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

EKST

Vol. 4 No. 1 (2024) 162-170
https://publisher.uthm.edu.my/periodicals/index.php/ekst

Prediction of Price Inflation for Education in Malaysia

Ashwaq Mohd Akhir[1], Kamil Khalid[1]*, Mohd Saifullah Rusiman[1]

[1] *Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh Edu Hub, 84600 Pagoh Muar, Johor, MALAYSIA*

*Corresponding Author: kamil@uthm.edu.my
DOI: https://doi.org/10.30880/ekst.2024.04.01.019

Article Info

Received: 27 December 2023
Accepted: 3 June 2024
Available online: 27 July 2024

Keywords

Forecast, ARIMA, Exponential smoothing, Naïve bayes

Abstract

This study addresses the critical issue of price inflation in the education sector in Malaysia, aiming to develop a robust forecasting model to predict future price trends as the cost of education continues to rise, understanding and predicting inflationary patterns becomes essential for policymakers, educational institutions, and stakeholders where the data set was taken from DOSM website. The data set contains monthly information about the consumer price index in Malaysia which from January 2015 to March 2023. This study aims to predict the price inflation for education in Malaysia for next four months to address the insufficiencies in capturing the true impact of education expenses on the Core CPI. Hence, the model has been chosen based on the percentage of accuracy value for the methods.

1. Introduction

Education is the acquisition of knowledge, skills, values, and attitudes through various methods, such as teaching, training, and research. It is a lifelong process that focuses on cognitive, emotional, social, and physical growth. Education can take place in formal settings or informally through independent study and hands-on learning [1]. It encompasses various courses and disciplines, such as math, science, language, history, arts, and physical education. The education system needs to be competitive in making changes due to the abundance of information sources. It emphasizes critical thinking, problem-solving, creativity, and effective communication. College tuition prices increased significantly between 1980 and 2004, making it crucial to understand the causes of tuition inflation [2]. Higher education in Malaysia has been influenced by international issues such as globalization and trade. The formation of the Ministry of Higher Education Malaysia (MOHE) in 2004 has led to a shift in the priority placed on the expansion of higher education.

Education is a crucial factor in social development, with affordability playing a significant role in ensuring equal access to quality education. It is seen as an investment with a long-term horizon, leading to increased output and economic success [3]. Education is also a key factor in determining economic growth, as emphasized by classical and neoclassical economists [4]. Time Series Forecasting (TSF) is a method for predicting future values, particularly useful for understanding and predicting price changes for education items in Malaysia.

This data relates to the growth in the cost of obtaining an education in Malaysia. It is a method for determining the state of the economy by analysing the extent to which a preset assortment of products and services has seen an increase or drop in price over the course of time. The core Consumer Price Index is used to extract underlying patterns in inflation by eliminating the costs of food and energy, which tend to be more volatile [5]. The index is constructed by utilising the average price change for items connected to education that are acquired from price collection centres located in each of Malaysia's states. These statistics are available on the website maintained by the Department of Statistics in Malaysia (DOSM). It includes monthly data that spans from January 2015 all the way through March 2023.

According to Mwanga, Y education is the process of acquiring knowledge, skills, values, and attitudes via a number of methods, such as teaching, training, or research. Education can also refer to the institution that facilitates this process. It is a process that one goes through their entire life, beginning when they are young and continuing till the end of their days [6]. As both society and schools become more institutionalized, the educational process will become less directly tied to day-to-day living, less about demonstrating and learning in the context of the real world, and more abstracted from practice, with a greater emphasis on distilling, telling, and learning things out of context. This will occur because education will focus less on showing and learning in the context of the real world.

The Core Consumer Price Index (CPI) in Malaysia is undervalued due to the insufficient inclusion of education expenses. This problem statement aims to predict price inflation for education in the next four months, enabling policymakers to make informed decisions and implement measures to mitigate the effects of inflation on education costs. Other previous studies were done in using forecasting method with the statistics technique in worldwide according to various fields of studies [7, 8, 9, 10].

There are three objectives for this research, first to construct a model of price inflation for education item in Malaysia using Exponential Smoothing and Naïve and Autoregressive Integrated Moving Average (ARIMA) models. Secondly, to identify the best model for price inflation for education item in Malaysia and lastly to predict for price inflation within the period of four months on education items in Malaysia by using the best model.

2. Methodology

2.1 Simple Exponential Smoothing

Simple Exponential Smoothing is a time series forecasting method used for making predictions based on historical data. It's particularly useful when you have a time series dataset and want to make short-term forecasts. The positive values are considered by selecting seasonality multiplicative. The elimination of the linear trend, however, greatly reduced the complexity of these problems; the NN was challenged with coming up with a trend that was highly unlikely to be linear. Furthermore, depending on the frequency of the data, either non-seasonal (annual and daily data), single-seasonal (monthly, quarterly, and weekly data), or double-seasonal (hourly data) models were utilised.

Here are the various updating formulas:

Non-seasonal models:

$$l_{t=}\alpha y_t + (1-\alpha)l_{t-1} \tag{1}$$

Single seasonality models:

$$l_t = \frac{\alpha y}{s_t} + (1-\ \alpha)l_{t-1} \tag{2}$$

$$s_{t+K} = \frac{\beta y_t}{l_t} + (1-\ \beta)s_t \tag{3}$$

Double seasonality models:

$$l_t = \frac{\alpha y_t}{s_t u_t} + (1-\ \alpha)l_{t-1} \tag{4}$$

$$s_{t+K} = \frac{\beta y_t}{l_t u_t} + (1-\ \beta)s_t \tag{5}$$

$$u_{t+L} = \frac{\gamma y_t}{l_t s_t} + (1-\gamma)u_t \tag{6}$$

If Y_t the value of the series at time t, l_t and s_t are the level and seasonality components, and u_t is the second seasonality component, K represents the total number of observations over a given time frame which four for quarterly data, twelve for monthly data, and fifty-two for weekly data, while L represents the total number of observations over a given time frame. Remember that s_t and u_t are always on the plus side, while the smoothing coefficients, and can range from 0 to 1. Applying exp ( ) to the parameters of the initial seasonality components and sigmoid ( ) to the smoothing coefficients makes it simple to adhere to these constraints.

2.2 ARIMA

For time series analysis, the Box-Jenkins method was developed by Box and Jenkins [11]. This technique applies ARIMA models to non-stationary data that are then made stationary. Depending on the type of data being considered and the most appropriate but constrained parameter among the available model alternatives, the Box-Jenkins technique selects an ARIMA model using the series difference operation. An ARIMA model can be comprehended by first breaking down each of its components as described in the following:

ACFs	PACFs	Model
Decay to zero with exponential pattern	Cuts off after lag p	AR(p)
Cuts off after lag q	Decay to zero with exponential pattern	MA(q)
Decay to zero with exponential pattern	Decay to zero with exponential pattern	ARMA(p,q)
Cuts off after lag q	Cuts off after lag p	AAR(p) or MA(q)

Then, the model of ARIMA is indicated by the notation ARIMA (p, d, q), where:

- p = The terms of autoregressive.
- d = The terms of difference for stationarity.
- q = The terms of moving average.

2.3 Naïve Bayes

The Naive Bayesian Classifier is a classification that determines which class is the most optimal by determining which class has the most opportunities. This is done by employing a statistical model to compute the opportunities of a class that contains each group of qualities that are present, and then finding out which class has the most opportunities overall. According to this strategy, each and every feature will play a role in the decision-making process. The same major weighting attributes will be employed for each attribute, and each attribute will operate independently from the others [12]. The Bayes formula, which will now be provided in the following style, is the conceptual underpinning for the Naive Bayes Classifier theorem, which is used in programming. This theorem is used to classify data.
Different naive Bayes classifiers differ from one another based on the distributional assumptions they make. Despite what seem to be its overly simplistic assumptions, naive Bayes classifiers have excelled in a variety of real-world situations, most notably document classification and spam filtering. They simply need a small quantity of training data to predict the necessary parameters. Compared to more sophisticated methods, naive Bayes learners and classifiers can work at breakneck speed [13]. Due to the decoupling of the class conditional feature distributions, each distribution can be independently estimated as a one-dimensional distribution. This therefore helps to solve problems caused by the dimensionality curse.

In honour of the 18th-century British mathematician Thomas Bayes, the formula known as Bayes' theorem is used to calculate conditional probability.

$$P(A|B) = [P(B|A)\ P(A)]\ /\ P(B) \tag{7}$$

Here, P(A|B) is the probability of the occurrence of event A when event B occurs, where:

- P(A) is the probability of the occurrence of A,
- P(B|A) is the probability of the occurrence of event B when event A occurs,
- P(B) is the probability of the occurrence of B.

3. Results and Discussions

Time series forecasting is widely utilized in various fields like weather prediction, electricity demand estimation, earthquake anticipation, and financial market projection. When forecasting time series data, the aims is to estimate how the sequence of observations will continue. The data that has been chosen are from the Department of Statistics Malaysia (DOSM). The researcher selected the data from the website to perform the time series plot. The data selected is from January 2015 until March 2023. The aim for this plotting is to identify the trend or pattern of the series.

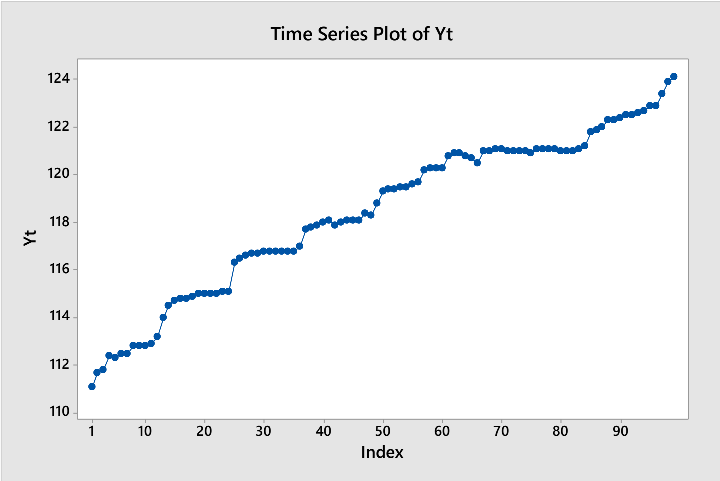

Fig. 1: *Time Series Plot for Education Price*

Fig.1 shows the time series chart has a clear upward trend. There may also be a slight curve in the data, as the increase in data values seems to accelerate over time. It indicates a consistent increase in the variable being measured over time.

3.1 ARIMA

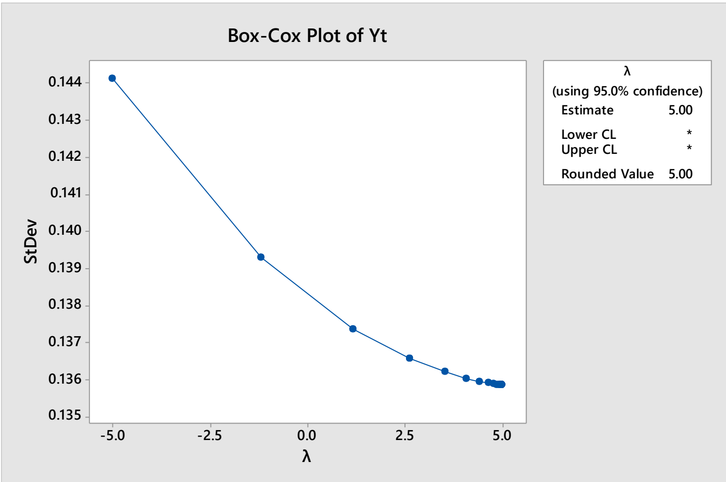

Fig. 2: *Box-cox Plot for Education*

Fig. 2 shows that the rounded value of λ is 5.00. Hence, the Box-Cox plot in Figure 2 indicates that the data's variance is not stable. For this reason, the Box-Cox transformation is needed. Since the rounded value of λ is 5.00, the price will be raised to the power of five in order to stabilize the variation.

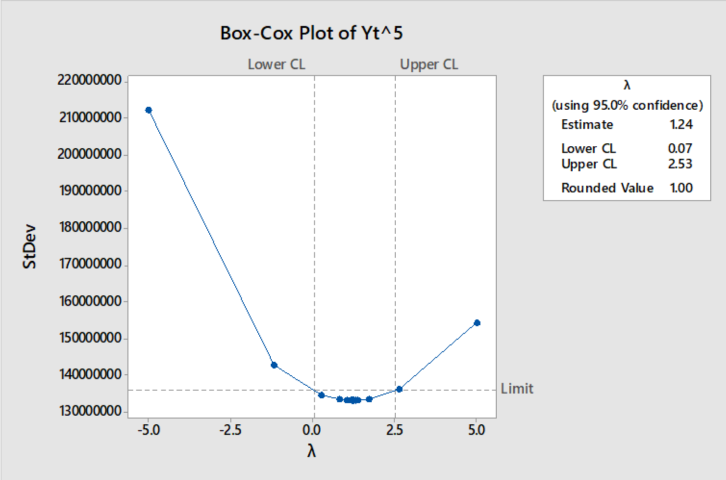

Fig. 3: *Box Cox Plot of Education Price after Transformation*

Fig. 3, the Box-Cox plot indicates that λ is equal to 2.53 for the upper control limit and 0.07 for the lower control limit. The variance of the data is steady since the rounded value of λ is equal to 1. Next, the autocorrelation function (ACF) and Partial Autocorrelation Function (PACF). The purpose of this plot is to confirm the steady nature of converted data.

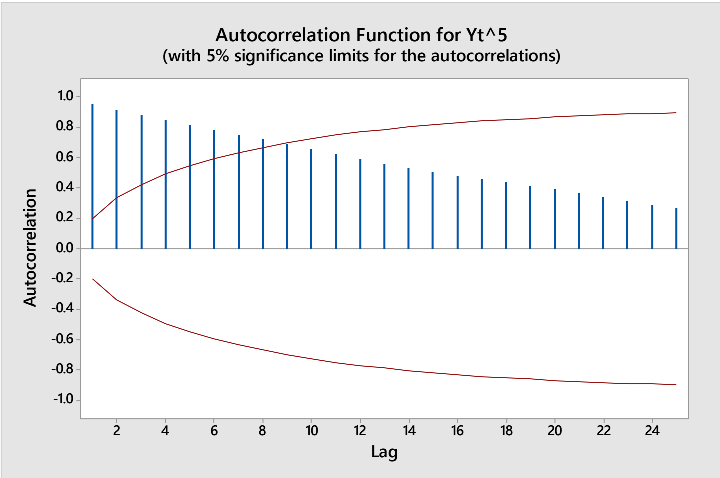

Fig. 4: *ACF and PACF Plot of Education Price After Transformation*

The data's ACF (Autocorrelation Function) plot is displayed in Fig. 4. The figure indicates that the data is decreasing to zero. Therefore, it can be concluded that there is non-stationarity in the mean. This requires differencing to guarantee that the mean remains steady.

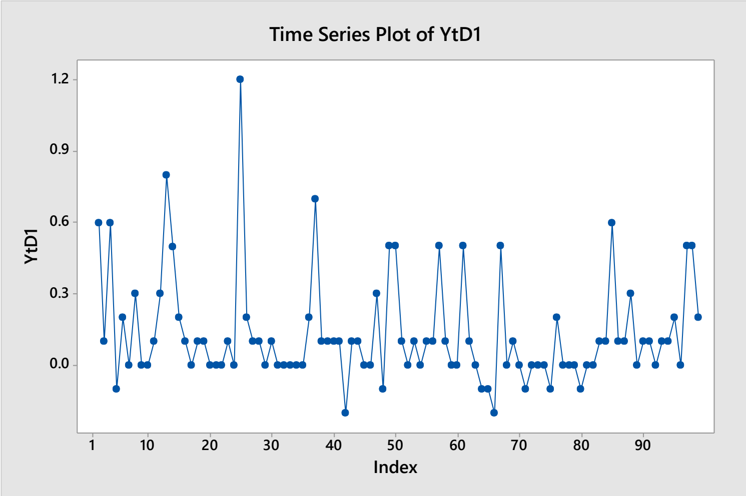

Fig. 5: *Time Series Plot of Education Price After Differencing*

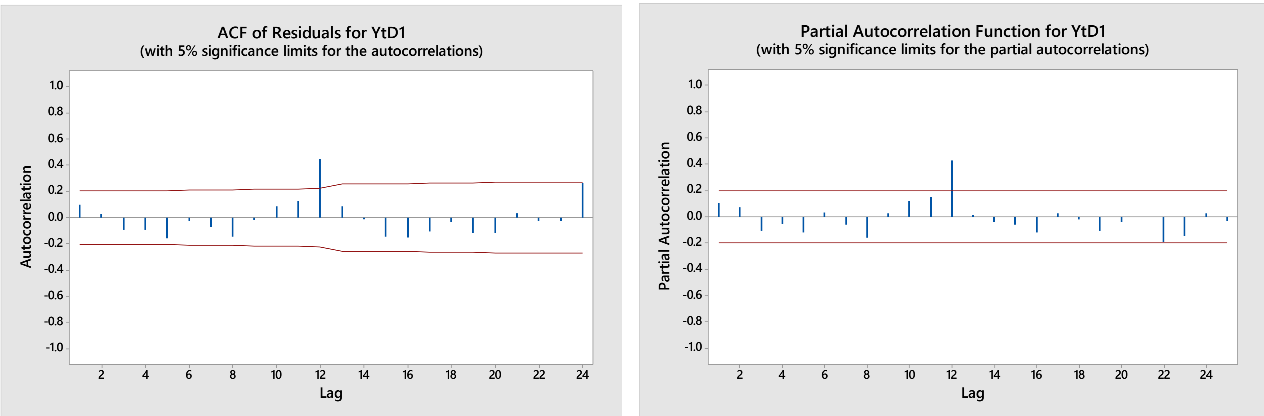

Fig. 6: *ACF and PACF Plot of Education Price After Differencing*

Model	*p*-value of parameters	*p*-value of Ljung-Box Test				MSE
		Lags 12	Lags 24	Lags 36	Lags 48	
(1,1,0)	AR (1) 0.000	0.000	0.000	0.000	0.000	0.07000
(1,1,1)	AR (1) 0.332 MA (1) 0.000	0.001	0.001	0.002	0.008	0.05170
(0,1,1)	MA (1) 0.000	0.000	0.000	0.000	0.001	0.05153

ARIMA Model (1,1,0) has AR (1) coefficient is highly significant (*p*-value = 0.000). There are no MA terms. Ljung-Box test results show high *p*-values (greater than 0.05) for all lags, indicating no significant autocorrelation in residuals with MSE value of 0.07000, which is the highest among the three models. While ARIMA Model (1,1,1), both AR (1) and MA (1) coefficients are significant (*p*-values close to 0.000). Ljung-Box test results show low *p*-values (0.001 to 0.008) at various lags, suggesting some residual autocorrelation with MSE value of 0.05170. Lastly, ARIMA Model (0,1,1) has MA (1) coefficient is highly significant which is *p*-value is 0.000. Ljung-Box test results show low *p*-values (0.000 to 0.001) at all lags, indicating potential issues with residual autocorrelation. The MSE value is 0.05153, which is slightly lower than the (1,1,1) model. Hence, (0,1,1) model might be the preferred choice due to its slightly lower MSE.

Enhanced Knowledge in Sciences and Technology Vol. 4 No. 1 (2024) p. 162-170

Forecasting using ARIMA (0,1,1)

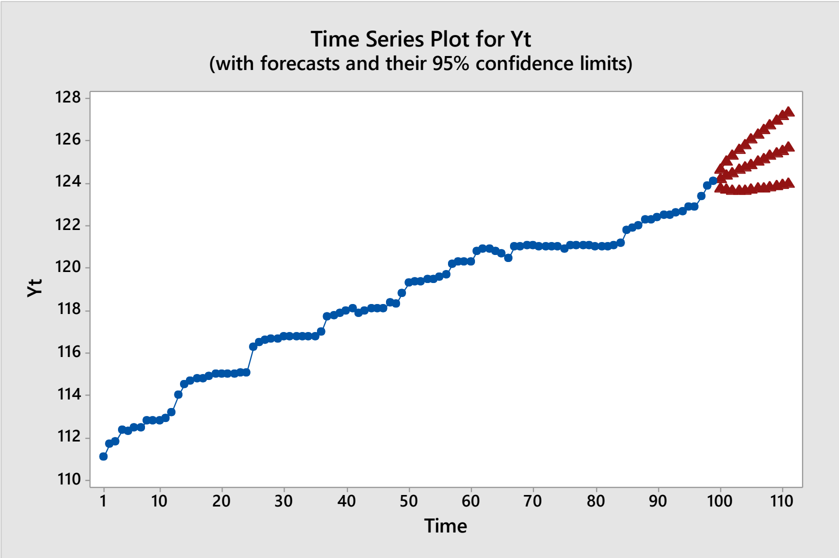

Fig. 7: *Time Series Plot for The Forecasting the Price Inflation of Education*

3.2 Simple Exponential Smoothing

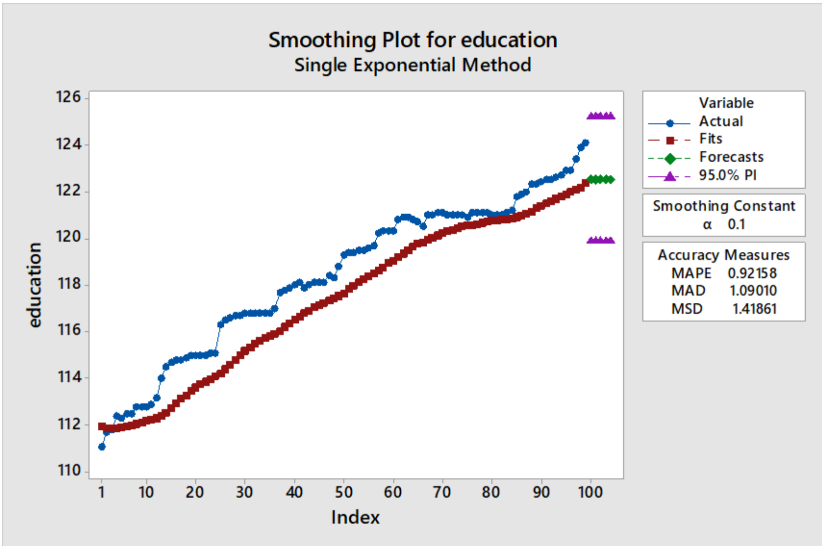

Fig. 8: *Time Series Plot for Education Item*

Fig. 8 shows that the predicting value for the next 30 days forward for price inflation for education item in Malaysia. By using simple exponential method, it shows that the pattern of the forecast value remains constant. In addition, the MSE value for the simple exponential method are 1.41861. In addition, the forward price inflation projection for education items in Malaysia is RM122.537 over the next 30 days, which is the same figure. This suggests that the simple exponential smoothing method is not the greatest way to estimate price inflation because the value for the following prediction is the same when implementing it.

3.3 Naïve Bayes

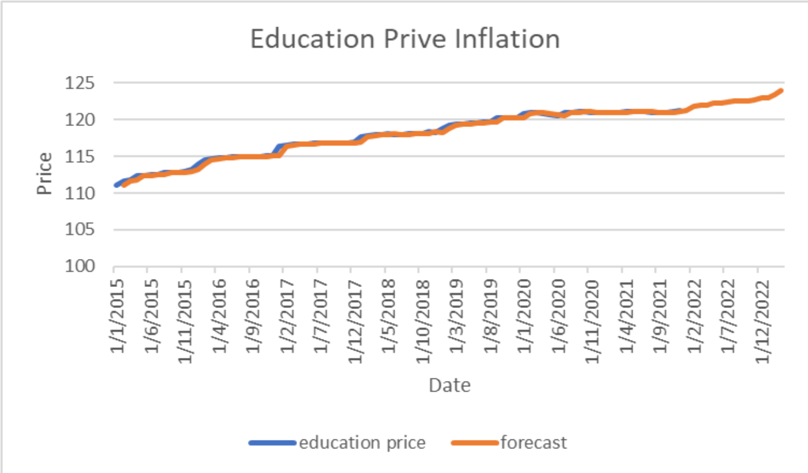

Fig. 9: *Naïve Bayes forecast for Education Item*

Fig. 9 shows that the forward price inflation projection for education items in Malaysia over the next 30 days, slightly different with the actual price. In addition, the MSE value for the naïve bayes method are 0.068757. The purpose of this study was to investigate the price inflation of educational products in Malaysia using the Naive Bayes classification approach. Known for its effectiveness and simplicity, the Naive Bayes algorithm performed admirably when it came to achieving our goals.

Table Comparison of forecasting performance

Accuracy Measure	Forecasting Method		
	Simple Exponential Smoothing	Naïve Bayes	ARIMA Model
MAE	1.09010	0.155102	0.00048
MSE	1.41861	0.068757	0.05153
MAPE	0.92158	0.211649	0.00482

The accuracy metric will be used to compare the three approaches the Naïve Bayes model, ARIMA model, and the Simple Exponential Smoothing model and ascertain which one predicts outcomes the best. To evaluate the forecasting capabilities of the Naïve Bayes method, ARIMA method, and Simple Exponential Smoothing method, three accuracy measures were used: mean absolute error (MAE), mean square error (MSE), and mean absolute percentage error (MAPE). The best options are the approaches with the lowest MAE, MSE, and MAPE values. The fact that accuracy metrics with lower values typically demonstrate more accuracy helps to explain this. Hence the best model among these three methods is ARIMA model.

4. Conclusions

In conclusion, a robust forecasting model to predict price inflation in the education sector of Malaysia has been constructed. This model incorporates Exponential Smoothing, Naïve, and Autoregressive Integrated Moving Average (ARIMA) techniques, ensuring a comprehensive and accurate representation of inflation trends within the education domain.

Moreover, these models have been evaluated using key performance metrics such as Mean Absolute Error (MAE), Mean Square Error (MSE), and Mean Absolute Percentage Error (MAPE). Through this assessment, we have identified the optimal model by selecting the approach with the lowest MAE, MSE, and MAPE values. This rigorous analysis guarantees the reliability and precision of our chosen forecasting method.

Furthermore, predictions of price inflation in Malaysia's education industry for the following 12 months have been construct by utilizing the ARIMA model's superior forecasting capabilities, we have produced priceless insights into the anticipated price movements that will help consumers, businesses, and policymakers navigate the economic landscape wisely and effectively while addressing the intricacies of Malaysia's education-related cost fluctuations.

For future researchers, here several suggestions can be made to improve the robustness and application of the study for this project, which outlier detection might be included for the next research. Outlier detection is a valuable tool for quality control, anomaly prevention, and gaining insights from data. Researchers and practitioners use a combination of methods and domain knowledge to identify and manage outliers effectively in their respective fields.

Acknowledgement

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

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Ashwaq Akhir:* ***Analysed and interpreted data.*** *Ashwaq Akhir, Kamil Khalid and Mohd Saifullah Rusiman. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Khan, S., & Alghulaiakh, H. (2020). ARIMA Model for Accurate Time Series Stocks Forecasting. International Journal of Advanced Computer Science and Applications, 11(7). https://doi.org/10.14569/ijacsa.2020.0110765

[2] Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. International Journal of Engineering Business Management, 10, 184797901880867. https://doi.org/10.1177/1847979018808673

[3] Bundick, B., & Pollard, E. (2019). The Rise and Fall of College Tuition Inflation. The Federal Reserve Bank of Kansas City Economic Review. https://doi.org/10.18651/er/1q19bundickpollard

[4] Clark, T. (n.d.). Comparing Measures of Core Inflation. https://core.ac.uk/download/pdf/6793514.pdf

[5] Mwanga, Y. (2020). Arima Forecasting Model for Uganda's Consumer Price Index. American Journal of Theoretical and Applied Statistics, 9(5), 238. https://doi.org/10.11648/j.ajtas.20200905.17

[6] Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. Data in Brief, 29, 105340. https://doi.org/10.1016/j.dib.2020.105340

[7] Loo, K., & Asrah, N. M. (2022). Survey on Customer Satisfaction Towards Courier Services in Johor. Enhanced Knowledge in Sciences and Technology, 2(2), 186-196.

[8] Zulkiflee, N. F., & Rusiman, M. S. (2021). Heart Disease Prediction Using Logistic Regression. Enhanced Knowledge in Sciences and Technology, 1(2), 177-184.

[9] Lim, L. S. Y. S., & Ismail, T. D. I. B. (2022). A Study On Volatility and Tail Risk of Small-Cap Companies in Comparison to Big-Cap Companies. Enhanced Knowledge in Sciences and Technology, 2(1), 241-249.

[10] Haron, N. A. A., & Kamardan, M. G. (2021). Queuing system of a busy restaurant using simulation software. Enhanced Knowledge in Sciences and Technology, 1(2), 66-71.

[11] Sidqi, F., & Sumitra, I. D. (2019). Forecasting Product Selling Using Single Exponential Smoothing and Double Exponential Smoothing Methods. IOP Conference Series: Materials Science and Engineering, 662(3), 032031. https://doi.org/10.1088/1757-899x/662/3/032031

[12] Parlina, I., Yusuf Arnol, M., Febriati, N. A., Dewi, R., Wanto, A., Lubis, M. R., & Susiani. (2019). Naive Bayes Algorithm Analysis to Determine the Percentage Level of visitors the Most Dominant Zoo Visit by Age Category. Journal of Physics: Conference Series, 1255(1), 012031. https://doi.org/10.1088/1742-6596/1255/1/012031

[13] Arif Yudianto, M. R., Agustin, T., Morgan James, R., Imannisa Rahma, F., Rahim, A., & Utami, E. (2021). Rainfall Forecasting to Recommend Crops Varieties Using Moving Average and Naive Bayes Methods. International Journal of Modern Education and Computer Science, 13(3), 23–33. https://doi.org/10.5815/ijmecs.2021.03.03