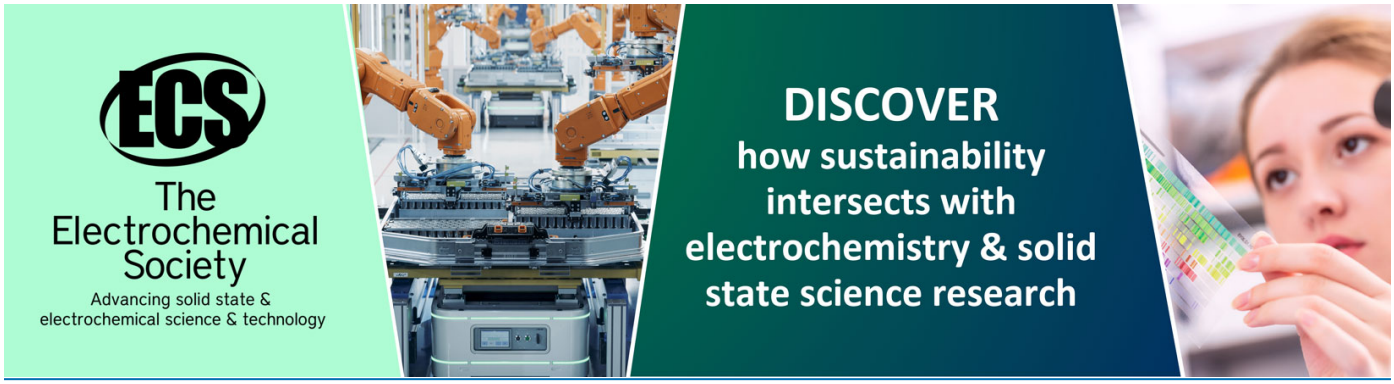
Implementation of Automatic Clustering Algorithm and Fuzzy Time Series in Motorcycle Sales Forecasting



IOP Conf. Series: Materials Science and Engineering **288** (2018) 012126 doi:10.1088/1757-899X/288/1/012126

Implementation of Automatic Clustering Algorithm and Fuzzy Time Series in Motorcycle Sales Forecasting

Rasim*, E Junaeti, and R Wirantika

Computer Science Department, Universitas Pendidikan Indonesia, Bandung, Indonesia

*rasim@upi.edu

Abstract. Accurate forecasting for the sale of a product depends on the forecasting method used. The purpose of this research is to build motorcycle sales forecasting application using Fuzzy Time Series method combined with interval determination using automatic clustering algorithm. Forecasting is done using the sales data of motorcycle sales in the last ten years. Then the error rate of forecasting is measured using Means Percentage Error (MPE) and Means Absolute Percentage Error (MAPE). The results of forecasting in the one-year period obtained in this study are included in good accuracy.

1. Introduction

Profit is one of criterions used to measure the performance of a company [1]. The company's rate profitability is directly proportional to the level of sales [2], if the company profits as might be expected, then the company can expand its business. One of the fundamental problems that often occur in making sales plans difficult to realization is the mistake in making the right sales prediction [3]. If the sales prediction is too large (Over Production) then the cost of production will swell and all asset which is invested into less efficient. Conversly, if the sales prediction is made too small (Under Production) then the company will face out of stock (Stock Out), so the customer was forced to wait too long for the desired product [4]. Then it requires an activity to estimate the amount of sales of goods and services by the manufacturer and distributor in a certain time period i.e. sales forecasting. Right forecasting has several important criteria, including accuracy, cost and convenience. The selection of appropriate forecasting methods is a matter of concern for accurate forecasting [5].

Sales forecasting will be influenced by previous sales data, then a suitable forecasting method used in sales forecasting is the time series method included the fuzzy time series method. This method is a fairly new method compared to other classical time series forecasting methods, such as Auto Regresive (AR), Moving Average (MA), ARMA, ARIMA, and so on. This method was used to forecast student enrollment at the University of Alabama by Shyi Ming Chen and Chia-Ching Hsu in 2004 [6]. The results of the comparison of the fuzzy time series method proposed by them can get a higher forecasting accuracy rate for forecasting enrollments than the existing methods [7]. A new method to forecast enrollments using fuzzy time series and clustering techniques [8]. The experimental results show that the proposed method gets a higher average forecasting accuracy rate than the existing methods. Furthermore, the process of this method does not require a learning system from a complex system, as it does in genetic algorithms and neural networks so the method is easy to develop [9].

In this research, the forecasting application of national motorcycle sales using method of high order fuzzy time series combined with interval determination using automatic clustering algorithm was

presented. The historical data of the motorcycle sales in Indonesia were used to illustrate the forecasting process of the proposed method. The experimental results show that the proposed method is included in good accuracy.

2. Methods

The research method is shown in the figure 1.

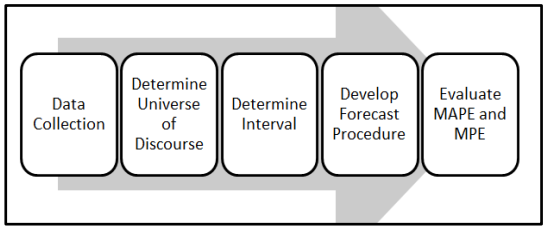


Figure 1. Research methods.

Based on the figure 1 in a simple forecasting method implemented in this study consists of five (5) stages as follows:

1. Collect monthly data on motorcycle sales in period of ten years.
2. Determine the universe of discourse of data sales for every month.
3. Determine interval data using automatics clustering algorithm.
4. Develop forecast procedure using high order fuzzy time series method without the process of determining interval.
5. Evaluated forecast result by calculating MPE and MAPE value.

3. Results and discussion

3.1. Historical data

The data used in this study are monthly data of the number of motorcycle sales of brand Honda, Suzuki, Yamaha, Kawasaki, and others brand in Indonesia. The data used are the monthly period for 10 years in period 2005 to 2014 so there are 120 periods for every band. Data of Suzuki motorcycle sales in January period 2005 to 2014 is illustrated in Table 1.

Data in January period 2005 to 2014 would be used to forecast motorcycle sales in January 2015. In addition, forecasting of motorcycle sales of others month would be forecast with the similar method.

Table 1. Historical Data of Suzuki motorcicle sales

No	Year	Data of Sales	No	Year	Data of Sales
1	2005	83.846	6	2010	40.225
2	2006	39.469	7	2011	47.614
3	2007	52.309	8	2012	53.337
4	2008	73.043	9	2013	35.758
5	2009	22.369	10	2014	30.012

3.2. Forecasting procedure

The following procedure is an example of forecasting calculation to predict motorcycle sales brand Suzuki in January 2015. Data in Table 1 was used as historical data.

1. Determine interval data using automatics clustering algorithm.
   a. Sorted the numerical data in ascending order and calculated average diff. Data in ascending order is shown in Table 2, and the average diff is 6.830,78.

Table 2. Historical Data of Suzuki Motorcicle Sales

No	Data	Order	No	Data	Order
1	22.369	d_1	6	47.614	d_6
2	30.012	d_2	7	52.309	d_7
3	35.758	d_3	8	53.337	d_8
4	39.469	d_4	9	73.043	d_9
5	40.225	d_5	10	83.846	d_{10}

b. Change the data into cluster using average diff value in [9]. The clusters are {22.369}, {30.012; 35.758; 39.469; 40.225}, {47.614; 52.309; 53.337}, {73.043}, and {83.846}.
c. Refine the clusters into: {22.369; 29.199,78}, {30.012; 40.225}, {47.614; 53.337}, {66.212,22; 79.873,78}, and {77.015,22; 83.846}.
d. Change the clusters into intervals: [22.369;29.199,78], [29.199,78;30.012], [30.012;40.225], [40.225;47.614], [47.614;53.337], [53.337;66.212,22], [66.212,22;79.874], [79.874;83.846].

2. Develop forecast procedure using high order fuzzy time series method without the process of determining interval.
   a. Determine universal of discourse and intervals using automatics clustering algorithm, which resulted intervals: [22.369; 29.199,78], [29.199,78;30.012] [30.012; 40.225], [40.225; 47.614], [47.614; 53.337], [53.337; 66.212,22], [66.212,22; 79.873,78], and [79.873,78; 83.846].
   b. Get a statistics of the distribution of the historical data in each interval to re-divided data into the new intervals. The new intervals are: [22.369; 25.784,39], [25.784,39; 29.199,78], [29.199,78; 29.605,89], [29.605,89 ; 30.012], [30.012; 32.565,25], [32.565,25; 35.118,50], [35.118,50; 37.671,75], [37.671,75 ; 40.255], [40.255 ; 43.919,50], [43.919,50 ; 47.614], [47.614 ; 49.521,67], [49.521,67 ; 51.429,33], [51.429,33 ; 53.337], [53.337 ; 57.628,74], [57.628,74 ; 61.920,48], [61.920,48 ; 66.212,22], [66.212,22 ; 73.043], [73.043 ; 79.873,78], [79.873,78 ; 81.859,89], [81.859,89 ; 83.846].
   c. Establish fuzzy logical relationships, which were:

$$A_1 = {}^{1}/_{u_1}, {}^{0,5}/_{u_2}, {}^{0}/_{u_3}, {}^{0}/_{u_4}, \dots, \dots, {}^{0}/_{u_{20}}$$
$$A_2 = {}^{0,5}/_{u_1}, {}^{1}/_{u_2}, {}^{0,5}/_{u_3}, {}^{0}/_{u_4}, \dots, \dots, {}^{0}/_{u_{20}}$$
$$A_3 = {}^{0}/_{u_1}, {}^{0,5}/_{u_2}, {}^{1}/_{u_3}, {}^{0,5}/_{u_4}, \dots, \dots, {}^{0}/_{u_{20}}$$
$$A_4 = {}^{0}/_{u_1}, {}^{0}/_{u_2}, {}^{0,5}/_{u_3}, {}^{1}/_{u_4}, \dots, \dots, {}^{0}/_{u_{20}}$$
$$A_{\dots} = {}^{0}/_{u_1}, {}^{0}/_{u_2}, {}^{0}/_{u_3}, \dots, \dots, {}^{1}/_{u_{19}}, {}^{0,5}/_{u_{20}}$$
$$A_{20} = {}^{0}/_{u_1}, {}^{0}/_{u_2}, {}^{0}/_{u_3}, \dots, \dots, {}^{0,5}/_{u_{19}}, {}^{1}/_{u_{20}}$$

   d. Get forecasting of year n+1, using modus trend value in the last 10 years, which was shown in figure 2 with comparison between actual data and forecasting result.

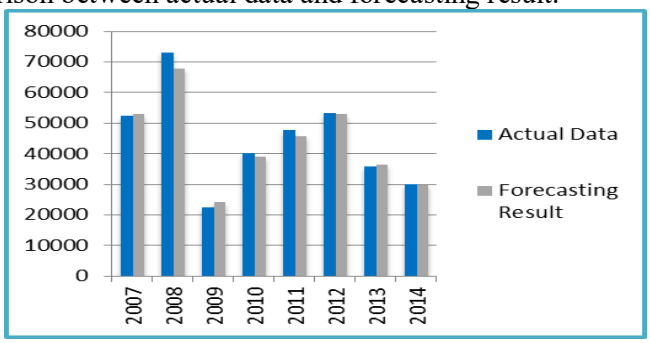


Figure 2. Comparison of data between actual data and forecasting result.

3. Evaluated forecast result by calculating MPE and MAPE value.

Using (1) and (2), MPE and MAPE was calculated, which results are MPE is 0,65% and MAPE is 3,26%. Based on criteria of MAPE, the forecasting results shown are included in good accuracy.

3.3. Results and discussion

Figure 3 showed the forecasting result of motorcycle sales for every brand. The total forecasting of sales are 7,820,612 units, with an average sales are 651,718 units per month. The highest sales for 2015 occurred in June of 743,696 units and the lowest sales occurred in July 530,929 units.

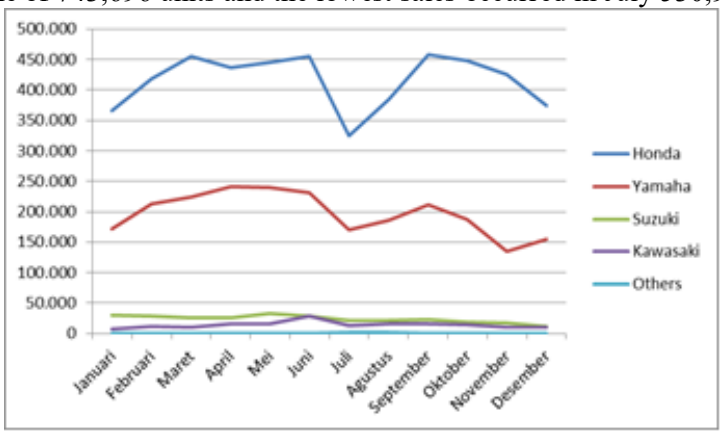


Figure 3. Forecasting result of motorcycle sales.

To evaluate the rate of accuracy of this forecasting methods, then motorcycle sales was forecasted in one year period. Afterward average of MAPE was calculated. The result show that the average value of MPE is 0.19% and MAPE is 2.15% which are included in good accuracy.

4. Conclussion

The fuzzy time series forecasting method using automatics clustering algorithm give good result in system forecasting of motorcycle sales in Indonesia. The accuracy of this method was better than the fuzzy time series forecasting method without automatics clustering algorithm.

References

[1] Storey D J, Keasey K, Watson R & Wynarczyk P 2016 *The performance of small firms: profits, jobs and failures* Routledge

[2] S I Mezhov, Y I Rastova & A D Shmatko 2014 Theoretical and methodological bases of forming the operational and innovation program of the company *Advances in Natural and Applied Sciences* **8** 5 451-458

[3] M McDonald & H Wilson 2016 Marketing Plans: How to prepare them, how to profit from them. John Wiley & Sons

[4] J Pan, S Liu, S Tuck, & A Alkuraiji, A Framework for Optimising Inventory Level of Global Critical Knowledge to Support Group Decision Making *In Joint International Conference on Group Decision and Negotiation* 81-89 Springer Cham

[5] Y Bai, P Wang, C Li, J Xie & Y Wang 2014 A multi-scale relevance vector regression approach for daily urban water demand forecasting *Journal of Hydrology* **517** 236-245

[6] CNN Indonesia, Aisi Kibarkan Bendera Putih Capai Target Penjualan Motor [Online] http://www.cnnindonesia.com/ekonomi/20150630074804-92-63169/aisi-kibarkan-bendera-putih-capai-target-penjualan-motor/. Accesed on 10 November 2016. 2015.

[7] S M Chen and C C Hsu 2014 A New Method to Forecast Enrollments Using Fuzzy Time Series *International Journal of Applied Science and Engineering* **2** 3 234-244

[8] K Tanuwijaya & S M Chen 2009 A new method to forecast enrollments using fuzzy time series and clustering techniques *In Machine Learning and Cybernetics, 2009 International Conference on* **5** 3026-3029 IEEE

[9] H T Jasim, A G Salim & K I Ibraheem 2014 A novel algorithm to forecast enrollment based on fuzzy time series *Applications and Applied Mathematics: An International Journal* **7** 1 385-397