PAPER • OPEN ACCESS

Sparepart sales clusterization and prediction using automatic clustering algorithm

To cite this article: Helmy Tirta *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **1007** 012191

IOP Conf. Series: Materials Science and Engineering 1007 (2020) 012191 doi:10.1088/1757-899X/1007/1/012191

Sparepart sales clusterization and prediction using automatic clustering algorithm

Helmy Tirta*, Novario Jaya Perdana and Bagus Mulyawan**

Informatics Engineering Department, Faculty of Information Technology, Universitas Tarumanagara, Indonesia

*helmy.535150074@stu.untar.ac.id
**bagus@untar.ac.id

Abstract. Creating an information system for a store becomes an important part in a business process as an effort to facilitate the running of the business process. Therefore, the existence of this information system will make it easier for the store to get information on products, recording the number of products entering and leaving, supplier information and can predict the number of sales in the coming month. Prediction of sales in the coming month is very important to facilitate the store in preparing stock of product in the coming month. The design of this program applies the automatic clustering method and fuzzy logical relationship that manages the historical data of the number of sales each month in the previous period, so we get a prediction of the number of sales in the coming month, which is useful for the store in preparing the amount of stock in the coming months. Based on the results of tests conducted, this program can provide predictions of the number of sales in the coming months. Tests are carried out on two product samples, where the products tested have an accuracy level of 86.20% and 90%, respectively.

1. Background

At present use of computer technology as a tool to support activities in the business sector makes it easy for humans to get information quickly, precisely, accurately so that the objectives of such work can be achieved effectively and efficiently. While business people have needs to utilize their "data warehouse", researchers see the opportunity to give birth to a new technology that responds to this need, namely data mining.

Mitra Jaya Harapan Indah store is a business engaged in the sale of engineering equipment spare parts. At the moment, Mitra Jaya Jaya Indah store is still running its business process manually, where the recording of information on the list of products, recording the number of product entering and leaving, information on suppliers and planning of supply of products is still done manually. To be able to overcome the problems that occur, the Mitra Jaya Harapan Indah store requires an information system that can facilitate the Mitra Jaya Harapan Indah store in finding information on each item, recording the number of products entering and leaving, information about the list of suppliers and predictions of sales in the coming month.

Prediction of sales in the future is very important to facilitate the store in preparing stock of products in the months to come, thus avoiding the availability of too much stock because it will result in high expenditure costs, otherwise a small stock availability will result in the store being unable to meet demand consumers, so that consumers do not feel satisfied.

Information systems are created using automatic clustering algorithms and fuzzy logical relationships. Automatic clustering algorithm is an algorithm of grouping a number of data into certain data groups (clusters). From the grouping of data, we will get a number of intervals. The interval will be used in the application of fuzzy logical relationships to predict the number of sales. So from the prediction of the sales amount, Mitra Jaya Harapan Indah store can prepare an inventory of the stock of products in the coming months.

2. Literatur Study

At present use of computer technology as a tool to support activities in the business sector makes it easy for humans to get information quickly, precisely, accurately so that the objectives of such work can be achieved effectively and efficiently. For achieving these objectives, the business need to be supported by reliable technology [16].While business actors have the needs to utilize the "data warehouse" they already have. From the information from the prediction of the number of sales, it can be used as a design for the future supply of goods. Forecasting is common for many people to know events that will occur in the future by looking at activities that have occurred before.[6]

There are several forecasting methods that use a quantitative approach, one of which is fuzzy logic 14] Fuzzy time series method was first applied by Song and Chissom[13] to solve forecasting problems by looking at previous data to form a fuzzy time series model. Song and Chissom's model uses Anggodo & Mahmudy, Forecasting Minimum Living Needs ... 95 min-max operations to predict the number of registrants at the University of Alabama.[13]In addition there are other fuzzy models, namely Chen et al's fuzzy time series model which is simpler also applied to predict the number of registrants at the University of Alabama.[5] In the previous year, several studies focusing on fuzzy time series for solving forecasting problems include [8]; [1]; [5]; [11]; [7]; [15]; [10]; [6].

From the research that has been done automatic clustering method is quite effective and very helpful in classifying data with a combination of various methods for its completion. The use of automatic clustering gets significantly more accuracy than without the use of data classification. Fuzzy logical relationship high-order forecasting methods solve TAIEX problems[2], whereas in Chen and Chen's research in 2015 on the same problem the fuzzy logical relationship method is optimized with the use of second-order and trend fuzzy probabilities in fuzzy logical relationship.[3] The use of high-order, second-order and fuzzy-trend in fuzzy logical relationships get accuracy results that are not too significant. Forecasting is also done to predict the number of University of Alabama students using automatic clustering and fuzzy logical relationships which generalize to get lower errors. [10]

In the same case the research of Cheng et al using fuzzy logical relationships developed with several other methods get lower errors than previous research.[6] From studies it can be concluded that fuzzy logical relationships can solve forecasting problems with various combinations of methods or stand alone. In this study forecasting will be done using automatic clustering and fuzzy logical relationships to predict the number of spare parts sales in the coming months. The first stage of the automatic clustering method will be used to classify historical data on the number of spare parts sales in the previous months. Forecasting is done by the fuzzy logical relationship method. This method is proven effective with better accuracy than the other methods. [10]; [6]

3. Method

The method used in this design are automatic clustering algorithm and fuzzy logical relationship methods. Automatic clustering algorithm is an algorithm for grouping data into certain data groups (clusters). In this design the data used is historical data on the number of spareparts sales in the previous months. From grouping data with the automatic clustering algorithm, interval values will be obtained. This interval will be used in the application of fuzzy logical relationships to predict the number of sales in the coming months. Automatic Clustering is used to classify data based on intervals. Convert data into numeric data groups, then numeric data groups are converted into intervals. Intervals that have different interval lengths at each interval. The steps of the Automatic Clustering algorithm include:

Step 1

Sort the data in ascending order and specify the average_diff value. Having n different numeric data, from the smallest datum to the largest datum without duplicate data (data with the same value). If there is duplicate data in data sorting, then retrieve one data from the duplicate data. For example, data that has been sorted without duplicate data is described as d1, d2, d3, ..., in, ..., dn. then calculate the value "average_diff" using equation (1).

$$average\_diff = \frac{\sum_{i=1}^{n-1}(\mathrm{d}i+1-di)}{n-1} \quad (1)$$

Notes :
d_{i+1}= next data
d_i= current data
n = amount of data
Where "average_diff" is the average value difference between each pair of ascending data sequences.

Step 2

Change data into cluster (group) form. Enter the smallest data that has been sorted without duplicate data into the first cluster. Based on the value of "average_dif", it is determined whether the numbers in the ascending data sequence are included in the current cluster or a new cluster is made up of data numbers with the following rules:

Rule 1: Assume that the current cluster is the first cluster and there is only one data d1 in it, and d2 is data whose value is greater and close to d1 which is described as follows: {d1}, d2, d3, ..., dn. If d2– d1≤ average_dif, then enter d2 inside.the current cluster where d1 is included, conversely form a new cluster for d2 and leave the new cluster formed where d2 is included as the current cluster.

Rule 2: Assume the current cluster is not the first cluster, and there is only one data that is the dj data in the current cluster. Assume dk is data whose value is larger and is close to dj data and di is the largest data in the cluster that existed before the current cluster, described as {d1}, ..., {..., at}, {dj}, dk, ..., dn . If dk– dj ≤ average_dif and dk-dj ≤ dj– di, then insert dk into the current cluster which is dj inside. Instead, form a new cluster for dk and let a new cluster form where dk is included as the current cluster.

Rule 3: Assume the current cluster is not the first cluster and there is more than one data cluster at this time. Assume di is the largest data in the current cluster and dj is the data whose value is greater and adjacent to di is described as follows: {d1}, ..., {...}, {..., di}, dj, ..., dn. If dj - in ≤ average_dif and dj - in ≤ cluster_dif, then insert dj into the current cluster that is a member of it, otherwise a new cluster with dj members is made and make the new cluster the current cluster. Where cluster-dif shows the difference in the average distance between each pair of data adjacent to the cluster. Cluster_dif calculation shown in equation (2)

$$cluster\_dif = \frac{\sum_{i=1}^{n-1}(ci + 1 - ci)}{n-1} \quad (2)$$

Notes :
c_{i+1}= next data
c_i= current data
n = amount of data
Where cluster_dif is the average of the data difference between the current cluster members that are close together and c1, c2, ..., cn are the data that are members of the current cluster

Step 3

Improve the contents of clusters (groups). Update the members in each cluster obtained from Step 2 based on the following three rules:

Rule 1: If the cluster has more than two data members, then take the smallest data and the largest data, then delete the other data in the cluster.

Rule 2: If the cluster has two data members, then maintain both.

Rule 3: If the cluster has one dq data member, then enter the values "dq - average_dif" and "dq + average_dif" into the cluster, and delete the dq data from the cluster. But it also has to adjust to the following situation:

Situation 1: if the first cluster, then delete "dq - average_dif" and maintain dq.

Situation 2: if the cluster is last, then delete "dq + average_dif" and keep dq.

Situation 3: if "dq - average_dif" is smaller than the smallest data value than the smallest value in the previous cluster, then rule 3 does not apply, so the cluster member remains dq.

Step 4
Change the cluster to an interval. Assume that the clustering results obtained from Step 3 are as follows: {d1, d2}, {d3, d4}, ..., {di, dj}, {dk, dl}, ... {dr}, {ds, dt}, ..., {dn-1, dn}. Change the results of the cluster into adjacent intervals into the following sub steps:

Sub Step 4.1 change the first cluster {d1, d2} to interval [d1, d2).

Sub Step 4.2 if the current interval [in, dj) and the current cluster are {dk, dl}, then:

1) If dj≥ dk, then change the current cluster {dk, dl} to the interval [dj, dl). Let [dj, dl) be the current interval and let the next cluster {dm, dn} be the current cluster.

2) If dj <dk, then change {dk, dl} into the interval [dk, dl) and create a new interval [dj, dk) between the intervals [di, dj) and [dk, dl). Now [dk, dl) becomes the current interval and let the next cluster {dm, dn} become the current cluster.

Sub Step 4.3 Repeat Sub step 4.1 and step 4.2 until all clusters become intervals.

4. Result & Discusion

Based on table 1 illustrates the comparison between the number of actual sales and the number of sales predictions per month with the automatic clustering algorithm method and fuzzy logical relationship on nankai rompi jaring hijau sparepart product.

Tabel 1 Final result for prediction sales

Product	NANKAI Rompi Jaring Hijau					
Year / Month	2018	Prediction	2019	Prediction	2020	Prediction
January	434		745	747	665	604
February	465	569	534	538	754	661
March	351	353	590	598	611	614
April	357	363	523	526	452	452
May	458	460	397	405	455	452
June	445	440	468	482		452
July	665	604	549	553		
August	563	627	633	637		
September	782	790	723	729		
October	682	684	719	720		
November	623	626	610	611		
December	556	560	689	697		

Based on table 2 is a comparison between the actual number of sales and the predicted number of sales using the MAPE method resulting an average error of 3.01%, which means the prediction of the number of sales using the automatic clustering method and fuzzy logical relationship has a high level of accuracy.

Tabel 2 Final result for calculation Error

				Nankai Rompi Jaring Hijau					
m y	2018	Calculation error	Result	2019	Calculation Error	Result	2020	Calculation Error	Result
1	434			745	(|745-747)/745|)*100	0,26%	665	(|665-604)/665|)*100	9,17%
2	465	(|465-569)/465|)*100	22,30%	534	(|534-538)/534|)*100	0,74%	754	(|754-661)/754|)*100	12,30%
3	351	(|351-353)/351|)*100	0,56%	590	(|590-598)/590|)*100	1,35%	611	(|611-614)/611|)*100	0,49%
4	357	(|357-363)/357|)*100	1,68%	523	(|523-526)/523|)*100	0,57%	452	(|452-452)/452|)*100	0,00%
5	458	(|458-460)/458|)*100	0,43%	397	(|397-405)/397|)*100	2,01%	455	(|455-452)/455|)*100	0,65%
6	445	(|445-440)/445|)*100	1,12%	468	(|468-482)/468|)*100	4,27%		Average Error	3,01%
7	665	(|665-604)/665|)*100	9,17%	549	(|549-553)/549|)*100	0,72%			
8	563	(|563-627)/563|)*100	11,36%	633	(|633-637)/633|)*100	0,63%			
9	782	(|782-790)/782|)*100	1,02%	723	(|723-729)/723|)*100	0,82%			
10	682	(|682-684)/682|)*100	0,29%	719	(|719-720)/719|)*100	0,13%			
11	623	(|623-626)/623|)*100	0,48%	610	(|610-611)/610|)*100	0,16%			
12	556	(|556-560)/556|)*100	0,71%	689	(|689-697)/689|)*100	1,16%			

Based on table 3 is a comparison between manual calculations with program output resulting in an accuracy rate of 86.20% which means that the information system designed has a good level of accuracy

Tabel 3 Final Result of comparison between manual calculation and output program

					NANKAI Rompi Jaring Hijau							
y m	2018	Output Manual	Output Program	Result	2019	Output Manual	Output Program	Result	2020	Output Manual	Output Program	Result
1	434				745	747	747	1	665	604	604	1
2	465	569	569	1	534	538	538	1	754	661	661	1
3	351	353	353	1	590	598	598	1	611	614	614	1
4	357	363	363	1	523	526	526	1	452	452	452	1
5	458	460	455	0	397	405	405	1	455	452	452	1
6	445	440	440	1	468	482	482	1		452	452	1
7	665	604	604	1	549	553	553	1		Average		86,20%
8	563	627	627	1	633	637	637	1				
9	782	790	778	0	723	729	729	1				
10	682	684	684	1	719	720	720	1				
11	623	626	626	1	610	611	610	0				
12	556	560	556	0	689	697	697	1				

5. Conclusion

The Conclusion are :

1. In the test carried out with a sample of 2 spare parts products Nankai rompi jaring hijau and maktec drill MT-60 by comparing between manual calculations and program output produces an accuracy rate of 86.20% and 90%

2. The automatic clustering method and fuzzy logical relationship can be applied as a method that can provide good sales predictions with a high degree of accuracy by comparing the actual sales amount and the predicted number of sales with MAPE method resulting in an error rate of only 3.01%.
3. Programs that have been created can run dynamically, despite the addition of new products and new sales data.

6. References

[1] Aznarte, J.L., Fdez, J.A. Azofra, A.A. dan Benitez, J. M., 2012. Financial time series forecasting with a bio-inspired fuzzy model. Expert Systems with Applications, **39**, 12302-12309.

[2] Chen, S.M. dan Chen, C.D., 2011. Handling forecasting problems based on high-order fuzzy logical relationships. Expert Systems with Applications, **38**, 3856-3864.

[3] Chen, S.M. dan Chen, S.W., 2015. Fuzzy forecasting based on two-factors secondorder fuzzy-trend logical relationship group and the probabilites of tren of fuzzy logical relatonships. IEEE Transactions on Cybernetics, Vol **45**, No. 3.

[4] Chen SM, Wang NY, Pan J R. Forecasting Enrollments Using Automatic Clustering Techniques and Fuzzy Logical Relationships. Upper Sddie River: Prentice Hall, 2009.

[5] Chen, S.M. dan Tunawijaya, K., 2011. Multivariate fuzzy forecasting based on fuzzy time series and automatic clustering technique. Expert System with Applications, **38**, 10594-10605.

[6] Cheng SH, Chen SM, Jian W S. Fuzzy Time Series Forecasting Based On Fuzzy Logical Relationships and similarity measures. Information Sciences, 2016.

[7] Hung, W.H. dan Kang, S.C.J, 2014. Automatic clustering method for real-time contruction simulation. Advanced Engineering Informatic, **28**, 138-152.

[8] Liu, T.K., Chen, Y.P. dan Chou, J.H., 2011. Extracting fuzzy relations in fuzzy time series model based on approximation concepts. Expert Systems with Applications, 2011, 11624-11629.

[9] Qiu, W., Liu, X. dan Li, H., 2011. A generalized method for forecasting based on fuzzy time series. Expert Systems with Applications, **38**, 10446-10453

[10] Qiu, W., Zhang, P. dan Wang, Y., 2015. Fuzzy time series forecasting model based on automatic clustering techniques and generalized fuzzy logical relationship. Hindawi Publishing Corporation Mathematical Problems in Engineering, 962597.

[11] Saha, S. dan Bandyopadhayay, S., 2013. A generalized automatic clustering algorithm in a multiobjective framework. Applied Soft Computing, **13**, 89-108.

[12] Satzinger, JW.Jackson RB.,& Burd SD. System Analysis And Design. Los Angeles : Course Tehcnology, 2010.

[13] Song, Q. dan Chissom, B.S., 1994. Forecasting enrollments with fuzzy time series – part II. Fuzzy Sets and Systems, Vol. **62**, Issue 1

[14] Wang G.P., Chen S.Y., Yang X. dan Liu, Jun., 2014. Modeling and analyzing of conformity behavior : A fuzzy logic approach. Optik, **125**, 6594-6598.

[15] Wang, W. dan Liu, X., 2015. Fuzzy forecasting based on automatic clustering and axiomatic fuzzy set classification. Information Sciences, **294**, 78-94.

[16] Mulyawan, B., Kosala, R., Ranti, B., Supangkat, S. H., 2020, *IOP Conference Series: Materials Science and Engineering* **852**-1-012178