

Forecasting World Crude Oil Prices using the Fuzzy Time Series Method with a Comparison of the Chen and Lee Model

Dian Rahman Syahputra ^{1*}, Riza Arifudin ¹

¹Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Semarang, Indonesia

*Corresponding author: dianrahmansyahputra78@gmail.com

ARTICLE INFO

ABSTRACT

Article history

Received 12 February 2020

Revised 17 March 2020

Accepted 3 April 2020

Keywords

Fuzzy Time Series

World Crude Oil

Chen Model

Lee Model

In this study, the Fuzzy Time Series (FTS) method compared to Chen and Lee models is used to predict world crude oil prices. The goal is to determine which model results are best between Chen and Lee models in the fuzzy time series method in predicting world crude oil prices. In the calculation of FTS number and width specified intervals beginning of the process, the process is very influential to the outcome prediction. The method for determining the number and width of the interval that effectively is by using Rules Sturgess. So that the formation of fuzzy logical relationships will be appropriate and effective yield predictive results. Of the 50 trials that have been done using daily data from the Organization of the Petroleum Exporting Countries (OPEC), it is known that the FTS Lee model can predict better than the Chen model with a comparison of the results of the AFER fuzzy time series Lee model by 97.4% and RMSE of 1.617 and the results of the AFER fuzzy time series Chen model by 97.2% and RMSE of 1.693.

This is an open access article under the [CC-BY-SA](#) license.



1 Introduction

Indonesia is a country that abounds in various types of natural resources such as coal, copper, nickel, iron ore, tin ore, and others, not to mention crude oil and natural gas (Natanael, Safitri, & Suparti, 2018). Historically, oil has been used by humans since time immemorial. Crude oil (petroleum) is often referred to as black gold, in the form of a dark brown viscous liquid that is flammable (Wulandari, Wahyuningsih, & Amijaya, 2017). In production, each individual or industry requires energy (oil) to meet daily life needs and produce.

The increasing oil demand is high enough but does not offset the high production yields also demand the community to make every effort to get the oil to the needs of everyday life (Fauzannissa, Yasin, & Ispriyanti, 2016). Thus, in order to meet government fuel consumption doing an import. Nevertheless, the problem reappeared in relation to the price of oil imports. Oil prices often fluctuated imports caused prices to change often, so the government requires predictions about the price of crude oil with the smallest error value level or close to the real price to consider the country's budget (Natanael *et al.*, 2018).

One method to predict crude oil price is by using the Fuzzy Time Series (FTS). This method can capture the pattern of past data is then used to project into the future. In Fuzzy Time Series There are various models are models of Song, Chissom, Chen, and Lee. Song and Chissom first proposed the definition of FTS in 1993. Song and Chissom present time-invariant and variant models for the prediction of the Fuzzy Time Series.

In his research, Lee and colleagues presented this method for predicting the temperature-based FTS high-level two-factor (Lee, Wang, Chen, & Leu, 2006), and Chen proposed arithmetic operations to replace complicated operations in Song and Chissom models. Enrollment at the University of

Alabama has been selected as a target in all these models. Chen's model provides the best prediction results (Huarng, 2001).

This study aimed to compare the results of the world crude oil price forecast using the Chen and Lee model on FTS. The number and width specified intervals beginning of the process because the number and width of the interval are very influential in the outcome prediction so that the formation of Fuzzy Logical Relationship (FLR) will be right. Average Forecasting Error Rate (AFER) and Root Mean Square Error (RMSE) was used to test the prediction error ketepatan results.

2 Methods

2.1 Fuzzy Time Series (FTS)

FTS is a new concept proposed by Song and Chissom. In a recent study, Song and Chissom in 1991 proposed the definition of FTS and outlined modeling using fuzzy logic reasoning for the first time (Song & Chissom, 1993). Based on the FTS theory, Song and colleagues presented several prediction methods for estimating enrolment at the University of Alabama (Chen & Hsu, 2004).

The definition of FTS, according to Song and Chissom described as follows (Song *et al.*, 1993):

Definition 1: If $Y(t)$, with values $t = 0, 1, 2, \dots$, as a subset of R . Let $Y(t)$ be the universal set explained by the fuzzy set $\mu_i(t)$. If $Y(t)$ consists of $\mu_i(t)$, with the value $i = 1, 2, \dots$, then $F(t)$ is called the Fuzzy Time Series on $Y(t)$.

Definition 2: If $F(t+1) = A_i$ and $F(t) = A_j$, something FLR can be written as $A_i \rightarrow A_j$, i.e., A_i and A_j the left and right sides of the FLR.

2.2 Fuzzy Time Series (FTS) Chen Model

The FTS Chen model's prediction stages can be illustrated in a flowchart, as shown in Figure 1.

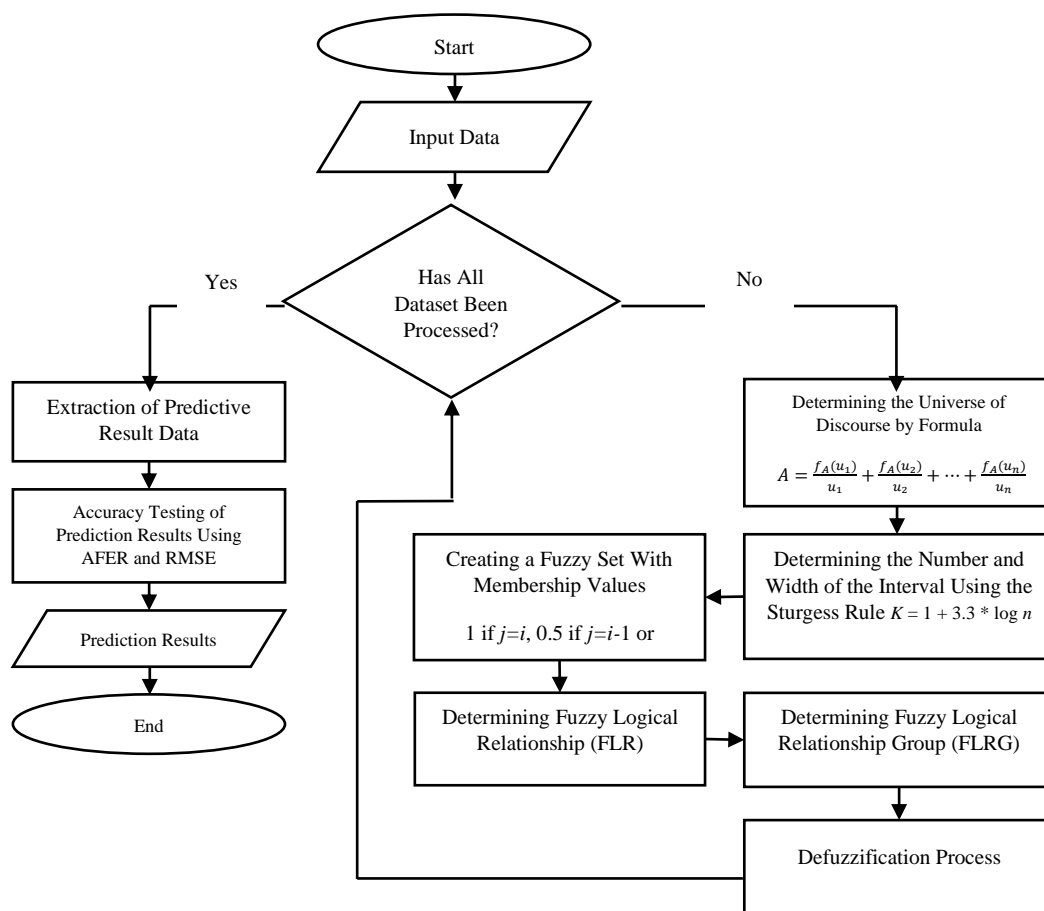


Figure 1. The flowchart of the FTS method model of Chen

The following is an explanation of the flowchart stages prediction of crude oil prices FTS model of Chen.

The steps in processing data using the servqual method in the triangular fuzzy number theory are as follows:

1. Determining the universe of discourse

The universe of discourse is usually expressed by the set U where $U = \{u_1, u_2, \dots, u_n\}$ then a fuzzy set A of U is defined as Equation 1.

$$A = (f_A(u_1))/u_1 + (f_A(u_2))/u_2 + \dots + (f_A(u_n))/u_n \quad (1)$$

Where f_A is the membership function of A , $f_A: U \rightarrow [0,1]$ and $f_A(u_i)$ show the degree of membership that is included in the fuzzy set A and $f_A(u_i) \in [0,1]$ with $1 \leq i \leq n$ [8].

2. Determine the number and width of the interval

The early stages of FTS calculation determine the interval's number and width because the process is very influential to the outcome prediction (Singh & Borah, 2013). The method for determining the interval's number and width is to use Sturgess rules that can be seen in Equation 2.

$$K = 1 + 3,3 * \log n \quad (2)$$

3. Creating a fuzzy set

Fuzzy set made to look for the membership value of the actual data that has been inputted with membership values 1 if $j = i$, 0.5 if $j = i-1$ or $j = i + 1$, and 0 if the others are.

4. Determining Fuzzy Logical Relationship (FLR)

FLR will give effect to the formation of Fuzzy Logical Relationship Group (FLRG) (Rukhansah, Muslim, & Arifudin, 2016). How to determine FLR $A_i \rightarrow A_j$ based on the value A_i that has been determined in the previous step, where A_i is year n and A_j is the year $n+1$ in the time series data (Tamrin, Noh, & Hamza, 2018).

5. Determining Fuzzy Logical Relationship Group (FLRG) based on grouping Chen model. Grouping Chen's model was to group all the results of FLR by taking one relationship if there are results of the same relation.

6. Defuzzification process and data extraction

Defuzzification process is a process in which researchers performed defuzzification and calculation of the predicted value from the calculation of FLR Group (Handayani & Anggriani). Extraction of the data was to look for predictive results that will be used in comparing the predicted results.

7. Evaluation

Evaluation is used to test the accuracy of prediction errors. Average Forecasting Error Rate (AFER) and Mean Square Error (MSE), which will be squared by a Root Mean Square Error (RMSE) are used in the evaluation process 8, 9, and 10.

$$AFER = \frac{|A_i - F_i|}{A_i} \times 100\% \quad (3)$$

$$MSE = \frac{\sum_{i=1}^n (A_i - F_i)^2}{n} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (A_i - F_i)^2}{n}} \quad (5)$$

2.3 Fuzzy Time Series (FTS) Lee Model

The prediction stages of the FTS Lee model can be illustrated in a flowchart as shown in Figure 2.

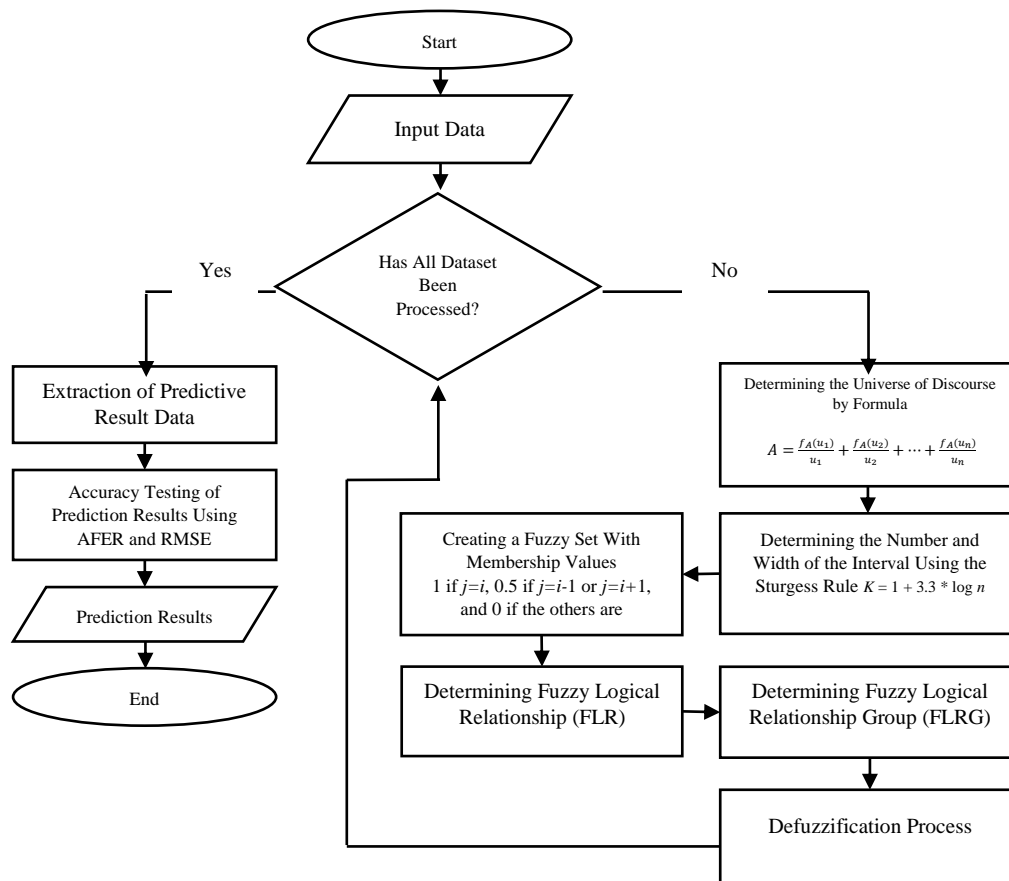


Figure 2. Flowchart of fuzzy time series method model of Lee

The following is an explanation of the flowchart stages prediction of crude oil prices FTS model of Lee.

1. Determining the Universe of Discourse

the universe of discourse is usually expressed by the set U where $U = \{u_1, u_2, \dots, u_n\}$ then, a fuzzy set A of U is defined as Equation 1.

$$A = \frac{f_A(u_1)}{u_1} + \frac{f_A(u_2)}{u_2} + \dots + \frac{f_A(u_n)}{u_n} \quad (6)$$

Where f_A is the membership function of A , $f_A: U \rightarrow [0,1]$ and $f_A(u_i)$ show the degree of membership that is included in the fuzzy set A and $f_A(u_i) \in [0,1]$ with $1 \leq i \leq n$.

2. Determine the number and width of the interval

The early stages of FTS calculation are to determine the interval's number and width because the process is very influential to the outcome prediction. The method for determining the interval's number and width is to use Sturgess rules that can be seen in Equation 2.

$$K = 1 + 3,3 * \log n \quad (7)$$

3. Creating a fuzzy set

Fuzzy set made to look for the membership value of the actual data that has been inputted with membership values 1 if $j = i$, 0.5 if $j = i-1$ or $j = i + 1$, and 0 if the others are.

4. Determining Fuzzy Logical Relationship (FLR)

Fuzzy Logical Relationship will give effect to the formation of Fuzzy Logical Relationship Group (FLRG) (Rukhansah *et al.*, 2016). How to determine Fuzzy Logical Relationship $A_i \rightarrow A_j$ based on the value A_i that has been determined in the previous step, where A_i is year n and A_j is the year $n+1$ in the time series data (Tamrin *et al.*, 2018).

5. Determining Fuzzy Logical Relationship Group (FLRG) based on the grouping Lee model. Grouping Lee model was to group all the Fuzzy Logical Relationship results by taking all relationships if there are results of the same relation.
6. Defuzzification Process and Data Extraction
Defuzzification process is a process in which researchers performed defuzzification and calculation of the predicted value from the calculation of FLRG. Extraction of the data was to look for predictive results that will be used in comparing the predicted results
7. Evaluation
Evaluation is used to test the accuracy of prediction errors. Average Forecasting Error Rate (AFER) and Mean Square Error (MSE), which will be squared by a Root Mean Square Error (RMSE) are used in the evaluation process 8, 9, and 10.

$$AFER = \frac{|A_i - F_i|/A_i}{n} \times 100\% \quad (8)$$

$$MSE = \frac{\sum_{i=1}^n (A_i - F_i)^2}{n} \quad (9)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (A_i - F_i)^2}{n}} \quad (10)$$

3 Results and Discussion

This study uses Fuzzy Time Series with a comparison of Chen and Lee's model to predict world crude oil prices as well as the accuracy of the prediction. These are calculated based on the smallest error AFER and RMSE with systems that have been created using PHP programming language with the help of application Sublime Text 3 and the Framework Laravel 5.7.

The data used in this research is secondary data, crude oil prices based on the price international (dollars). Data obtained from the Organization of the Petroleum Exporting Countries (OPEC) as many as 772 data reported daily world crude oil prices in 2016-2018 (three years). The dataset can be seen in Table 1.

Table 1. Dataset of world crude oil prices

Date	Price (USD)	Date	Price (USD)	Date	Price (USD)
04/01/2016	31,79	03/01/2017	53,13	13/12/2018	58,67
05/01/2016	31,21	04/01/2017	52,71	14/12/2018	59,07
...
15/01/2016	24,74	16/01/2017	52,17	28/12/2018	51,55

Once it is done, predictions using FTS Chen and Lee models by determining the set of the universe first. To determine the set of universes should be sought beforehand, the minimum and maximum values of the actual data of world crude oil prices entered. In this study, the data shown in Table 1 obtained a minimum value of 22.48 USD and a maximum of 84.09 USD, so $D_{min} = 22.48$ USD and $D_{max} = 84.09$ USD. Based on these values by using the formula in Equation 1 then the universe of discourse (U) can be defined as follows $U = (22,48,84,09)$.

Next, determine the number and width of the interval, as the number and width of the interval are very influential on the accuracy of the predicted outcome value. In this study, the authors use Sturges Rules to determine the number and width of the interval. Of the 772 sought from the data, minimum

and maximum values obtained a minimum value of 22.48 USD and a maximum of 84.09 USD. Having obtained the minimum and maximum values, subsequently applying Equation 2 to calculate the number of intervals. Then obtained a number of intervals for the data world crude oil prices by 11, with the following calculation:

$$K = 1+3,3*\log 772 = 10,52$$

The result of the calculation of the number of intervals using Rules Sturges is 10.52. Because the number of intervals must be in the form of an integer, then the calculation will be rounded to 11.

After a number of intervals are obtained, the next step is to find a wide interval. Then obtained the width of the interval for data on world crude oil prices of 5,6, with the calculation process as follows:

$$\text{Interval Width} = \frac{84,09 - 22,48}{11} = 5,6$$

Table 2. The number and width of the interval

No	Interval	Median (USD)
1	U ₁ = [22,48, 28,08]	25,28
2	U ₂ = [28,08, 33,68]	30,88
...
11	U ₁₁ = [78,48, 84,08]	81,28

After the number and width of the interval is obtained, then determine the fuzzy set with the fuzzy A_i membership function, which is between 0, 0.5, and 1. Where $1 = i = n$, n is the number of previously divided intervals. The matrix value is 1 if $j = i$, the matrix value is 0.5 if $j = i-1$ or $i+1$ and the matrix value is 0 if j has another value. The matrix of the formation of fuzzy sets can be seen in Table 3.

Table 3. matrix fuzzy set

A _{ij}	A _{1j}	A _{2j}	A _{3j}	A _{4j}	A _{5j}	A _{6j}	A _{7j}	A _{8j}	A _{9j}	A _{10j}	A _{11j}
A _{1j}	1	0,5	0	0	0	0	0	0	0	0	0
A _{2j}	0,5	1	0,5	0	0	0	0	0	0	0	0
...
A _{11j}	0	0	0	0	0	0	0	0	0	0,5	1

The matrix produces fuzzy set based on the following formula:

$$A_1 = \frac{a_{11}}{u_1} + \frac{a_{12}}{u_2} + \dots + \frac{a_{1n}}{u_n}$$

$$A_2 = \frac{a_{21}}{u_1} + \frac{a_{22}}{u_2} + \dots + \frac{a_{2n}}{u_n}$$

the actual data (preliminary data) with the data of world crude oil price forecasts from 2016 to 2018 using Lee's Fuzzy Time Series model.

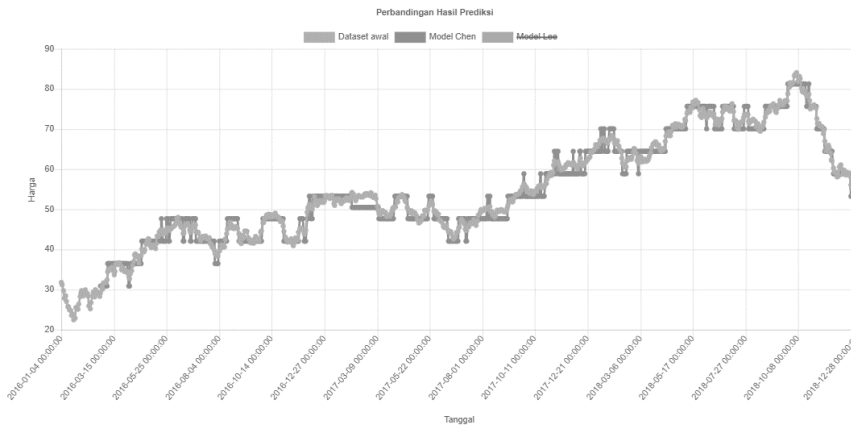


Figure 3. Comparison graph of actual data and FTS prediction data from Chen model

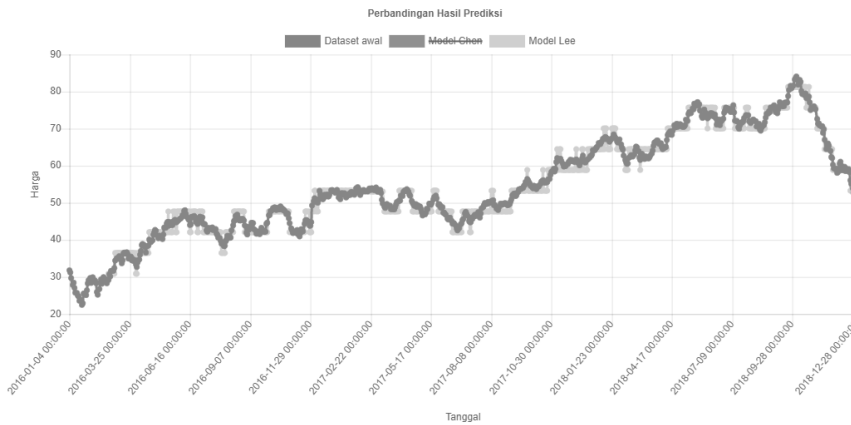


Figure 4. Comparison graph of actual data and FTS prediction data from lee model

To determine the level of error in the prediction results of the FTS Chen and Lee model and to determine which model is better in predicting world crude oil prices from 2016 to 2018 based on the smallest error precision level, the researchers used Average Forecasting Error Rate (AFER) and Root Mean Square Error (RMSE). The results of testing the FTS order 38 Chen model's prediction accuracy can be seen in Table 11, and the results of testing the prediction accuracy of the FTS order 38 Lee model can be seen in Table 12.

Table 11. Results of testing the accuracy of prediction of FTS order 38 Chen model

No	Date	Price (USD)	Prediction Results (USD)	Absolute Value	AFER	MSE
1	04/01/2016	31,79	-	-	-	-
...
283	03/02/2017	54,24	50,48	3,76	0,069	14,138
...
772	28/12/2018	51,55	53,28	1,73	0,034	2,993
On Average AFER and MSE					2,8%	2,865
RMSE Value						1,693

Table 12. Results of testing the accuracy of prediction of FTS order 38 Lee model

No	Date	Price (USD)	Prediction Results (USD)	Absolute Value	AFER	MSE
1	04/01/2016	31,79	-	-	-	-
...
283	03/02/2017	54,24	53,065	1,175	0,022	1,381
...
772	28/12/2018	51,55	53,28	1,73	0,034	2,993
On Average AFER and MSE					2,6%	2,615
RMSE Value						1,617

After testing the accuracy of FTS Order 38 Chen and Lee models' prediction, then obtained AFER Chen results by 2.8%, MSE of 2.865 and RMSE of 1.693, and AFER Lee results by 2.6%, MSE of 2.615, and RMSE of 1.617.

4 Conclusion

In predicting world crude oil prices using the FTS method with a comparison of Chen and Lee models on the world crude oil price data begins by entering data that will be predicted. After that, determine the universe of discourse, followed by determining the number and width of the interval using the Sturges Rules. The next step is to create a Fuzzy set to determine the membership value of each actual data entered, then determine FLR, then determine FLRG Chen model and Lee model. The next process is defuzzification and continued with the extraction of predictive result data, then testing the predicted results' accuracy using AFER and RMSE. After all the processes have been carried out, it can be concluded that the FTS Lee model can predict better than the FTS Chen model with the comparison of the AFER value Fuzzy Time Series Lee model by 97.4% and the RMSE of 1.617 and AFER value Fuzzy Time Series Chen model by 97.2% and the RMSE of 1.693.

References

- Chen, S. M. & Hsu, C. C. (2004). A New Method to Forecast Enrollments using Fuzzy Time Series. *International Journal of Applied Science and Engineering*, 2(3), 234-244. Retrieved from [http://ir.lib.cyut.edu.tw:8080/bitstream/310901800/8202/1/%E9%99%B3%E9%8C%AB%E6%98%8E\(234-244\).pdf](http://ir.lib.cyut.edu.tw:8080/bitstream/310901800/8202/1/%E9%99%B3%E9%8C%AB%E6%98%8E(234-244).pdf)
- Fauzannissa, R. A., Yasin, H. & Ispriyanti, D. (2016). Peramalan Harga Minyak Mentah Dunia menggunakan Metode Radial Basis Function Neural Network [Forecasting World Crude Oil Prices using the Radial Basis Function Neural Network Method]. *Jurnal Gaussian*, 5(1):193-202. Retrieved from <https://ejournal3.undip.ac.id/index.php/gaussian/article/view/11049>
- Handayani, L. & Anggriani, D. (2015). Perbandingan Model Chen Dan Model Lee pada Metode Fuzzy Time Series untuk Prediksi Harga Emas [Comparison of Chen's Model and Lee's Model on the Fuzzy Time Series Method for Gold Price Prediction]. *Jurnal Pseudocode*, 2(1), 28-36. doi:10.33369/pseudocode.2.1.28-36
- Huang, K. (2001). Heuristic Models of Fuzzy Time Series for Forecasting. *Elsevier Fuzzy Sets and Systems*, 123(3), 369-386. doi:10.1016/S0165-0114(00)00093-2

- Lee, L. W., Wang, L. H., Chen, S. M. & Leu, Y. H. (2006). Handling Forecasting Problems Based on Two-Factors High-Order Fuzzy Time Series. In *IEEE Transactions on Fuzzy Systems*, 14(3), 468-477. doi:10.1109/TFUZZ.2006.876367
- Natanael, D. K., Safitri, D. & Suparti. (2018). Prediksi Harga Minyak Dunia dengan Metode Autoregressive Fractionally Integrated Moving Average (ARFIMA) [World Oil Price Prediction Using the Autoregressive Fractionally Integrated Moving Average (ARFIMA) Method]. *Jurnal Statistika*, 6(1), 65-72. Retrieved from <https://jurnal.unimus.ac.id/index.php/statistik/article/view/3425>
- Rukhansah, N., Muslim, M. A. & Arifudin, R. (2016). Peramalan Harga Emas menggunakan Fuzzy Time Series Markov Chain Model [Gold Price Forecasting using Fuzzy Time Series Markov Chain Model]. *KOMPUTAKI*, 1(1), 56-74. Retrieved from <https://www.unaki.ac.id/ejournal/index.php/komputaki/article/view/113>
- Singh, P. & Borah, B. (2013). An Efficient Time Series Forecasting Model based on Fuzzy Time Series. *Elsevier Engineering Applications of Artificial Intelligence*, 26(10), 2443-2457. doi:10.1016/j.engappai.2013.07.012
- Song, Q. & Chissom, B. S. (1993). Forecasting Enrollments with Fuzzy Time Series - Part I. *Elsevier Fuzzy Sets and Systems*, 54(1), 1-16. doi:10.1016/0165-0114(94)90067-1
- Tamrin, H., Noh, J. & Hamza, S. (2018). Perbandingan Model Chen Dan Model Lee pada Metode Fuzzy Time Series untuk Prediksi Jumlah Ikan [Comparison of Chen's Model and Lee's Model in the Fuzzy Time Series Method for Predicting the Number of Fish]. *Jurnal Teknologi Informatika (J-TIFA)*, 1(1), 8-17. Retrieved from <http://jurnal.umm.ac.id/index.php/J-TIFA/article/view/116>
- Wulandari, A., Wahyuningsih, S. & Amijaya, F. D. T. (2017). Peramalan Harga Minyak Mentah Dunia (Crude Oil) Menggunakan Metode Radial Basis Function Neural Network (RBFNN). *Jurnal Eksponensial*, 8(2), 161-168. Retrieved from <http://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/37>