International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-7 Issue-5s, January 2019

Forecasting Prices of Fish and Vegetable using Web Scraped Price Micro Data

Mazliana Mustapa, Raja Rajeswari Ponnusamy, Ho Ming Kang

Abstract: In Malaysia, price statistics that are used as a proxy for inflation is the Consumer Price Index (CPI). The web scraped data has the possibility to become new source of compiling the CPI. The benefits using the web scraped data is can get the price information on a daily basis as compared to traditional data collection which takes on weekly or monthly basis. Price movement of the web scraped data can be monitored in real time and can benefits to policy makers. Forecasting price using the web scraped data helps the official statistics office to predict future value and can be used to control the situation of supply and demand side. Forecasting using web scraped data allow the policy makers to make the quick and right decision at the right time. Numerous studies have been conducted by the other National Statistics Office regarding the web scraped data, however studies on forecasting using web scraped is deficient. Thus, this study aims to utilize the web scraped data in forecasting ten selected fish and vegetables in Malaysia using Auto Regressive Integrated Moving Average (ARIMA) approach. The main objective of this study is to explore and evaluate the dependability of the alternative online data prices to forecast using ARIMA approach. The outcome of this research wills benefits to the Department of Statistics, Malaysia (DOSM). The forecasting model will be used to forecast price in the CPI compilation. This information offers better estimation and more timely. The modernization of the data collection by using the web scraped data will helps to reduce the burden of the establishments/supermarkets/wet markets. The coverage of CPI will be extended and will produce good quality statistics. The forecasting using web scraped data will improve understanding or perception of price behavior. Price forecasting will be an input to the policy makers when the price is increasing.

Keywords: Consumer Price Index, Web scraped data, Forecasting, ARIMA

I. INTRODUCTION

The Consumer Price Index (CPI) measures the percentage change purchasing cost by the residents of Malaysia in “fixed basket” of goods and services over a specific time period. There are 12 groups in CPI in order to calculate price index, namely Clothing and Footwear; Housing, Water, Electricity, Gas and Other Fuels; Health; Education; Communication; Food and Non-alcoholic Beverages; Restaurants and Hotels, Recreation Services and Culture; Alcoholic Beverages and Tobacco; Transport; Miscellaneous Goods and Services; and Furnishings, Household Equipment and Routine Household Maintenance (Department of Statistics, Malaysia, 2017). Official price statistics in Malaysia which are used to measure for inflation is CPI. The CPI is published monthly and it was released by the Department of Statistics, Malaysia (DOSM) after conducting monthly price data collection. The method of data collection is by visiting the supermarkets or outlets which a quite conventional. The monthly price data collection was done by the DOSM field officers, which they are collected the price information by visiting the supermarkets, wet markets or company of 14 states in Malaysia consists of urban and rural. The price data collection is based on selected products and services in the “fixed baskets”. The frequency of perishable products such as fish and vegetables is weekly. Meanwhile, the monthly price collection was done for the remaining products or services. E-commerce website is growing with the trends of consumers purchase online. The e-commerce is in line with tools and applications, which make consumer easy to purchase online. Thus, e-commerce website has a rich information about price. E-commerce web site has impacted on big data, which allows us to have a potential new data source. We can get more insights about the patterns of web scraped data. The online retailer will provide pricing information on their website on a daily basis. In this era, consumers like to purchase goods through the online channel. As explained by Cavallo and Rigobon (2016), online price offers many benefits such as real-time, the frequency of data is daily basis, economical, full of information, easy to collect anywhere and comparable in the country. Online inflation statistics are used in Argentina using web data begins when there is a discrepancy of inflation rate between National Statistic Office of Argentina. The inflation rate which is published by the National Statistic Office Argentina did not represent the actual changes of price. The online inflation rate is 20%, meanwhile the real inflation rate is 8%. Online inflation rate shows the best actual inflation rate and in line with the local economist estimations and several agencies. Thus, the important of online price inflation or index is needed and the index has released every day since March 2008. The author also mentioned the other elements features of online price data which is price level information. Web scrapping is the process of getting information from the website using special tools named web scraper. The web scraped data has the possibility to become new source of compiling the CPI. The benefits using the web scraped data is we can get the price information on a daily basis as compared to traditional data collection

Manuscript published on 30 January 2019.
***Correspondence Author(s)**
Mazliana Mustapa is working in School of Computing, Asia Pacific University of Technology and Innovation Malaysia .
Raja Rajeswari Ponnusamy is working in School of Mathematics, Actuaries and Quantitative Studies, Asia Pacific University of Technology and Innovation, Malaysia.
Ho Ming Kang is working in School of Mathematics, Actuaries and Quantitative Studies, Asia Pacific University of Technology and Innovation, Malaysia .

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

which takes on weekly or monthly basis. Price movement of the web scraped data can be monitored in real time and can benefits to policy makers. Forecasting price using the web scraped data helps the official statistics office to predict future value and can be used to control the situation of supply and demand side. Forecasting using web scraped data allow the policy makers to make the quick and right decision at the right time. Based on the ideas of Polidoro et al. (2015), web scraped data on selected items such as electronic products and airlines ticket are tested. The Italian National Statistical Institute (ISTAT) did the research with the main objective to have an alternative data source through online data collection. The outcome of this research shows good results where the process of data collection becomes efficient. Using the web scraped data helps decrease of the error of measurement and sampling error. The web scraping technique presents coverage error as neutral.Mayhew (2016) highlighted on the missing value of web scraped data by applying the imputation technique. Web scraped data has a few problems; scraper issue and the items is not available on the website. The author stated a simulation study is conducted in order to decrease the price imputation bias. There are three techniques were applied in the study, firstly computes the item or class mean, then using the previous price and imputation ratio. The results of the simulation study, the geometric average growth is recommended to impute price as the data is part of CPI. The author recommends the number of days for the imputation of the price. At least, three days is the number of days that can do price imputation before it is being eliminated from the dataset. The effects of applying the price imputation are the time series is smooth, regular and identical. Based on the results, carry forward previous price is the best method of price imputation with small bias.Metcalfe et al. (2016) found the Clustering Large dataset into Price Indices (CLIP), the statistical method. The CLIP method is to produce price index with web scraped data. The author mentioned that CLIP is a research index, which not suitable in analysing the economy. The assumption in CLIP model is that the consumer will buy a different product. Therefore, it only caters for the price change of cluster and group of services and goods. CLIP has a limitation where it depends on the product availability.

Powell et al. (2017) discussed on the average prices of web scraped price data of 33 food items of CPI products. The author describes that the analysis has two datasets which are 33 products of web scraped price of 14 months. The supermarket policy in United Kingdom stated that they have the same price across the country. Out of CPI weights, web scraped data contribute 13%. The second dataset is obtained from the CPI disaggregated data which are at the price level. The new model is develops to address issues of web scraped data. Hyper parameters estimates are produced to tackle the different product prices. The estimation helps to determine the products which can influence to CPI and determine the web scraped price data which has a correlation with the normal price collection survey.

Cavallo (2013) outlined the online price index, which is a mixture of official methods and web price data. The author claims that the online price index is equivalent to official price statistics (inflation) in Colombia, Venezuela, Argentina and Brazil.Chuanyang and Joseph (2016) stated Singapore Department of Statistics has a pilot work of collecting price information through web scraping. There are two types of web crawler, "point-and-click" and customized. Two issues were discussed along the pilot work which are no consistency of product type and lack of staff, which do not have the capability to handle the web crawler.Breton et al., (2016) highlights the work of scraping price from three websites in United Kingdom namely Waitrose, Tesco and Sainsbury. About 6,500 price quotation is collected on a daily basis of 35 CPI products within 13 months. The Gini, Eltetö, Köves and Szulc (GEKS) index is proposed to overcome the chain issue. The GEKS is better than the unit price index.

In Malaysia, official price statistics, namely CPI was released three weeks after the calendar month. For example, CPI December 2017 is published on January 24, 2018. CPI December 2017 increased by 3.5 percent as compared to the same month of the previous year. Food and Non-alcoholic beverages is the second main group shows an increase of 4.1%. A time lag of the publishing CPI has influenced decision to the policy makers or researchers on goods and services inflation. Currently, food price, especially fish and vegetables have become a hot issue after the government implemented subsidy rationalization. Numerous studies have been conducted by the other National Statistics Office regarding the web scraped data, however studies on forecasting using web scraped is deficient. Thus, this study aims to utilize the web scraped data in forecasting ten selected fish and vegetables in Malaysia using ARIMA approach.

The main aim of this study is to utilize the web scraped price data as an alternative data source and to forecast or predict the selected fish and vegetables using time series forecasting technique. The main objective of this study is to explore and evaluate the dependability of the alternative online data prices to forecast using ARIMA approach. There are three specific objectives in this study. The first objective is to determine the ARIMA model of the ten selected fish and vegetables. The second objective is to forecast prices of the selected fish and vegetables and finally the third objective is to identify the best ARIMA model based on performance measures.

II. METHODS AND MATERIALS

For the short term forecasting daily fish and vegetables, an ARIMA forecasting technique was implemented. This study applied for the ten selected fish and vegetables namely, bawal, kembong, selar kuning, cencaru, red bream, kangkung, round cabbage, long beans, green spinach and sawi jepun. The time series frame is built for selected of fish and vegetables, ten time series frame was built from 1 July to 30 November 2017 (5 months of daily price data). The imputation process of the missing value was based on three methods which are taking the mean, median and previous price.

The ARIMA is the forecasting technique that has high accuracy on forecast based on past value. ARIMA is applied to non-stationary time series with ARIMA (p, d, q). Michinaka et al., (2016) researched on price forecasting of Japanese logs using ARIMA and exponential smoothing method (ETS). The study is conducted on three Japanese logs which are karamatsu, hinoki and sugi. The aim of the research is to forecast monthly price of the selected logs of 6 and 12 months ahead. Based on the accuracy, ARIMA is outperformed as compare to ETS method. Paul, Hoque and Rahman, (2013) identified the best ARIMA model for daily share price. The Square Pharmaceuticals Limited (SPL) share price is used in this research. The author claims that the serial data is not stationary even though it has transformed to log price. Differencing the time series has overcome the stationary issue. The best ARIMA model is determine using the AME, RMSE, AIC, MAPE, SIC, AICc based on least value of the measures. The best model for forecasting share price is ARIMA (2, 1, 2). Abdullah (2012) described that ARIMA models have the capability to present many types of time series with the flexibility criteria. The researcher stated that the objective of the ARIMA model to predict the price of gold bullion coin (selling). The study undertaken is developed based on pre and post period of the forecast. The ARIMA (2, 1, 2) model is outperformed with the minimum error 10%. It can conclude that the model is good with the highest accuracy in predicting gold bullion coin.

R software is chosen as the tools in this study, considering a few factors. R Studio is the open source and free of data analysis. Three libraries or packages were used when performing the time series analysis namely ‘tseries’, ‘auto.arima’ and ‘forecast’. Two statistical test was performed using the Box.test() function for Ljung Box test and adf.test() function for Augmented Dickey Fulley test.The time series which is stationary must meet the three condition, mean is equal to zero, variance constant and covariance is stabilize. The Augmented Dickey Fuller test is to verify the time series is non-stationary or stationary. The null and alternative hypothesis is outlined as

Ho : The series is not stationary and

H1 : The series is stationary.

The P-value is obtained automatically using the software as 0.05 (95% confidence interval). If the p-value of ADF test is less than 0.05, Ho is rejected.

Model adequacy is checked by applying a statistical test which is Ljung-Box test. The hypothesis of Ljung-Box is outlined as

H0: autocorrelations of chosen lags = 0 and

H1: autocorrelations of chosen lags $\neq 0$.

The null hypothesis was rejected if the p-value is less than 0.05 (level of significance). Thus, one of the autocorrelation is not zero. If the p-value is greater than 0.05, we accept the null hypothesis with the assumption there is no autocorrelation of the chosen lags. The accuracy of the model is measured using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). This performance measure is used to calculate the forecast error (et). The best forecast error should be least value to identify the good fit model.

III. RESULTS

Table 1 shows the results of the Augmented Dickey-Fulley test. Based on the results, all ten selected fish and vegetables not stationary, referring to p-value which are greater than level of significance, 0.05.

Table 1: Augmented Dickey Fuller (ADF) Test

Item	Original Time Series			First Difference		
	Dickey Fuller	**Lag Order**	**p-value**	**Dickey Fuller**	**Lag Order**	**p-value**
Bawal	-2.9603	35	0.1762 **	-5.5360	5	0.01 *
Cencaru	-2.2159	35	0.4864 **	-7.6904	5	0.01 *
Kembong	-2.3993	5	0.41 **	-6.6710	5	0.01 *
Red Bream	-3.1999	37	0.09052 **	-6.4815	5	0.01 *
Selar Kuning	-2.758	35	0.2605 **	-5.8303	5	0.01 *
Green Spinach	-2.2854	15	0.4574 **	-5.2735	5	0.01 *
Kangkung	-3.0936	36	0.1206 **	-5.0658	5	0.01 *
Long Beans	-3.284	10	0.07652 **	6.2167	5	0.01 *
Round Cabbage	-2.8829	10	0.2084 **	-5.8983	5	0.01 *
Sawi Jepun	-2.5199	10	0.3597 **	-6.3604	5	0.01 *

Note: ** denotes not significant at 0.05 level of significance, indicates that time series is notstationary by accepting the Null Hypothesis

* the series is significant at 5% level, null hypothesis is rejected, Hence the series is stationary

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Table 2 depicted the best ARIMA model by using auto.arima function in R. From the results, it shows that Bawal, Cencaru, Kembong, Round Cabbage and Sawi Jepun has ARIMA (0, 1, 0) which is ARMA (0,0) if the difference is removed. This situation is called Random Walk or White Noise. Meanwhile, the other remaining items such as Red Bream, Selar Kuning, Green Spinach and Kangkung has ARIMA (2,0,1) with non-zero mean. Round Cabbage has ARIMA model with ARIMA (1,0,2) without differencing the series.

Table 2: ARIMA model

Item	ARIMA (p,d,q)
Bawal	ARIMA (0, 1 ,0)
Cencaru	ARIMA (0, 1 ,0)
Kembong	ARIMA (0, 1 ,0)
Red Bream	ARIMA (2, 0 ,1) with non-zero mean
Selar Kuning	ARIMA (2, 0 ,1) with non-zero mean
Green Spinach	ARIMA (2, 0 ,1) with non-zero mean
Kangkung	ARIMA (2, 0 ,1) with non-zero mean
Long Beans	ARIMA (1, 0 , 2) with non-zero mean
Round Cabbage	ARIMA (0, 1 ,0)
Sawi Jepun	ARIMA (0, 1 ,0)

The best ARIMA models are tabulated in Table 3. The model is developed using the best ARIMA fit model for the selected fish and vegetables. Bawal, Cencaru, Kembong, Round Cabbage and Sawi Jepun are using the same model as seen in Table 3 below.

Table 3: ARIMA best model

Item	Model
Red Bream	$Y_t = 2.4205 - 0.0333Y_{t-1} + 0.6382Y_{t-2} + 0.0594 + 0.9688e_{t-1}$
Selar Kuning	$Y_t = 1.4794 - 1.6648Y_{t-1} - 0.7431\ Y_{t-2} + e_t - 0.7924e_{t-1}$
Green Spinach	$Y_t = 1.1813 + 1.7246Y_{t-1} - 0.7774\ Y_{t-2} + 0.1008 - 0.8126e_{t-1}$
Kangkung	$Y_t = 1.0995 + 1.7293\ Y_{t-1} - 0.7776\ Y_{t-2} + 0.1023 - 0.8315\ e_{t-1}$
Long Beans	$Y_t = 0.5460 + 0.3667\ Y_{t-1} + 0.0780 + 0.4322\ e_{t-1} + 0.2638\ e_{t-2}$
Bawal, Cencaru, Kembong, Round Cabbage and Sawi Jepun	$\mathbf{Y_t = \mu + Y_{t-1}}$ whereμ : mean of the changes of period to period

Table 4 shows the performance measures of RMSE, MAE and MAPE of ten selected fish and vegetables. The ARIMA model of Selar Kuning has the least RMSE (0.04) compared to others. Meanwhile, the model of Selar Kuning (0.02) and Sawi Jepun (0.02) has the minimum MAE. The MAPE of Kembong and Red Bream has the lowest error (0.81) respectively.

Table 4: RMSE, MAE and MAPE of ARIMA Fit Model

Item	RMSE	MAE	MAPE
Bawal	0.32	0.09	1.12
Cencaru	0.14	0.04	1.39
Kembong	0.11	0.03	0.81
Red Bream	0.06	0.03	0.81
Selar Kuning	0.04	0.02	1.57
Green Spinach	0.10	0.05	4.67
Kangkung	0.10	0.05	5.07
Long Beans	0.08	0.05	10.57
Round Cabbage	0.18	0.05	3.43
Sawi Jepun	0.06	0.02	3.84

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Table 5: Forecasted Price of Fish (Malaysia Ringgit)

Date	Bawal	Cencaru	Kembong	Red Bream	Selar Kuning	Green Spinach	Kangkung	Long Beans	Round Cabbage	Sawi Jepun
1/12/2017	0.6	2.84	4.07	2.49	1.51	1.33	1.25	0.61	1.94	0.6
2/12/2017	0.6	2.84	4.07	2.46	1.5	1.28	1.2	0.58	1.94	0.6
3/12/2017	0.6	2.84	4.07	2.46	1.49	1.24	1.16	0.56	1.94	0.6
4/12/2017	0.6	2.84	4.07	2.45	1.49	1.2	1.12	0.55	1.94	0.6
5/12/2017	0.6	2.84	4.07	2.45	1.48	1.17	1.09	0.55	1.94	0.6
6/12/2017	0.6	2.84	4.07	2.44	1.48	1.15	1.07	0.55	1.94	0.6
7/12/2017	0.6	2.84	4.07	2.44	1.47	1.14	1.06	0.55	1.94	0.6
8/12/2017	0.6	2.84	4.07	2.43	1.47	1.13	1.05	0.55	1.94	0.6
9/12/2017	0.6	2.84	4.07	2.43	1.47	1.12	1.04	0.55	1.94	0.6
10/12/2017	0.6	2.84	4.07	2.43	1.47	1.12	1.04	0.55	1.94	0.6
11/12/2017	0.6	2.84	4.07	2.43	1.47	1.12	1.04	0.55	1.94	0.6
12/12/2017	0.6	2.84	4.07	2.42	1.47	1.13	1.05	0.55	1.94	0.6
13/12/2017	0.6	2.84	4.07	2.42	1.47	1.14	1.05	0.55	1.94	0.6
14/12/2017	0.6	2.84	4.07	2.42	1.48	1.14	1.06	0.55	1.94	0.6

The forecasted value of the ten selected fish and vegetables from 1 December 2017 to 14 December 2017 is given in Table 5. As seen from the table, Red Bream and Selar Kuning show variability of the prices with the average price per pieces. Meanwhile Bawal (RM0.60/pieces), Cencaru (RM 2.84/pieces) and Kembong (RM4.07/pieces) show the constant price for next 14 days. Vegetables such as Green Spinach, Kangkung and Long Beans showed varies price between 14 days. In the other hand, Round Cabbage (RM1.94/pieces) and Sawi Jepun (RM0.60/100g) showed the constant price for next 14 days.

IV. DISCUSSION

The ARIMA technique which applied for forecasting ten selected fish and vegetables is success by three different model; ARIMA (0, 1, 0) for Bawal, Cencaru, Kembong, Round Cabbage and Sawi Jepun; ARIMA (2, 0, 1) with non zero mean for Red Bream, Selar, Green Spinach and Kangkung; and ARIMA (1,0, 2) with non zero mean for Long Beans. The estimated ARIMA model is determined by the auto.arima function in R.

When comparing to other results, with other researcher, Abdullah (2012) has the best ARIMA model (2,1,2) of forecasting the bitcoin gold price. Meanwhile, Paul, Hoque and Rahman, (2013) also have the same ARIMA (2,1,2) model. The similarity between this two researches is both used the daily share price and bitcoin gold price which has more price movements as compared to the fish and vegetables price. The period of the time series also influenced the results of the estimation model. Since this study only covers 152 days (5 months). Long time series gives a better forecasting because ARIMA take account the number of periods in estimating the model.

Based on price behaviour of all ten selected items of fish and vegetables, the price quite stagnant for a long time (more than 14 days) and dropped if there are promotions. Overall, the price range of the selected items is within the minimum and maximum price range. Tesco is a big supermarkets, which can buy items in bulk and restock the items when the items is finished, it also helps the retailer to control the price which it still in the profit margin even though the discount is given.

CONCLUSION

As a conclusion, the web scraped data can used to forecast fish and vegetables using the ARIMA approach. The estimated ARIMA model can forecast for the next 14 days, as example. The forecast value of fish and vegetables can be used in data collection of CPI, in order to monitor the price behaviour. The government will have the signal when there is an increase of the price. Based on the forecast, appropriate solutions can be draw in order to control the inflation in the country.

REFERENCES

1. Abdullah, L. (2012). ARIMA Model for Gold Bullion Coin Selling Prices Forecasting. *International Journal of Advances in Applied Sciences*. 1 (4). 153-158.
2. Breton, R., Flower, T., Mayhew, M., Metcalfe, E., Milliken, N., Payne, C., Smith, T., Winton, J. and Woods,A. (2016). *Research indices using web-scraped data*: [Online]. 2016. Office for National Statistics, Newport. Available from: http://inflationmatters.com/wp-content/uploads/2015/10/ONS-web-scraped-data-article-01092015.pdf [Accessed: 9 December 2017]
3. Cavallo, A. (2013). Online and official price indexes: Measuring Argentina's inflation. *Journal of Monetary Economics*, 60(2), 152-165.
4. Cavallo, A. and Rigobon, R. (2016). The Billion Prices Project: Using Online Prices for Measurement and Research. *Journal of Economic Perspectives*, 30(2), 151-178.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

5. Chuanyang F. and Joseph L.W.H (2016). Experiences with the use of Online Prices in Consumer Price Index. [Online]. 2016. Singapore Department of Statistics. Available from: https://www.singstat.gov.sg/docs/default-source/default-document/library/publications /publications_and_papers/prices/ssnsep16-pg1-4.pdf [Accessed: 8 December 207].
6. Department of Statistics, Malaysia (2017). Consumer Price Index Malaysia December 2017 Available from: https://newss.statistics.gov.my/newss- portalx/ep /epFreeDownloadContentSearch.seam?cid=35096 [Accesed: 4 Jan 2018].
7. Dickey, D. and Fuller, W. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74(366), 427-431.
8. Fayyad, U., Piatetsky, G., Smyth, P. and Uthurusamy, R. (1996). *Advances in knowledge discovery and data mining*. 1st ed. Menlo Park, Calif.: AAAI Press/The MIT Press.
9. Kasyok, A. (2015). Simple Steps for Fitting Arima Model to Time Series Data for Forecasting Using R. *International Journal of Science and Research*. 4 (3). 318-321.
10. Mayhew, M. (2016). *Imputing Web Scraped Prices*. [Online]. 2016. Office of National Statistics. Available from:https://www.ons.gov.uk/economy/inflationandpriceindices/ methodologies/imputingwebscrapedprices. [Accessed: 7 November 2017].
11. Metcalfe, E., Flower, T., Lewis, T., Mayhew, M. and Rowland, E. (2016). *Research indices using web scraped price data: clustering large datasets into price indices (CLIP)*. [Online]. 2016. Office for National Statistics. Available from: https://www.ons.gov.uk/economy.
12. Michinaka, T., Kuboyama, H., Tamura, K., Oka, H. and Yamamoto, N. (2016). Forecasting Monthly Prices of Japanese Logs. *Forests*, 7(5), 94.
13. Paul, J., Hoque, M. and Rahman, M. (2013). Selection of Best ARIMA Model for Forecasting Average Daily Share Price Index of Pharmaceutical Companies in Bangladesh: A Case Study on Square Pharmaceutical Ltd. *Global Journal of Management and Business Research Finance*, 13(3), 14-25.
14. Polidoro, F., Giannini, R., Conte, R., Mosca, S. and Rossetti, F. (2015). Web scraping techniques to collect data on consumer electronics and airfares for Italian HICP compilation. Statistical *Journal of the IAOS*, 31(2), 165-176.
15. Powell, B., Nason, G., Elliott, D., Mayhew, M., Davies, J. and Winton, J. (2017). Tracking and modelling prices using web-scraped price microdata: towards automated daily consumer price index forecasting. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 1-20.

AUTHORS PROFILE

Mazliana Mustapa is working in School of Computing, Asia Pacific University of Technology and Innovation Malaysia .

Raja Rajeswari Ponnusamy is working in School of Mathematics, Actuaries and Quantitative Studies, Asia Pacific University of Technology and Innovation, Malaysia.

Ho Ming Kang is working in School of Mathematics, Actuaries and Quantitative Studies, Asia Pacific University of Technology and Innovation , Malaysia .

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)