382	5273,9794	72,6222	1,6627	6,9786
April	108,01	538,88	224,37	8820,5716	93,9179	1,7576	6,8356
May	93,34	444,99	203,8559	5609,5177	74,8967	1,4112	6,0556
June	112,78	474,36	230,9486	7859,2319	88,6523	1,0148	3,8402
July	111,48	434,4	209,1532	5387,7657	73,4014	1,2029	5,0726
August	105,58	532,06	230,8282	9129,2901	95,5473	1,4682	5,6772
September	113,92	356,33	202,5277	3189,4362	56,4751	0,6726	3,6953
October	105,07	567,6	230,7041	9843,1243	99,2125	1,8015	7,1485
November	114,3	489,76	213,6318	5959,7064	77,1991	2,094	8,6065
December	111,26	621,72	222,0477	11843,8742	108,8296	2,3321	9,3665

Given the above, it is necessary to determine the characteristics of the DGP of each of the series. For this purpose, the existence or not of seasonal unitary roots in the series must be identified. For this, as mentioned above, the HEGY seasonal unit root test of [12] for the monthly series.

Following [21], the need to consider dummy variables that capture the non-stochastic seasonality in the series is initially taken into account. Therefore, the test must be performed together with these dummy variables. The results of the test for the series of prices of the sample under study are presented in table 4.

Table 4. Test of Hylleberg, Engle, Granger and Yoo for the future of the sample

Null hypothesis	Test statistic
$\pi 1 = 0$ (non-seasonal unit root)	-0,866
$\pi 2 = 0$ (bi-monthly root)	-1,194
$\pi 3 = \pi 4 = 0$ (unit root for periods of 4 months)	692,315 ***
$\pi 5 = \pi 6 = 0$ (unit root quarterly)	421,117 ***
$\pi 7 = \pi 8 = 0$ (semi-annual unit root)	608,359 ***
$\pi 9 = \pi 10 = 0$ (unit root at the frequency $5\pi / 6$)	145,757 ***
$\pi 11 = \pi 12 = 0$ (annual unit root)	406,645 ***
$\pi 2 = \pi 3 = \cdots = 0$ (all the unit roots are present seasonal)	454,508 ***
$\pi 1 = \pi 3 = \cdots = 0$ (all unit roots are present, seasonal and non-seasonal)	419,209 ***

Note: rejects Ho with a level of significance of: 10% (*), 5% (**), 1% (***).

The seasonal dichotomous variables are not significant, so they are excluded to perform the unit root test.

To determine if the non-stochastic seasonality detected contributes to forecast the price of the sample, 3 ARIMA models will be used to generate price forecasts and compare them with the real values. In particular, after integrating the series to find stationary processes and perform a control for possible problems of heteroskedasticity and autocorrelation of the series, we proceed to model these filtered series (stationary series) in 2 different ways: with an ARMA model and a model SARMA. On the other hand, a SARIMA model is estimated for the unfiltered series; that is, without taking into account the non-stationary seasonality [15]. The best models for each case are reported in table 5.

Table 5. Estimated models

Filtered series	Best model SARIMA	Best ARIMA model
Crude sugar	SARIMA (9,0,2) (0,0,1) with mean 0	ARIMA (10,0,2) with mean 0
Refined sugar	SARIMA (2,0,2) (3,0,0) with mean 0	ARIMA (10,0,2) with mean 0
Unfiltered series	Best model SARIMA	
Crude sugar	SARIMA (4,1,2) (2,0,1) with mean 0	
Refined sugar	SARIMA (1,1,1) (2,0,0) with mean 0	

Finally, Table 6 shows the monthly seasonality indices of real prices. The period of increase in prices began between the months of April-May-June and ended in November. For the males of 1 and 11/4 of the year, prices reached increases of up to 5% and 3.5% in the month of August and decreased to -6.7 and -4.6 percentage points in February. For males aged 11/2 and 13/4, the high price season began in April and May with price increases of up to 3% and 2% in the month of June, respectively. In the low price season real prices fell to -4% in the month of February.

Table 6. Seasonality of real monthly prices

Months	P1	P2	P3	P4	P5	P6	P7
January	0,957	0,977	0,982	0,995	0,852	0,745	0,794
February	0,933	0,954	0,964	0,961	0,8584	0,847	0,876
March	0,963	0,966	0,967	0,971	0,741	0,725	0,723
April	0,986	0,986	0,995	1,002	0,8154	0,832	0,812
May	0,999	0,999	1,009	1,017	0,9584	0,921	0,941
June	1,041	1,035	1,030	1,019	1,1124	1,190	1,190
July	1,027	1,026	1,020	1,010	1,312	1,3412	1,306
August	1,050	1,035	1,020	1,010	1,214	1,257	1,268
September	1,034	1,023	1,019	1,013	1,154	1,178	1,137
October	1,022	1,014	1,013	1,008	1,175	1,199	1,188
November	1,004	1,010	1,002	1,004	1,185	1,124	1,163
December	0,987	0,989	0,981	0,991	0,851	0,812	0,840

The reference price of each product and the threshold for each warning status used in this study are presented in Table 7, according to [16] together with DANE [21].

Table 7. Reference price, interval price increase, and the levels of early warning status.

Commodity	Interval Percentage of Price Increase Relative to Reference Price			
	Normal	Advisory	Watch	Warning
P1	≤5%	(5%,10%]	(10%,15%]	>15%
P2	≤25%	(25%,50%]	(50%,75%]	>75%
P3	≤10%	(10%,20%]	(20%,30%]	>30%
P4	≤5%	(5%,10%]	(5%,10%]	>15%
P5	≤5%	(5%,10%]	(10%,15%]	>15%
P6	≤25%	(25%,50%]	(50%,75%]	>75%
P7	≤25%	(25%,50%]	(50%,75%]	>75%

According to the Garzon´s coefficient, the most relevant internal variables are the price and year, while in the macroeconomic policies are the foreign investment and the range of corruption, due to that both destabilize the price of the dollar.

5. Conclusions

Using monthly data corresponding to the sale of perishable products of the family basket during the 2016-2018 period, it is found that there is no deterministic seasonal pattern in the series. However, this study finds the existence of seasonal unitary roots.

In other words, a "summer" can turn into a "winter" due to unforeseen shocks. This result is used to generate forecasts outside the sample for the 12 months of each year. Said forecasts with models with filtered series that take into account the non-stationary stochastic seasonality behave better in terms of measures such as the Mean Absolute Error, the Mean Absolute Percentage Error and the Root Mean Square Error. That is, the finding of the existence of non-stationary stochastic seasonality allows us to improve the performance of forecast models.

Thus, the results imply that although there is seasonality, this is not deterministic. In this order of ideas, the proposed model presents an improvement over others available in the literature that do not take into account the "stochastic" seasonality due to the presence of seasonal roots.

References

1. Fonseca, Z., Heredia, A.P., Ocampo, P.R., Forero, Y., Sarmiento, O.L., Álvarez, M., Rodríguez, M.: Encuesta Nacional de la Situación Nutricional en Colombia 2010. Bogotá: Da Vinci (2011).
2. Instituto Colombiano de Bienestar Familiar (ICBF).: Ministerio de Salud y Protección Social, Instituto Nacional de Salud (INS), Departamento Administrativo para la Prosperidad Social, Universidad Nacional de Colombia. The National Survey of the Nutritional Situation of Colombia (ENSIN) (2015).
3. Food and Agriculture Organization of the United Nations (FAO).: Pan American Health Organization (PAHO), World Food Programme (WFP), United nations International Children's Emergency Fund (UNICEF). Panorama of Food and Nutritional Security in Latin America and the Caribbean. Inequality and Food Systems. Santiago (2018).
4. Frank, R. J., Davey, N., Hunt, S. P.: Time Series Prediction and Neural Networks. Journal of Intelligent and Robotic Systems **31** (3), 91-103 (2001).
5. Haykin, S.: Neural Networks and Learning Machines. New Jersey: Prentice Hall International (2009).
6. Jain, A. K., Mao, J., Mohiuddin, K. M.: Artificial neural networks: a tutorial. IEEE Computer **29** (3), 1-32 (1996).
7. Kulkarni, S., Haidar, I.: Forecasting Model for Crude Oil Price Using Artificial Neural Networks and Commodity Future Prices. International Journal of Computer Science and Information Security **2** (1), 81-89 (2008).
8. McNelis, P.D.: Neural Networks in Finance: Gaining Predictive Edge in the Market. Massachusetts: Elsevier Academic Press **59** (1), 1-22 (2005).
9. Mombeini, H., Yazdani-Chamzini, A.: Modelling Gold Price via Artificial Neural Network. Journal of Economics, Business and Management **3** (7), 699-703 (2015).
10. Sevim, C., Oztekin, A., Bali, O., Gumus, S., Guresen, E.: Developing an early warning system to predict currency crises. European Journal of Operational Research **237** (1), 1095-104 (2014).
11. Zhang, G.P.: Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing **50** (1), 159-75 (2003).
12. Horton, N.J., Kleinman, K.: Using R For Data Management, Statistical Analysis, and Graphics. CRC Press, Clermont (2010).
13. Chang, P.C., Wang, Y.W.: Fuzzy Delphi and backpropagation model for sales forecasting in PCB industry. Expert Syst. Appl. **30**(4), 715–726 (2006).

14. Lander, J.P.: R for Everyone: Advanced Analytics and Graphics. Addison-Wesley Professional, Boston (2014).
15. Chopra, S., Meindl, P.: Supply Chain Management: Strategy, Planning and Operation. Prentice Hall, NJ, (2001).
16. Acosta-Cervantes, M.C., Villarreal-Marroquín, M.G., Cabrera-Ríos, M.: Estudio de validación de un método para seleccionar técnicas de pronóstico de series de tiempo mediante redes neuronales artificiales. Ing Invest Tecnología **14** (1), 53–63 (2013).
17. Babu C.N., Reddy, B.E.: A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. Appl Soft Comput **23** (1), 27–38 (2014).
18. Cai, Q., Zhang, D., Wu, B., Leung, S.C.: A novel stock forecasting model based on fuzzy time series and genetic algorithm. Procedia Comput Sci **18** (1),1155–1162 (2013).
19. Egrioglu, E., Aladag, C.H., Yolcu, U.: Fuzzy time series forecasting with a novel hybrid approach combining fuzzy c-means and neural networks. Expert Syst Appl **40** (1), 854–857 (2013).
20. Kourentzes, N., Barrow, D.K., Crone, S.F.: Neural network ensemble operators for time series forecasting. Expert Syst Appl **41** (1), 4235–4244 (2014).
21. Departamento Administrativo Nacional de Estadística-DANE.: Manual Técnico del Censo General. Bogotá: DANE (2018).
22. Fajardo-Toro, C.H., Mula, J., Poler, R.: Adaptive and Hybrid Forecasting Models—A Review. In: Ortiz, Á., Romano, A.C., Poler, R., García-Sabater, J.P. (eds) Engineering Digital Transformation. Lecture Notes in Management and Industrial Engineering. Springer, Cham (2019).
23. Deliana, Y., Rum, I.A.: Understanding consumer loyalty using neural network. Polish Journal of Management Studies **16** (2), 51-61 (2017). Sekmen, F., Kurkcu, M.: An Early Warning System for Turkey: The Forecasting of Economic Crisis by Using the Artificial Neural Networks. Asian Economic and Financial Review **4**(1), 529-43 (2014).
24. Chang, O., Constante, P., Gordon, A., Singana, M.: A novel deep neural network that uses space-time features for tracking and recognizing a moving object. Journal of Artificial Intelligence and Soft Computing Research **7** (2), 125-136 (2017).
25. Scherer, M.: Waste flows management by their prediction in a production company. Journal of Applied Mathematics and Computational Mechanics, **16** (2), 135-144 (2017).
26. Sekmen, F., Kurkcu, M.: An Early Warning System for Turkey: The Forecasting of Economic Crisis by Using the Artificial Neural Networks. Asian Economic and Financial Review **4**(1), 529-43 (2014).
27. Ke, Y., Hagiwara, M.: An English neural network that learns texts, finds hidden knowledge, and answers questions. Journal of Artificial Intelligence and Soft Computing Research **7** (4), 229-242 (2017).