Automatic Clustering and Fuzzy Logical Relationship to Predict the Volume of Indonesia Natural Rubber Export

Widya Aprilini [a,1,*], Dian Palupi Rini [b,2], Hadipurnawan Satria [b,3]

[a] Student of Department of Informatics, Sriwijaya University, Palembang, Indonesia
[b] Lecturer of Department of Informatics, Faculty of Computer Science, Sriwijaya University, Palembang, Indonesia
[1] widyaprilini@gmail.com*; [2] dprini@unsri.ac.id; [3] hadipurnawan.satria@gmail.com
* corresponding author

ARTICLE INFO

Article history
Received July 5, 2022
Revised Jan 16, 2023
Accepted Feb 25, 2023

Keywords
Automatic Clustering
Exports
Fuzzy Time Series
Predicting
Natural Rubber

ABSTRACT

Natural rubber is one of the pillars of Indonesia's export commodities. However, over the last few years, the export value of natural rubber has decreased due to an oversupply of this commodity in the global market. To overcome this problem, it is possible to predict the volume of Indonesia natural rubber exports. Predicted values can also help the government to compile market intelligence for natural rubber commodities periodically. In this study, the prediction of the export volume of natural rubber was carried out using the Automatic Clustering as an interval maker in the Fuzzy Time Series or usually called Automatic Clustering and Fuzzy Logical Relationship (ACFLR). The data used is 51 data per year from 1970 to 2020. The purpose of this study is to predict the volume of Indonesia natural rubber exports and compare the prediction results between the Automatic Clustering and Fuzzy Logical Relationship (ACFLR) and Chen's Fuzzy Time Series. The results showed that there was a significant difference between the two methods, ACFLR got 0.5316% MAPE with $p = 11$ and Chen's Fuzzy Time Series model got 8.009%. Show that the ACFLR method performs better than the pure Fuzzy Time Series in predicting volume of Indonesia natural rubber exports.

1. Introduction

As a developing country, Indonesia certainly participates in exports and imports in the international market to create stable national economic growth. Economic growth can occur when export demand increases, thereby triggering an increase in domestic production [1,2]. Indonesia's export commodities are divided into oil and gas; and non-oil and gas exports. Where non-oil and gas exports over the last few years experienced a surplus, which is a favorable situation for Indonesia [1]. Non-oil and gas sector has a plantation sub-sector, which contains palm oil, natural rubber, coffee, and others [4].

Natural rubber as one of Indonesia's leading export commodities controls 27.95% of the world's natural rubber market share [3]. The average development of natural rubber exports shows a positive number, namely 2.29% [4]. Even so, the export value continued to decline due to the oversupply of natural rubber in the international market. This incident resulted in the loss of the mutually beneficial nature of export and import activities [5]. Therefore, in 2019, Indonesia and the two largest natural rubber supplying countries in the world agreed to reduce the volume of natural rubber exports. It aims to increase its value in the global market and maximize the absorption of natural rubber commodities for local industries [6]. Prediction of the volume of exports of natural rubber can be done as a reference figure for the reduction in the volume of exports. The figures obtained can also help the authorities in conducting marketing intelligence for natural rubber commodities periodically [1, 7, 8].

fuzzy time series is a method that can do a prediction based on time series data. The fuzzy time series method is a simple method that can do a prediction with only one variable and few data [9]. However, the interval formed is an interval with a static length [10]. In fact, according to [11], the optimal interval can improve the level of prediction accuracy. In this study, automatic clustering[1] is used as an interval maker in the fuzzy time series to get the optimal interval so that the level of prediction accuracy can increase.

2. Literature Study

a. Natural Rubber

Natural rubber is a processed product derived from thickening the sap of the rubber plant or commonly called latex [12, 15]. There are various kinds of rubber-producing plants, but *Havea brasiliensis* dominates the natural rubber market [13, 14, 15]. *Havea brasiliensis* or rubber plant was originally a wild plant that was first found among canopied tree groups in the Amazon river basin, South America [16]. In 1876, Henry A. Wickham, an English explorer, brought rubber seeds from South America for distribution to South Asia and Asia [17]. Significant growth of rubber plantations occurred in 1910 after planting in Southeast Asia including Bogor, Indonesia. Rubber plantations then developed commercially in 1918 on the East Coast of Sumatra along with the increasing demand for natural rubber in the market [16, 18].

Natural rubber processing begins with mixing the sap with chemicals to control viscosity and color in a large tank. Then to clot it, the sap is given a coagulant (formic acid). The solid latex resulting from the coagulation is then processed according to the desired semi-finished natural rubber form. Be it in the form of sheets, crepes, or rubber blocks [19]. In addition, there is also natural rubber with special technical specifications or commonly called Technically Specified Natural Rubber (TSNR). TSNR originated from the standards set by the International Standards Organization (ISO). The indicators assessed in determining the standard are dirt content, ash content, volatile matter, nitrogen content, initial Wallace plasticity (Po), plasticity retention index (PRI), and Mooney viscosity.

In international trade, every year, around 70% of Indonesia's natural rubber production is diverted for export [19]. Natural rubber exports have been recorded since 1970 with an export volume of 581,190 tons and a value of 185,164 USD. This figure continues to increase until 2020, Indonesia's natural rubber export volume reaches 2,280,090 tons with a value of 3,010,245 USD [4]. Several factors that determine Indonesia's natural rubber exports are: the performance of domestic natural rubber production; absorption of natural rubber for domestic industry; exchange rate; inventory, price, and export volume of the previous period; the policy of the authorities; and domestic prices in importing countries [20].

b. Fuzzy Time Series

In this section, the steps of Chen's fuzzy time series are described as follows [21, 22, 23]:

Step 1. Define Universe of Discourse (U).

$$U = [D_{min} - d_1, D_{max} + d_2] \tag{1}$$

Where:

U = universe of discourse
D_{min} = smallest data
D_{max} = largest data
d_1, d_2 = two proper positive numbers

Step 2. Determine the intervals and their length. The number of intervals is rounded off from the calculation result with the sturgess rule, namely:

$$k = (1 + 3{,}322 \times log\, n) \tag{2}$$

Where:

[1] Chen, S. M., Wang, N. Y., & Pan, J. S. (2009). Forecasting Enrollments Using Automatic Clustering Techniques and Fuzzy Logical Relationships. Expert Systems with Applications, 36(8), 11070–11076. https://doi.org/10.1016/j.eswa.2009.02.085

k = number of interval
n = amount of data

Then determine the length of each interval by:

$$i = \frac{(D_{max} - D_{min})}{k} \quad (3)$$

Where:
i = interval length
D_{min} = smallest data
D_{max} = largest data

Step 3. Data fuzzification. Based on interval formed $(u_1, u_2, \dots, u_n)$, define each fuzzy set A_k where $1 \le k \le n$, as follows:

$$A_1 = \frac{1}{u_1} + \frac{0,5}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \cdots + \frac{0}{u_{n-1}} + \frac{0}{u_n},$$
$$A_2 = \frac{0,5}{u_1} + \frac{1}{u_2} + \frac{0,5}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \cdots + \frac{0}{u_{n-1}} + \frac{0}{u_n},$$
$$A_3 = \frac{0}{u_1} + \frac{0,5}{u_2} + \frac{1}{u_3} + \frac{0,5}{u_4} + \frac{0}{u_5} + \cdots + \frac{0}{u_{n-1}} + \frac{0}{u_n},$$
$$A_4 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0,5}{u_3} + \frac{1}{u_4} + \frac{0,5}{u_5} + \cdots + \frac{0}{u_{n-1}} + \frac{0}{u_n},$$
$$A_n = \frac{0}{u_1} + \frac{0,5}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \cdots + \frac{0,5}{u_{n-1}} + \frac{1}{u_n} \quad (4)$$

If the datum belongs to u_k, then the datum is fuzzified into A_k

Step 4. Derive a fuzzy logical relationship (FLR). If (d) is fuzzified as A_i and $(d+1)$ fuzzified as A_j, then the fuzzy logical relationship is $Ai \rightarrow Aj$.

Step 5. Collect the same relations in one group. The group called fuzzy logical relationship group (FLRG).

Step 6. Calculate the prediction of data 'd' by following principle:

Principle 1. If the fuzzified of data 'd' is A_i and there is only one relation of A_i or known as $Ai \rightarrow Aj$., then the prediction of data '$(d+1)$' is m_j or known as middle point of interval u_j.

Principle 2. If the fuzzified of data 'd' is A_i and there are more than one relations of A_i or known as: $A_i \rightarrow A_{j1}(x_1), A_{j2}(x_2), A_{j3}(x_3), \dots, A_{jn}(x_n)$, then the prediction of data '$(d+1)$' shown as follows:

$$\frac{(x_1 \times m_{j1} + x_2 \times m_{j2} + x_3 \times m_{j3} + \cdots + x_n \times m_{jn})}{x_1 + x_2 + x_3 + \cdots + x_n} \quad (5)$$

Principle 3. If the fuzzified of data 'd' is A_i and there is no relations of A_i or known as $A_i \rightarrow \#$, then the prediction of data '$(d+1)$' is m_i or known as middle point of interval u_i.

c. Automatic Clustering

Algorithm of automatic clustering based on [10, 24, 25] shown as follows:

Step 1. Sort the data in an ascending sequence with no duplicate data. Then calculate the value of average_dif. Where average_dif is an average of differences between every pair of data. Average_dif calculate by equation 6:

$$\text{average\_dif} = \frac{(\sum_{i=1}^{n-1} d_{i+1} + d_i)}{n-1} \quad (6)$$

Step 2. First data (smallest data) is a member of first cluster. The cluster then known as current cluster. The determination of the next clusters is based on:

Principle 1. Assume that the current cluster is a first cluster and there is only one datum d_1 in it.

$$\{d_1\}, d_2, d_3, d_4, \dots, d_n$$

If $d_2 - d_1 \le average\_dif$, then d_2 is belongs to current cluster. Otherwise, create a new cluster for d_2 and this cluster is now a current cluster.

Principle 2. Assume that the current cluster is not the first cluster and there is only one datum d_j in it. Assume that d_i is the largest datum in the 'before current cluster' cluster and d_k is the datum to be clustered.

$$\{d_1, d_2\}, \{d_3, \ldots, d_i\}, \{d_j\}, d_k, \ldots, d_n$$

If $d_k - d_j \leq average\_dif$ and $d_k - d_j < d_j - d_i$, then d_k is belongs to current cluster with d_j. Apart from that, create a new cluster for d_k and this cluster is now a current cluster.

Principle 3. Assume that the current cluster is not the first cluster and tere are more than one datum in it. Assume that d_i is the largest datum in current cluster and d_j is the datum to be clustered.

$$\{d_1, d_2\}, \{d_3, \ldots, d_i\}, d_j, d_k, \ldots, d_n$$

If $d_j - d_i \leq average\_dif$ and $d_j - d_i \leq cluster\_dif$, then put d_j in current cluster. Otherwise, create a new cluster for d_j and this cluster is now a current cluster. Cluster_dif denotes the average difference of the distances between every pair of adjacent data in the cluster or in other words, cluster_dif is an average_dif for cluster.

$$cluster\_dif = \frac{(\sum_{i=1}^{n-1} c_{i+1} + c_i)}{n-1} \tag{7}$$

Where c_{i+1} and c_i denotes the data in the current cluster.

Step 3. Based on clustering result obtained in **step 2**, adjust the contents of clusters by following principles:

Principle 1. If there are more than two data in the cluster, then we keep the smallest datum, keep the largest datum and remove the others.

Principle 2. If the cluster has exactly two data, then leave it unchanged.

Principle 3. If there is only one datum d_j, then put the value of "$d_j - average\_dif$" and "$d_j + average\_dif$" to the cluster. After that, adjust the cluster by following conditions:

Condition 1. If this happen in the first cluster, then remove the value of "$d_j - average\_dif$" from the cluster.

Condition 2. If this happen in the last cluster, then remove the value of "$d_j + average\_dif$" from the cluster.

Condition 3. If the value of "$d_j - average\_dif$" is smaller than the smallest value in its antecedent cluster, then undo all the action in **Principle 3**.

Step 4. Assume that the clusters obtain from **Step 3** is shown as follows:

$$\{d_1, d_2\}, \{d_3, d_4\}, \{d_5, d_6\}, \ldots, \{d_i, d_j\}, \{d_k\} \ldots, \{d_n\}$$

Transform these clusters into intervals by following sub-step:

Step 4.1. Transform the first cluster $\{d_1, d_2\}$ into interval $[d_1, d_2$int)

Step 4.2. If the current interval $[d_g, d_h)$ dan current cluster $\{d_i, d_j\}$, then adjust by following condition:

Condition 1. if $d_h \geq d_i$, then create an interval $[d_h, d_j)$ as we know as current interval now and the next cluster $\{d_k, d_l\}$ is current cluster.

Condition 2. If $d_h < d_i$, then create an interval $[d_i, d_j)$ as we know as current interval now and one another interval before that as $[d_h, d_i)$. If the cluster after current interval $\{d_k\}$, then create a new interval $[d_i, d_k)$, let it be the current interval and let the next cluster be current cluster.

Condition 3. Let the last interval be $[d_m, d_n]$.

Step 4.3. Repeatedly do the step until all clusters have been transformed into intervals.

Step 5. Divide each obtained interval into p sub-intervals, where $p \geq 1$.

3. Methodology

a. Data Collection

The data used for this research is in the form of annual time series data on the volume of Indonesian natural rubber exports. The primary data was collected as many as 51 data, of which the data from 1970 to 2019 came from the publications of the Central Statistics Agency which were summarized by the Directorate General of Plantation, Ministry of Agriculture of the Republic of Indonesia in the plantation statistics book and for 2020 it was taken from the publication of the

Central Statistics Agency[2] entitled "Statistical Bulletin Export Foreign Trade According to HS, December 2020" because it has not been summarized in the plantation statistics book. The data obtained are in the form of the year and volume of Indonesia's natural rubber exports in tons.

b. Data Processing

Table 1 shows the data used in this research, followed by steps in automatic clustering and fuzzy logical relationship applied to the data.

Table 1. Actual data

Year	Data	Year	Data	Year	Data	Year	Data	Year	Data
1970	581,190	1981	812,800	1992	1,267,605	2003	1,662,210	2014	2,623,471
1971	580,232	1982	797,608	1993	1,214,568	2004	1,874,261	2015	2,630,313
1972	755,960	1983	938,032	1994	1,244,950	2005	2,024,593	2016	2,578,791
1973	866,638	1984	1,009,558	1995	1,324,295	2006	2,286,897	2017	2,991,909
1974	794,741	1985	987,771	1996	1,434,285	2007	2,407,972	2018	2,812,105
1975	788,292	1986	958,692	1997	1,404,010	2008	2,283,158	2019	2,503,671
1976	789,892	1987	1,092,525	1998	1,641,186	2009	1,991,533	2020	2,280,090
1977	781,967	1988	1,132,132	1999	1,494,543	2010	2,351,915		
1978	865,960	1989	1,151,409	2000	1,379,612	2011	2,556,233		
1979	865,321	1990	1,077,331	2001	1,453,382	2012	2,444,503		
1980	976,131	1991	1,220,020	2002	1,495,987	2013	2,701,995		

Automatic Clustering and Fuzzy Logical Relationship Algorithm is applied by combining automatic clustering algorithm and fuzzy time series (step 3 – step 6) or shown by "Fig. 1".

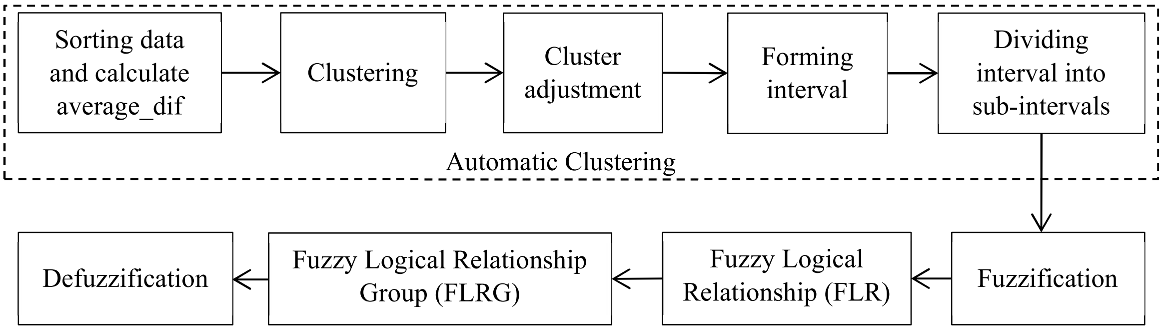

Fig. 1. Automatic Clustering and Fuzzy Logical Relationship Algorithm

[Step 1] Sorting data to the ascendant and calculate the average_diff "(6)". If there is duplicate data, remove one of them.

$$average_{dif} = \frac{(581190 - 580232) + \cdots + (2991909 - 2812105)}{51 - 1} = \frac{2411677}{50} = 47286.784$$

[Step 2] Do the clustering process based on principles of **Step 2**.
[Step 3] Adjust the content of clusters obtained from [**Step 2]** based of three principles of **Step 3.** Results showns as follows:
{580232, 581190}, {755960, 797608}, {812800}, {865321, 865960},{866638}, {938032, 987771}, {1009558}, {1077331, 1092525}, {1132132, 1151409}, {1214568, 1220020}, {1244950, 1267605}, {1277008.216, 1371581.784}, {1379612, 1404010}, {1434285, 1453382}, {1494543, 1495987}, {1641186, 1662210}, {1826974.216, 1921547.784}, {1991533, 2024593}, {2280090, 2283158}, {2286897}, {2304628.216, 2399201.784}, {2407972, 2444503}, {2456384.216, 2550957.784}, {2556233, 2578791}, {2623471, 2630313}, {2654708.216, 2749281.784}, {2764818.216, 2859391.784}, {2944622.216, 2991909}

[2] www.bps.go.id

[Step 4] Forming the intervals and the mid point of the interval based on **Step 4.** The intervals obtained shown as follows:

$u_1 = [580232, 581190)$ $m_1 = 580711$ $u_{44} = [2749281.784, 2764818.216)$ $m_{44} = 2757050$
$u_2 = [581190, 755960)$ $m_2 = 668575$ $u_{45} = [2764818.216, 2859391.784)$ $m_{45} = 2812105$
$u_3 = [755960, 812800)$ $m_3 = 784380$ $u_{46} = [2859391.784, 2944622.216)$ $m_{46} = 2902007$
$u_4 = [865321, 866636)$ $m_4 = 865979.5$ $u_{47} = [2944622.216, 2991909]$ $m_{47} = 2968265.608$
...

[**Step 5]** Divide each obtained interval into p sub-intervals, where $p \geq 1$. If the value of p is two, then we divide one interval into two intervals.
E.g. $u_1 = [580232, 581190)$ $m_1 = 580711$, become $u_1 = [580232, 580471.5)$ $m_1 = 580351.5$ and $u_2 = [580471.5, 581190)$ $m_2 = 580830.75$ and so on. We use $p = 11$ because based on MAPE value in **Table 2** and **Table 3**, $p = 11$ produce the smallest MAPE value out of 12 sub-intervals (p). In addition, too many intervals causes complexity and reduces the essence of fuzzy time series [26, 27, 28, 29].

Table 2. Prediction based on Value of *p*

Year	**Actual Data**	**Value of *p***					
		***p*=1**	***p*=2**	***p*=3**	***p*=4**	***p*=5**	***p*=6**
1970	581190	-	-	-	-	-	-
1971	580232	580711	580471,4	580391.5	580351.8	580237.8	580311.8
1972	755960	784380	770170	765433.3	763065	761644	760696.7
1973	866638	849208.3	875068,4	878537	875562.3	873777.4	872587.5
...	...	...	...	...	...	...	...
2017	2991909	2968265.6	2968265.6	2968265.6	2968265.6	2968265.6	2968265.6
2018	2812105	2812105	2812105	2796342.7	2800283.3	2793190.3	2796342.7
2019	2503671	2503671	2503671	2487908.7	2491849.3	2484756.3	2487908.7
2020	2280090	2283493	2281791.8	2281224.5	2280940.9	2280770.7	2280657.3
MAPE		2.5352%	1.7640%	1.2767%	1.3017%	1.0785%	0.8555%

Table 3. Prediction based on Value of *p* - continued

Year	**Actual Data**	**Value of *p***					
		***p*=7**	***p*=8**	***p*=9**	***p*=10**	***p*=11**	***p*=12**
1970	581190	-	-	-	-	-	-
1971	580232	580300.4	580291.9	580285.2	580279.9	580275.6	580271.9
1972	755960	760020	759512.5	759117.8	758802	758543.6	758328.3
1973	866638	871737.6	871100.1	870604.3	870207.7	869883.2	869612.8
...	...	...	...	...	...	...	...
2017	2991909	2968265.6	2968265.6	2968265.6	2968265.6	2968265.6	2968265.6
2018	2812105	2791839.2	2794372.5	2791088.7	2793190.3	2790611	2792402.2
2019	2503671	2483405.2	2485938.5	2482654.7	2484756.3	2482177	2483968.2
2020	2280090	2280576.2	2280515.4	2280468.2	2280430.4	2280399.4	2280373.6
MAPE		0.7155%	0.8467%	0.8470%	0.5409%	**0.5316%**	0.8202%

[Step 6] Fuzzify each datum into fuzzy sets they belong. The fuzzified data shown in Table 4.

Table 4. Fuzzification

Data	Fuzzy Set	Data	Fuzzy Set	Data	Fuzzy Set	Data	Fuzzy Set	Data	Fuzzy Set
581,190	A_{12}	812,800	A_{34}	1,267,605	A_{166}	1,662,210	A_{276}	2,623,471	A_{441}
580,232	A_{1}	797,608	A_{31}	1,214,568	A_{133}	1,874,261	A_{292}	2,630,313	A_{452}
755,960	A_{23}	938,032	A_{67}	1,244,950	A_{155}	2,024,593	A_{320}	2,578,791	A_{430}
866,638	A_{56}	1,009,558	A_{78}	1,324,295	A_{182}	2,286,897	A_{342}	2,991,909	A_{517}
794,741	A_{30}	987,771	A_{74}	1,434,285	A_{221}	2,407,972	A_{375}	2,812,105	A_{490}
788,292	A_{29}	958,692	A_{70}	1,404,010	A_{210}	2,283,158	A_{335}	2,503,671	A_{402}
789,892	A_{29}	1,092,525	A_{100}	1,641,186	A_{265}	1,991,533	A_{309}	2,280,090	A_{331}
781,967	A_{28}	1,132,132	A_{111}	1,494,543	A_{243}	2,351,915	A_{358}		
865,960	A_{50}	1,151,409	A_{122}	1,379,612	A_{199}	2,556,233	A_{419}		
865,321	A_{45}	1,077,331	A_{89}	1,453,382	A_{232}	2,444,503	A_{386}		
976,131	A_{72}	1,220,020	A_{144}	1,495,987	A_{254}	2,701,995	A_{468}		

[Step 7] Denote the fuzzy logical relationship (FLR).
$A_{12} \rightarrow A_{1}$, $A_{1} \rightarrow A_{23}$, $A_{23} \rightarrow A_{56}$, $A_{56} \rightarrow A_{30}$, $A_{30} \rightarrow A_{29}$, $A_{29} \rightarrow A_{29}$, $A_{29} \rightarrow A_{28}$, $A_{28} \rightarrow A_{50}$, $A_{50} \rightarrow A_{45}$, $A_{45} \rightarrow A_{72}$, $A_{72} \rightarrow A_{34}$, $A_{34} \rightarrow A_{31}$, $A_{31} \rightarrow A_{67}$, $A_{67} \rightarrow A_{78}$, $A_{78} \rightarrow A_{74}$, $A_{74} \rightarrow A_{70}$, $A_{70} \rightarrow A_{100}$, $A_{100} \rightarrow A_{111}$, $A_{111} \rightarrow A_{122}$, $A_{122} \rightarrow A_{89}$, $A_{89} \rightarrow A_{144}$, $A_{144} \rightarrow A_{166}$, $A_{166} \rightarrow A_{133}$, $A_{133} \rightarrow A_{155}$, $A_{155} \rightarrow A_{182}$, $A_{182} \rightarrow A_{221}$, $A_{221} \rightarrow A_{210}$, $A_{210} \rightarrow A_{265}$, $A_{265} \rightarrow A_{243}$, $A_{243} \rightarrow A_{199}$, $A_{199} \rightarrow A_{232}$, $A_{232} \rightarrow A_{254}$, $A_{254} \rightarrow A_{276}$, $A_{276} \rightarrow A_{292}$, $A_{292} \rightarrow A_{320}$, $A_{320} \rightarrow A_{342}$, $A_{342} \rightarrow A_{375}$, $A_{375} \rightarrow A_{335}$, $A_{335} \rightarrow A_{309}$, $A_{309} \rightarrow A_{358}$, $A_{358} \rightarrow A_{419}$, $A_{419} \rightarrow A_{386}$, $A_{386} \rightarrow A_{468}$, $A_{468} \rightarrow A_{441}$, $A_{441} \rightarrow A_{452}$, $A_{452} \rightarrow A_{430}$, $A_{430} \rightarrow A_{517}$, $A_{517} \rightarrow A_{490}$, $A_{490} \rightarrow A_{402}$, $A_{402} \rightarrow A_{331}$, $A_{331} \rightarrow A\#$.

[Step 8] Collect the same relations in one group. The group called fuzzy logical relationship group (FLRG). The group obtained shown in Table 5.

Table 5. Fuzzy Logical Relationship Group

No	Fuzzy Set	No	Fuzzy Set	No	Fuzzy Set	No	Fuzzy Set	No	Fuzzy Set
1	$A_{1} \rightarrow A_{23}$	**11**	$A_{56} \rightarrow A_{30}$	**21**	$A_{133} \rightarrow A_{155}$	**31**	$A_{254} \rightarrow A_{276}$	**41**	$A_{375} \rightarrow A_{335}$
2	$A_{12} \rightarrow A_{1}$	**12**	$A_{67} \rightarrow A_{78}$	**22**	$A_{144} \rightarrow A_{166}$	**32**	$A_{265} \rightarrow A_{243}$	**42**	$A_{386} \rightarrow A_{468}$
3	$A_{23} \rightarrow A_{56}$	**13**	$A_{70} \rightarrow A_{100}$	**23**	$A_{155} \rightarrow A_{182}$	**33**	$A_{276} \rightarrow A_{292}$	**43**	$A_{402} \rightarrow A_{331}$
4	$A_{28} \rightarrow A_{50}$	**14**	$A_{72} \rightarrow A_{34}$	**24**	$A_{166} \rightarrow A_{133}$	**34**	$A_{292} \rightarrow A_{320}$	**44**	$A_{419} \rightarrow A_{386}$
5	$A_{29} \rightarrow A_{28}(1)$, $A_{29}(1)$	**15**	$A_{74} \rightarrow A_{70}$	**25**	$A_{182} \rightarrow A_{221}$	**35**	$A_{309} \rightarrow A_{358}$	**45**	$A_{430} \rightarrow A_{517}$
6	$A_{30} \rightarrow A_{29}$	**16**	$A_{78} \rightarrow A_{74}$	**26**	$A_{199} \rightarrow A_{232}$	**36**	$A_{320} \rightarrow A_{342}$	**46**	$A_{441} \rightarrow A_{452}$
7	$A_{31} \rightarrow A_{67}$	**17**	$A_{89} \rightarrow A_{144}$	**27**	$A_{210} \rightarrow A_{265}$	**37**	$A_{331} \rightarrow A\#$	**47**	$A_{452} \rightarrow A_{430}$
8	$A_{34} \rightarrow A_{31}$	**18**	$A_{100} \rightarrow A_{111}$	**28**	$A_{221} \rightarrow A_{210}$	**38**	$A_{335} \rightarrow A_{309}$	**48**	$A_{468} \rightarrow A_{441}$
9	$A_{45} \rightarrow A_{72}$	**19**	$A_{111} \rightarrow A_{122}$	**29**	$A_{232} \rightarrow A_{254}$	**39**	$A_{342} \rightarrow A_{375}$	**49**	$A_{490} \rightarrow A_{402}$
10	$A_{50} \rightarrow A_{45}$	**20**	$A_{122} \rightarrow A_{89}$	**30**	$A_{243} \rightarrow A_{199}$	**40**	$A_{358} \rightarrow A_{419}$	**50**	$A_{517} \rightarrow A_{490}$

[Step 9] Defuzzification based on "(5)". Assume that we want to to defuzzify A_1, then based on Table 5, we can see that there is a $A_1 \rightarrow A_{23}$ in group 1. Therefore, the value of A_1 can be calculated referring to principles in **Step 6** FTS:

Known $m_{23} = 758543.64$ so the value of A_1 is 758543.64

Calculate the value of every group in Table 5 then the result of this defuzzifications are intended with the results of the fuzzification of data per year in Table 4.

E.g. Assume we need to predict the data of year 1972 (3rd data), we need to see the fuzzified value of the year before 1972, which is 1971 (2nd data). Based on Table 4, data of year 1971 belongs to A_1 so the prediction of year 1972 is the defuzzification value of A_1 (i.e. 758543.64).

4. Result and Discussion

Automatic clustering and fuzzy logical relationship was tested on 50 data (data from 1971 to 2020). The test results then are compared with the test result of Chen's Fuzzy Time Series as the basic method of automatic clustering and fuzzy logical relationship. The comparison between the error values will show the effect of automatic clustering as an interval maker in fuzzy time series. Table 6 shown the comparison between those two methods followed by the visualisation of it on "Fig.2". In order to calculate the error values, Mean Absolute Percentage Error (MAPE) is used, defined as follows:

$$MAPE = \frac{\sum_{i=1}^{n} \frac{|Actual_i - Predicted_i|}{Actual_i}}{n} \times 100 \qquad (8)$$

Table 6. Predicted Value

Year	Actual Value	Predicted Value - FTS	Predicted Value - ACFLR ($p = 11$)
1970	581190	-	-
1971	580232	809915.52	580275.55
1972	755960	809915.52	758543.64
1973	866638	809915.52	869883.18
1974	794741	809915.52	776629.09
1975	788292	809915.52	774045.46
1976	789892	809915.52	772753.64
1977	781967	809915.52	865680.18
...	...	...	...
2013	2701995	2475121.07	2680501.01
2014	2623471	2589962.83	2623782
2015	2630313	2475121.07	2631421.87
2016	2578791	2475121.07	2580821.91
2017	2991909	2475121.07	2968265.61
2018	2812105	2589962.83	2790611.01
2019	2503671	2589962.83	2482177.01
2020	2280090	2475121.07	2280399.41
MAPE		8,6009%	**0.5316%**

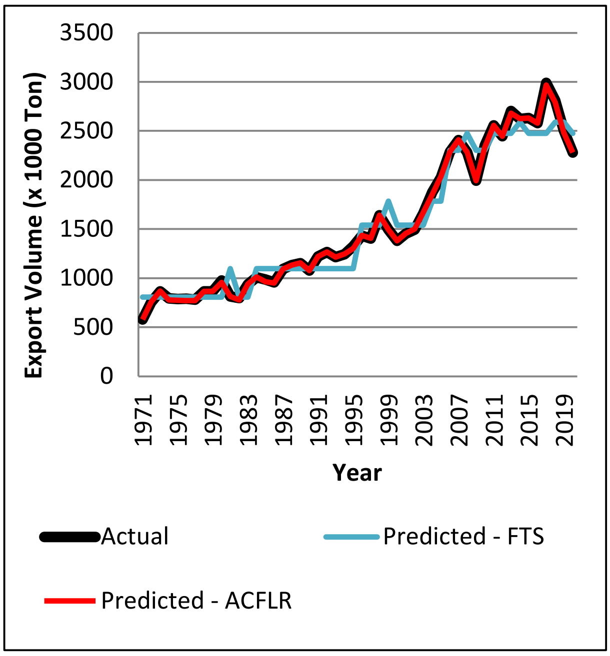

Fig. 2. Comparison between actual and prediction

Based on Table 6, it can be seen that there is a significant difference between the FTS without automatic clustering and the FTS that uses automatic clustering as the interval maker. It can be seen that ACFLR produced smaller MAPE value than the Chen FTS with a large enough difference, 8.0693%. These results prove that ACFLR is better at predicting the volume of Indonesia's natural rubber exports. Based on these results, it is evident that the factors that influence the difference in MAPE values lie in the length of the interval in each method. In the ACFLR, interval formation is carried out from data clusters where data clusters are formed based on the level of proximity between data so that an effective interval is produced. In addition, the interval that has been formed can be further divided into several appropriate sub-intervals so that the fuzzification process can be carried out more optimally. In FTS, the interval formation is only based on the number of classes calculated by equation (2), not measured by the proximity between the data. The number of intervals used in the FTS is 11 intervals, while the ACFLR uses 517 intervals.

In Fig.3 a line graph is presented to visualize the comparison of the data. The black line shows the actual data, the red line shows the predicted data using the ACFLR, and the blue line shows the predicted data using the Chen's FTS method. In Fig.3, the difference between the red line and the blue line is clearly visible, where the red line is the ACFLR's that almost coincides with the black line (actual data). This indicates that ACFLR is better at predicting the volume of Indonesian natural rubber exports due to the influence of the different interval formation processes between the two methods.

5. Conclusion

Based on the tests carried out on 50 data, it can be seen that the Automatic Clustering and Fuzzy Logical Relationship methods are superior to the MAPE value compared to Chen's Fuzzy Time Series method in predicting the volume of Indonesian natural rubber export. The comparison of MAPE values for each method is 0.5316% and 8.6009%. Therefore, it can be concluded that the length of the interval can affect the Fuzzy Time Series method in making predictions so the use of the Automatic Clustering method optimization for interval formation is proven to reduce the MAPE value

References

[1] Kertayuga, D. (2021). *Prediksi Nilai Ekspor Impor Migas dan Non-Migas Indonesia Menggunakan Extreme Learning Machine (ELM). Edy Santoso, S. Si., M. Kom. dan Nurul Hidayat, S. Pd., M. Sc* (Doctoral dissertation, Universitas Brawijaya).

[2] Floranica, P.B., B2A219051 (2020) *Prediksi Nilai Ekspor Migas dan Non-Migas di Jawa Timur dengan Artificial Neural Network Conjugate Gradient Fletcher-Reeves.* Undergraduate thesis, Muhammadiyah University, Semarang.

[3] Lindung, L., & Jamil, A. S. (2018). *Posisi Daya Saing Dan Tingkat Konsentrasi Pasar Ekspor Karet Alam Indonesia Di Pasar Global. Jurnal AGRISEP: Kajian Masalah Sosial Ekonomi Pertanian Dan Agribisnis*, *17*(2), 119-128.

[4] Kementerian Pertanian. 2019. *Statistik Perkebunan Unggulan Nasional 2019 – 2021.* Direktorat Jenderal Perkebunan Kementerian Pertanian

[5] Perdana, R. P. (2020, July). *Kinerja Ekonomi Karet dan Strategi Pengembangan Hilirisasinya di Indonesia*. In Forum penelitian Agro Ekonomi (Vol. 37, No. 1, pp. 25-39).

[6] Kementerian Perdagangan. 2019. Keputusan Menteri Perdagangan Republik Indonesia Nomor 779 Tahun 2019

[7] Atika, S., & Afifuddin, S. (2015). *Analisis Prospek Ekspor Karet Indonesia ke Jepang. Jurnal Ekonomi dan Keuangan*, 3(1), 14835.

[8] Al Mahkya, D. (2016). *Prediksi Nilai Ekspor Jawa Tengah Menggunakan Pendekatan Hierarchical Time Series* (Doctoral dissertation, Institut Teknologi Sepuluh Nopember).

[9] Nugroho, K. (2016). *Model Analisis Prediksi Menggunakan Metode Fuzzy Time Series*. Infokam, 12(1).

[10] Chen, S. M., Wang, N. Y., & Pan, J. S. (2009). *Forecasting Enrollments Using Automatic Clustering Techniques and Fuzzy Logical Relationships*. Expert Systems with Applications, 36(8), 11070–11076. https://doi.org/10.1016/j.eswa.2009.02.085

[11] Huarng, K. (2001). Effective lengths of intervals to improve forecasting in fuzzy time series. *Fuzzy sets and systems*, *123*(3), 387-394.

[12] Wahyudy, H. A. (2018). *Perkembangan Ekspor Karet Alam Indonesia*. Dinamika Pertanian, *34*(2), 87-94.

[13] Kohjiya, S. (2015). *NATURAL RUBBER*. Smithers Rapra.

[14] Junaidi. 2019. *Jenis Tanaman Penghasil Karet dan Produk yang Dihasilkan.* https://penasultra.com

[15] Anonim, 2008. *Panduan Lengkap Karet*. PENEBAR SWADAYA. Bogor

[16] Priyadarshan, P. M. (2011). *BIOLOGY OF HEVEA RUBBER* (PP. 1-6). Wallingford, UK: CABI.

[17] Dean, W. (2002). *Brazil and The Struggle for Rubber*. Department of history New York University.

[18] Dr.M. Subandi, Ir., M. (2011). *Budidaya Tanaman Perkebunan Unggal*. In Jakarta: Penebar Swadaya.

[19] Kementerian Pertanian. (2021). *SEJARAH KARET.* http://museum.pertanian.go.id/berita/sejarah-karet-.9568256

[20] Soleh, A. (2016). *Analisis Ekspor dan Produksi Karet di Indonesia (Aplikasi Model Lag Terdistribusi)*. EKOMBIS REVIEW: Jurnal Ilmiah Ekonomi dan Bisnis, 4(1).

[21] Chen, S. M. (1996). *Forecasting Enrollments Based on Fuzzy Time Series. Fuzzy sets and systems*, 81(3), 311-319.

[22] Hidayatullah, M. A. (2015). *Model Hibrida Arima dan Fuzzy Time Series untuk Meramalkan Data Berpola Trend.*

[23] Fauziah, N., Wahyuningsih, S., & Nasution, Y. N. (2016). *Peramalan Mengunakan Fuzzy Time Series Chen (Studi Kasus: Curah Hujan Kota Samarinda).* Jurnal Statistika Universitas Muhammadiyah Semarang, 4(2).

[24] Sitohang, S. (2018). *Analisis Peramalan Harga Emas dengan Metode Automatic Clustering And Fuzzy Logic Relationship.* Journal Information System Development (ISD), 3(2).

[25] Van Tinh, N. (2016). *A Forecasting Method Based on Combining Automatic Clustering Technique and Fuzzy Relationship Groups*.

[26] Gao, R.; Duru, O. (2020). *Parsimonious Fuzzy Time Series Modelling*. Expert Systems with Applications, 156(), 113447–.

[27] Panigrahi, S., & Behera, H. S. (2020). *FUZZY TIME SERIES FORECASTING: A SURVEY*. Computational Intelligence in Data Mining, 641-651.

[28] Abdullah, L., & Ling, C. Y. (2012). *Intervals in Fuzzy Time Series Model Preliminary Investigation for Composite Index Forecasting.* ARPN Journal of Systems and Software, 2(1), 7-11.

[29] Kamal S. Selim, Gihan A. Elanany. (2013). *"A New Method for Short Multivariate Fuzzy Time Series Based on Genetic Algorithm and Fuzzy Clustering",* Advances in Fuzzy Systems, vol. 2013, Article ID 494239, 10 pages. https://doi.org/10.1155/2013/494239