ENHANCED KNOWLEDGE IN SCIENCES AND TECHNOLOGY
e-ISSN: 2773-6385

EKST

Vol. 4 No. 2 (2024) 233-240
https://publisher.uthm.edu.my/periodicals/index.php/ekst

Predictive Modelling of Unemployment Rate and Youth Unemployment Rate in Malaysia: A Forecasting Approach

Ling Shi Hong[1], Azme Khamis[1*]

[1] *Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, UTHM Kampus Cawangan Pagoh, Hab Pendidikan Tinggi Pagoh, KM 1, Jalan Panchor, 84600 Pagoh, Muar, Johor, MALAYSIA*

*Corresponding Author: azme@uthm.edu.my
DOI: https://doi.org/10.30880/ekst.2024.04.02.027

Article Info

Received: 27 December 2023
Accepted: 11 January 2024
Available online: 12 December 2024

Keywords

Unemployment Rate, Youth Unemployment Rate, SES, Holt's Linear Trend Method, ARIMA, Evaluation Metrics

Abstract

Unemployment is a global phenomenon that is regarded as a problem throughout the world. Unemployment occurs when a person who is unemployed and looking for work is unable to find job while youth unemployment refers to those aged 15 to 24, or 15 to 29. The study focused on forecasting Malaysia's overall unemployment rate and youth unemployment rate. The monthly data sets applied for both the unemployment rate and youth unemployment rate are starting from January 2010 to March 2023 (159 data points) and January 2016 to March 2023 (87 data points) respectively. This study employs the performance of three forecasting models: Simple Exponential Smoothing (SES), Holt's Linear Trend Method, and Autoregressive Integrated Moving Average (ARIMA). The objective is to construct forecast models for unemployment rate and youth unemployment rate using the three methods, determine the best performing model among them using the performance metrics such as mean absolute erro (MAE), mean square error (MSE) and mean absolute percentage error (MAPE). Lastly, the third objective is to forecast the future value of unemployment rate and youth unemployment rate by implementing the best forecast model. By employing accuracy measures, the study identify that Holt's Linear Trend method emerges as the optimal model for forecasting the monthly unemployment rate with the minimum MAE and MAPE value of 0.005648 and 1.2354 separately. On the other hand, SES is identified as the best method for forecasting monthly youth unemployment rate with MAE, MSE and MAPE value of 0.05220, 0.1850 and 2.4361 respectively. In conclusion, this study underscores the significance of selecting reliable forecasting models to enhance economic planning, workforce strategies, and overall economic stability.

1. Introduction

Winkelmann [1] stated that unemployment occurs when a person wishes to work but fails to find work. The unemployment rate is sometimes expressed as a percentage of the entire labor force, which is a significant indication of a country's labour market and economy. Unemployment has far-reaching social, economic, and psychological upshot for individuals, families, communities, and the economy as a whole, and it is a major source of concern for policymakers, employers, and job seekers alike. Malaysia has recorded an unemployment rate of 3.6% in December 2022 and this value remained constant up to January 2023 based on the information from DOSM.

When young people, often between the ages of 15 and 24, actively seek employment but are unable to secure acceptable positions, this is referred to as youth unemployment. High youth unemployment rates can have a long-term negative impact on people, societies, and economies, which is a major worry in many nations. Detrimental macroeconomic performance is one of the key factors that contributes to this issue. According to [2] youth in particular are impacted by the economy's lack of significant economic growth. Inadequate economic growth performance also contributes to rising rates of youth unemployment, particularly during times of crisis. A sharp reduction in economic performance may be responsible for half of the increase in youth unemployment during the crisis [3]. Several factors can contribute to this dilemma. Firstly, the economic factor and lacking work experience can have impact on the unemployment rate [4]. This situation occurs because there is reduction and limitation in job opportunity and thus increase in unemployment rate when the economy of a country does not perform well as expected.

Predicting future unemployment rates in Malaysia is crucial especially in the developing country as the unemployment rate keep fluctuating from time to time. Hence, many past research and investigation had been conducted to forecast the unemployment rate in Malaysia. For instances, research by Nor et al. reveals that that Holt's model is the best at projecting the overall yearly unemployment rate, male yearly unemployment rate, and overall quarterly unemployment rate. Furthermore, both the female yearly unemployment rate and the overall monthly unemployment rate, SES was the greatest predictor [5]. Furthermore, analysis by Ismail et al. tells that the best time series models found were ARIMA (2, 1, 2) and ARFIMA (0, −0.2339, 0) [6]. Research done by Ramli et al. to forecast the unemployment rate in Malaysia exposed that ARIMA (2,1,2) is the best model compared to Holt's exponential smoothing method [7].

The unemployment rate, a major economic statistic, is critical in predicting job success and driving policymaking initiatives [8]. Accurate unemployment rate forecasting enables policymakers to design effective solutions, such as job creation program and skill development efforts, to address and mitigate the negative effects of unemployment on individuals and the larger economy. Furthermore, this research helps governments and businesses anticipate labor market changes and modify fiscal, monetary, and investment policies accordingly. Accurate forecasting provides critical for proactive decision-making and ensuring the resilience of national and regional economies.

The purpose of this study is to apply SES, Holt's Linear Trend Method and ARIMA modelling to forecast the future unemployment rate and youth unemployment rate. The best model for each unemployment rate and youth unemployment rate will be evaluated through the accuracy measure like Mean Absolute Error (MAE), Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE).

2. Methodology

The study makes use of historical data on Malaysia's unemployment rate and young unemployment rate. Monthly unemployment rate statistics in percentage are available from January 2010 to March 2023 (159 data points), whereas youth unemployment rate data are available from January 2016 to March 2023 (87 data points). These datasets, collected from the website of the Malaysian Department of Statistics (DOSM), are divided into training and testing sets. The unemployment rate training set includes January 2010 to December 2021, whereas the youth unemployment rate spans January 2016 to September 2021. Unemployment rate dataset testing period runs from January 2022 to March 2023 while youth unemployment rate testing period runs from October 2021 to March 2023. For data analysis, statistical software such as Minitab 21, Excel, and SPSS are used. Three main methods have been applied in this study which are SES, Holt's Linear Trend method and ARIMA modelling.

2.1 Simple Exponential Smoothing (SES)

SES is a time series data forecasting method that uses past values to predict future ones and used for time series data without trend and seasonality. It computes a weighted average, with greater weight given to current observations than to older ones. The key concept is to emphasise fresh data while gradually decreasing the importance of older data. The smoothing parameter alpha (0 < α <1) controls how quickly the weights decline, influencing the significance of historical observations. A higher alpha weight the most recent observation more heavily, whereas a lower alpha weight prior observation more evenly. SES can be used to predict time series data with a consistent or slowly changing pattern [6]. SES model can be express as shown in equation (1):

$$F_t = \alpha Y_t + (1-\alpha) F_{t-1} \tag{1}$$

where

F_t : Forecast value for period t

α : Smoothing constant $(0 < \alpha < 1)$

Y_t : Actual value for period t

F_{t-1} : Forecast value for period $t-1$

2.2 Holt's Linear Trend Method

Double Exponential Smoothing, often known as Holt's approach, is a forecasting methodology for time series data that adds a trend aspect to simple exponential smoothing. Holt's technique includes two components: level and trend, which is especially beneficial when there is a clear pattern in the data over time. The level indicates the time series' average value, whereas the trend represents the rate of change. Holt's approach estimates these components and projects future values using smoothing parameters, alpha and beta [9]. A forecast equation (2) and two smoothing equations (3) and (4) are involved in this model [10].

$$\hat{Y}_{t+h|t} = \ell_t + hb_t \tag{2}$$

$$\ell_t = \alpha Y_t + (1-\alpha)(\alpha_{t-1} + b_{t-1}) \tag{3}$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \tag{4}$$

Where;

$\hat{Y}_{t+h|t}$: Forecast value for period $t+h$

ℓ_t : Estimate of level of the series at time t

h : number of periods ahead to be forecast

b_t : Estimate of trend of the series at time t

α : Smoothing parameter for the level

Y_t : Actual value at time t

β^* : Smoothing parameter for the trend

2.3 ARIMA Model

ARIMA refers to the autoregression process (AR), integration (I), and moving average process (MA). ARIMA is a method used in the Box-Jenkins methodology. Box-Jenkins is a method for discovering and estimating models that encompasses both AR and MA models. Box-Jenkins requires a stable variance and a stationary mean in the data. To meet these conditions prior to the model identification procedure, Box-Cox transformation and differencing techniques are applied. When the data does not have a constant variance, the Box-Cox transformation is employed, followed by a back-transformation to yield the anticipated values. The differencing technique is used to determine whether non-stationary data can be transformed into a stationary mean using the differencing procedure.

The ARIMA model is written as ARIMA (*p*,*d*,*q*), where *p* represents the order of the AR component, *d* represents the degree of initial differencing involved, and *q* represents the order of the MA part [11]. Time series data must be stationary in order to proceed with time series. To begin, a time series graph is drawn to discover the graph's components or pattern. Before beginning with the identification model, if the dataset exhibits non-stationary behaviour, it must be turned into a stationary series. To determine if the mean and variance are stationary, the autocorrelation function (ACF), partial autocorrelation function (PACF), and Box-Cox transformations were used. A full model of non-seasonal ARIMA model can be expressed in equation (5).

$$\phi_p(B)\nabla^d y_t = \theta_q(B)e_t \tag{5}$$

Where:

$\nabla^d y_t$ is the degree of differencing, can express as below:

$$\nabla^d = (1-B)^d \tag{6}$$

$\phi_p(B)$ is the non-seasonal autoregressive operator of order p, as below:

$$\phi_p(B) = 1 - \phi_1(B) - \phi_2 B^2 - \ldots \phi_P B^P \tag{7}$$

$\theta_q(B)$ is the non-seasonal moving average operator of order q, as below:

$$\phi_q(B) = 1 - \theta_1(B) - \theta_2 B^2 - \ldots \theta_q B^q \tag{8}$$

By substitute equations (6), (7) and (8) into equation (5), the ARIMA model can derived as (9):

$$(1-\phi_1 B-\phi_1 B^2-...\phi_p B^p)(1-B)^d y_t=(1-\phi_1 B-\phi_2 B^2-...\phi_q B^q)e_t \tag{9}$$

where:

y_t : differenced series.

ϕ : parameter of differencing AR model.

θ : parameter of differencing MA model.

p : order of the autoregressive part.

d : degree of first differencing involved.

q : order of the moving average part.

e_t : white noise

If the data does not exhibit seasonality, regular differencing can be used to analyse it. Seasonal differencing can be applied to data that exhibits seasonality behaviour. After the mean has reached stationary, the pattern of the autocorrelations function (ACF) and partial autocorrelation function (PACF) can be used to identify the tentative models. The model identification criteria are shown in Table 1.

Table 1: Model Identification

Model	ACF	PACF
AR(*p*)	Decay to 0 with exponential pattern	Cut off at lag *p*
MA(*q*)	Cut off at lag *q*	Decay to 0 with exponential pattern
AR(*p*) or MA(*q*)	Cut off at lag *q*	Cut off at lag *p*
ARMA(*p*,*q*)	Decay to 0 with exponential pattern	Decay to 0 with exponential pattern

Following that, the parameter values of the selected models must be approximated using the *p*-value as a reference. The *p*-value of the estimated parameter value must be less than 5%, indicating that the estimated parameter value is significant. The following phase is model checking. The Ljung-Box test is used to determine the lack of fit of a time series model. The test was applied to time series residuals after fitting an ARIMA model to the data.

2.4 Accuracy Measure

According to [12], prediction accuracy is a critical metric for assessing model forecasting ability. [13] defined accuracy measure as a method that provides information about the performance of forecasting methods in order to determine whether or not the model is suited to the data. The forecasting ability of the ARIMA model, SES method, and Holt's Linear Trend approach will be assessed in this study utilising accuracy measures. Mean Absolute Error (MAE), Mean Square Error (MSE), and Mean Absolute Percentage Error (MAPE) are three forms of statistics errors used to assess the accuracy of forecasting methodologies. The lower the value of error, the more accurate the forecasting systems [14]. Formulae for the MAE, MSE and MAPE are shown in equations (10), (11) and (12). Table 2 shows the interpretation of MAPE value. The accuracy of forecast value is higher when the MAPE value is lower.

$$MAE=\frac{\sum_{t=1}^{n}\left|y_t-y_t\right|}{n} \tag{10}$$

$$MSE=\frac{\sum_{t=1}^{n}(y_t-y_t)^2}{n} \tag{11}$$

$$\mathrm{MAPE}=\frac{\sum_{t=1}^{n}(\frac{y_t-y_t}{y_t})}{n}\times 100 \tag{12}$$

where:

y_t : Actual observation at period t

$\hat{y}_t$: Forecast value at period t

n : Number of period used in forecast

Table 2: Interpretation of MAPE value

MAPE Value	Interpretation
Less than 10%	Highly Accurate Forecast
11% to 20%	Good Forecast
21% to 50%	Reasonable Forecast
More than 51%	Inaccurate Forecast

3. Result And Discussion

The anticipated value of models is obtained first. The calculated forecasted value is used to compute the model accuracy measures. To assess the performance of the models, the accuracy measures for unemployment rate and youth unemployment rate are provided in Table 3 and Table 4, respectively. The lower the error, the more accurate the anticipated results.

Table 3: Accuracy measures for Unemployment Rate (UR)

Accuracy Measure	Simple Exponential Smoothing	Holt's Linear Trend Method	ARIMA Model
MAE	0.05099	0.005648	0.497776
MSE	0.006187	0.04662	0.305969
MAPE	1.3492	1.2354	13.64809

Table 4: Accuracy measures for Youth Unemployment Rate (YUR)

Accuracy Measure	Simple Exponential Smoothing	Holt's Linear Trend Method	ARIMA Model
MAE	0.05220	0.05899	1.086164
MSE	0.1850	0.1999	1.468571
MAPE	2.4361	2.6359	14.94677

According to the results obtained in Table 3, Holt's Linear Trend method has lower values of MAE and MAPE when compared SES and ARIMA modelling method for Unemployment Rate (UR). This indicates that the Holt's Linear Trend method has higher tendency to be more accurate in forecasting. Hence, this makes the Holt's Linear Trend method to have the better forecasting performance compared to SES and ARIMA modelling method.

For Youth Unemployment Rate (YUR), SES has lowest values of all MAE, MSE and MAPE when compared to Holt's Linear trend method and ARIMA modelling method according to Table 4. This indicates that the SES method has higher tendency to be more accurate in forecasting. Therefore, this causes the SES model to have the better forecasting performance compared to ARIMA modelling method and Holt's Linear trend method.

According to Table 5, the predicted value using ARIMA (1,1,1) is significantly different from the actual value. This shown that the accuracy of ARIMA (1,1,1) is relatively poor, yielding fewer accurate findings. The anticipated value generated by SES and Holt's Linear Trend, on the other hand, is quite close to the actual value. This suggests that both models outperform ARIMA (1,1,1) since their errors are less. Holt's Linear Trend method is chosen as its error is the lowest among other method. This can be proven through Fig. 1 which the forecasted data for Holt's method is the closest to the actual data.

Table 6 shows that the anticipated value using ARIMA (1,1,1) differs significantly from the actual value. This demonstrates that the accuracy of ARIMA (1,1,1) is quite low, resulting in fewer accurate discoveries. In contrast, the expected value created by SES and Holt's Linear Trend is very close to the actual value. Because their errors are smaller, both models outperform ARIMA (1,1,1). The SES approach was chosen since it has the lowest error among the other methods. This can be proven through Fig. 2 which the forecasted data for SES method is the closest to the actual data.

Table 5: Comparison of the Forecasting Performance for Unemployment Rate (UR)

Date	Actual Value	SES	Holt's Linear Trend	ARIMA (1,1,1)
1/1/2022	4.2	4.2076	4.1983	4.2135
2/1/2022	4.1	4.2006	4.1957	4.2249
3/1/2022	4.1	4.1076	4.0983	4.2346
4/1/2022	3.9	4.1006	4.0957	4.2427
5/1/2022	3.9	3.9151	3.9009	4.2495
6/1/2022	3.8	3.9011	3.8958	4.2553
7/1/2022	3.7	3.8076	3.7983	4.2602
8/1/2022	3.7	3.7081	3.6984	4.2643
9/1/2022	3.6	3.7006	3.6957	4.2677
10/1/2022	3.6	3.6076	3.5983	4.2706
11/1/2022	3.6	3.6006	3.5957	4.2731
12/1/2022	3.6	3.6000	3.5957	4.2752
1/1/2023	3.6	3.6000	3.5957	4.2769
2/1/2023	3.5	3.6000	3.5957	4.2784
3/1/2023	3.5	3.5076	3.4983	4.2796

Table 6: Comparison of the Forecasting Performance for Youth Unemployment Rate (YUR)

Date	Actual Value	SES	Holt's Linear Trend	ARIMA (1,1,1)
10/1/2021	8.1	8.5334	8.5844	8.5395
11/1/2021	8.2	8.1524	8.2336	8.5684
12/1/2021	8.5	8.1942	8.2172	8.5896
1/1/2022	8.3	8.4630	8.4360	8.6051
2/1/2022	8.1	8.3197	8.3439	8.6164
3/1/2022	8.0	8.1266	8.1713	8.6247
4/1/2022	7.7	8.0153	8.0523	8.6307
5/1/2022	7.4	7.7381	7.7985	8.6352
6/1/2022	7.4	7.4409	7.5098	8.6384
7/1/2022	7.4	7.4049	7.4351	8.6408
8/1/2022	7.1	7.4006	7.4157	8.6426
9/1/2022	7.1	7.1363	7.1877	8.6438
10/1/2022	7.4	7.1044	7.1287	8.6448
11/1/2022	7.0	7.3643	7.3364	8.6454
12/1/2022	6.9	7.0440	7.0928	8.6459
1/1/2023	7.0	6.9174	6.9554	8.6463
2/1/2023	7.0	6.9900	6.9942	8.6466
3/1/2023	7.1	6.9988	7.0042	8.6468

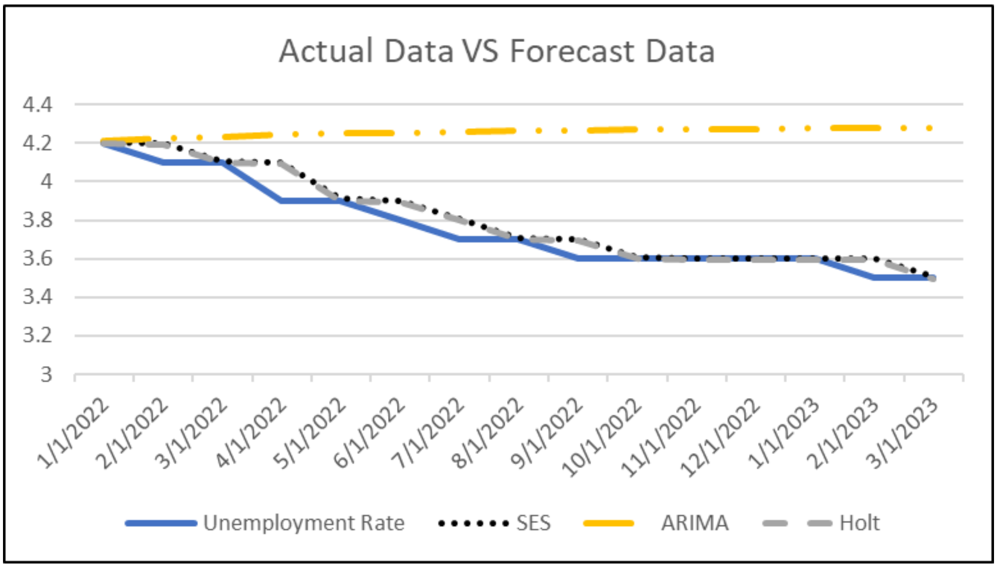

Fig. 1 Actual Data VS Forecast Data for Unemployment Rate (UR)

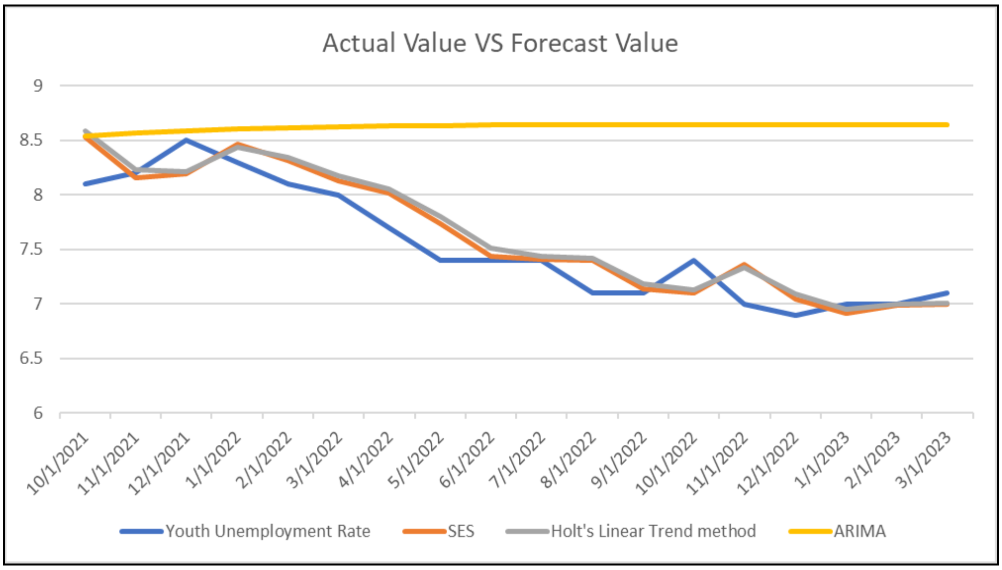

Fig. 2 Actual Data VS Forecast Data for Youth Unemployment Rate (YUR)

4. Conclusion

This study provides a detailed exploration of unemployment rate and youth unemployment rate forecasting using SES, ARIMA, Holt's Linear Trend methods, achieving its three primary objectives. The first objective is to build forecasting model on monthly unemployment and youth unemployment rate in Malaysia by implementing the three methods. The forecasting models showed different trends. SES method and Holt's Linear Trend method stay close to actual data while ARIMA model shows an increasing trend for both the data sets. The second objective involves performance evaluation utilizing metrics such as MAPE, MAE and MSE confirmed the superior accuracy of Holt's Linear Trend method compared to SES and ARIMA for unemployment rate (UR) while SES method for youth unemployment rate (YUR) compared to Holt's Linear Trend method and ARIMA. These findings establish Holt's Linear Trend method as the most effective method for forecasting unemployment rate while SES as the most effective method for forecasting youth unemployment rate, providing crucial insights for investors and policy makers in making the informed decisions. ARIMA method is not chosen because the value of accuracy measure is higher compared to the other two methods. The third objective to forecast the future unemployment and youth unemployment rate by using the best forecasting model. Based on the final output, the forecasted unemployment rate and youth unemployment rate for April 2023 were 3.5% and 7.1% respectively.

The research focuses on forecasting Malaysia's unemployment rate using both the SES approach, Holt's Linear Trend approach and ARIMA modelling. The authors do, however, note the limitations of these models, particularly in the face of unforeseen events such as the COVID-19 pandemic. Despite the data showing a consistent trajectory till 2020, the outbreak prompted a different forecasting technique. To address the impact of such unexpected events, the authors urge that future study incorporate an intervention time series model. They recommend looking into additional time series models and categorising data into before and after 2020 periods for more accurate modelling and forecasting. This comparative analysis seeks to identify the best forecasting approach for various time periods. It had been proved that Holt's Linear Trend method produce the best forecast for Unemployment Rate (UR) and SES method produce the best forecast for Youth Unemployment Rate (YUR) because of its lower value of accuracy measures.

Acknowledgement

The authors would thank the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia for its support.

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Ling Shi Hong, Azme Khamis;* ***data collection:*** *Ling Shi Hong;* ***analysis and interpretation of results:*** *Ling Shi Hong, Azme Khamis;* ***draft manuscript preparation:*** *Ling Shi Hong. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Winkelmann, R. (January 2014). *Unemployment*. Springer eBooks. https://doi.org/10.1007/978-94-007-0753-5_3078

[2] Ryan, P. (2001). The School-to-Work Transition: A Cross-National Perspective. *Journal of Economic Literature*, 39(1): 34-92.

[3] Banerji, Saksonovs, Lin, & Blavy. (2014, December). Youth Unemployment in Advanced Economies in Europe: Searching for Solutions. https://www.imf.org/external/pubs/ft/sdn/2014/sdn1411.pdf

[4] Hudson, P. (2013). How Unemployment Rates Affect The Economy. Retrieved on 21, 2013 from http://elitedaily.com/news/business/how-unemployment-rates- affect-the-economy/
Nor, M. E., Saharan, S., Lin, L. S., Salleh, R. M., & Asrah, N. M. (2018, November 30). *Forecasting of Unemployment Rate in Malaysia Using Exponential Smoothing Methods*. International Journal of Engineering & Technology. https://doi.org/10.14419/ijet.v7i4.30.22365

[5] ISMAIL, N., RAMZI, N., & Mah, P. (2022). Forecasting The Unemployment Rate In Malaysia During Covid-19 Pandemic Using Arima And Arfima Models. *Malaysian Journal Of Computing, 7*(1), 982-994. doi:10.24191/mjoc.v7i1.14641Jo, C., Kim, D. H., & Lee, J. W. (2023). *Forecasting unemployment and employment: A system dynamics approach*. Technological Forecasting and Social Change. https://doi.org/10.1016/j.techfore.2023.122715

[6] Ramli, S. F., Fidaus, M., Uzair, H., Khairi, M. & Zharif, A (2018). Prediction of the Unemployment Rate in Malaysia. International Journal of Modern Trends in Social Sciences, 1(4), 38-44.
Ravinder, H.V. (2013). Forecasting With Exponential Smoothing – What's The Right Smoothing Constant? https://core.ac.uk/download/pdf/268106307.pdf

[7] Holt, C.C. (1957). Forecasting trends and seasonals by exponentially weighted averages, *O.N.R. Memorandum* 52/1957, Carnegie Institutte of Technology.

[8] Wilson, J.H. and Keating, B. (2007). *Business Forecasting with Accompanying Excel-Based ForecastX Software*. 5th Ed. Boston, Mass: McGraw-Hill

[9] G.P. Zhang, "Time series forecasting using hybrid ARIMA and neural network model," Neurocomputing, vol. 50, pp. 159-175, 2003.

[10] M. Askari, & H. Askari, "Time series grey system prediction-based models: Gold price forecasting," Trends in Applied Sciences Research, vol. 6, no. 11, pp. 1287-1292, 2011.

[11] S. Makridakis, "Accuracy measures: theoretical and practical concerns," International Journal of Forecasting, vol. 9, no. 4, pp. 527-529, 1993.
[12] M.H. Lee, M.E. Nor, and H.J. Sadaei, "Fuzzy Time Series: An Application to Tourism Demand Forecasting," vol. 9, no. 1, pp. 132-140, 2011.