An approach for passengers forecasting using Fuzzy Time Series

U Syaripudin[1], W B Zulfikar[2,*], W Uriawan[3] and R A Nugraha[2]

[1] Department of ICT, Asia E-University, Malaysia
[2] Department of Informatics, UIN Sunan Gunung Djati Bandung, Indonesia
[3] LIRIS INSA de Lyon, France

*wildan.b@uinsgd.ac.id

Abstract. At certain times the number of passengers is very difficult to predict. There was a time when the number of passengers occurred a very significant surge. However, there are times when the number of passengers is drastically reduced. This is caused by many factors including time including the holiday season, holidays, and so on. As for other factors such as natural disasters, endemic diseases, and others. The number of passengers must be proportional to the number of vehicles that have been prepared. The company must consider the condition of the vehicle and also the physical condition of the driver. The aim of this study is to conduct passenger forecasting on a coming day. The methodology used is Fuzzy Time Series. The result of the experiment shows that this model has the accuracy of the difference between predictions with real data using PE which is equal to 34.6%.

1. Introduction

Nowadays, transportation is one of the main needs. A company engaged in public transportation can at least dispatch hundreds to thousands of passengers every day. Transportation is the transfer of people or goods using vehicles driven by humans or machines. Transportation is used to facilitate humans in carrying out daily activities [1–4]. Good transportation will play an important role in the development of the region, especially in accessibility, while what is meant by accessibility is the ease and ability of an area or space to be accessed or reached by parties from outside the area, either directly or indirectly [5–9]. When determining travel schedules transportation companies set each route in a day so that in one day trips can occur without passengers or even have to add to the trip other than that the number of cars and drivers that are not appropriate i.e. 10 cars and 17 drivers found it difficult to determine the driver's holiday and car use.

In previous works, this method mainly used in order to assess multiple case especially time series [10–14]. In fuzzy time series, the length of the interval has been determined at the beginning of the process [15–20]. In this process the length of the interval is very influential for the predicted results [21]. The method for determining an effective interval length is the average-based method. Then to calculate the accuracy of the prediction of the original data using PE.

The purpose of this study is to minimize losses incurred by the company with a prediction model of the number of passengers. This work implements fuzzy time series which predicts the number of passengers in the coming day based on previous travel data.

2. Research methods

An Indonesian transport company departs at least 6 departures per day with a capacity of 7 seats and 2 for packages with an average per day being able to carry around 3 passengers per trip based on travel data in July 2019. If the company is able to predict the number of passengers, the results will be different. Because with the prediction of the number of passengers the company can set the schedule in accordance with the prediction of the number of passengers so that the expected average can be increased. In addition, the company can also arrange driver holidays and additional fleets if needed.

In producing good information, good data is also needed. In the implementation of data mining to predict the graduation level of these students can be classified into 2 parts, namely: internal data and data extraction. Internal data is data that comes from within the agency or company. Internal data obtained from travel data owned by the company. The data extraction process will produce this data mining application database. Data extraction includes training data and testing data and will be managed in a DBMS.

In this study, we collected travel data from October 2018 to July 2019 with a total data of 1835 travel data in which contained date and number of passengers. Then the data will be processed using Fuzzy Time Series. Fig 1 describes that on several weeks the number of passengers significantly reduced and the other one expand significantly.

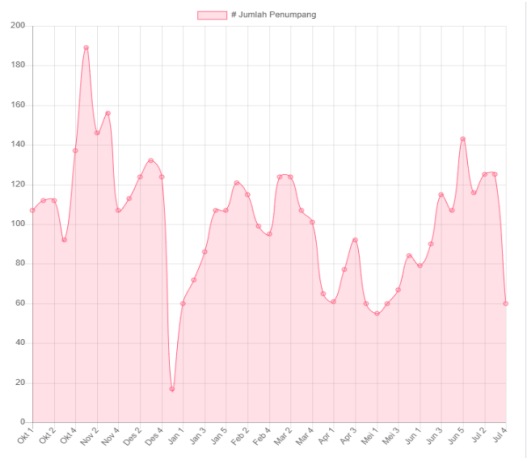

Figure 1. Number of passengers per week.

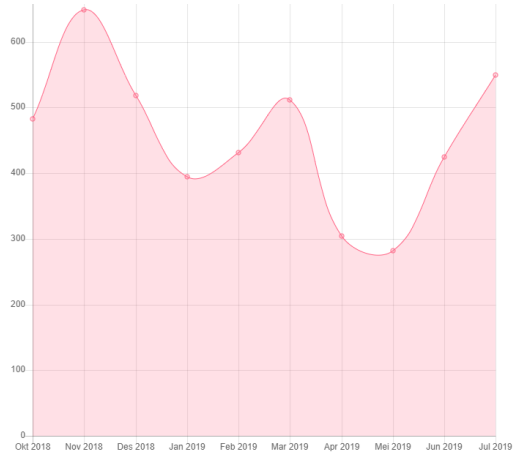

Figure 2. Number of passengers per month.

Fig. 1 and Fig. 2 explain that the number of passengers varies from month to month. The highest number of passengers occurred in November 2018 with a number of 649 passengers or an average of 21.63 passengers a day and the lowest occurred in May 2019 with 282 passengers or an average of 9.4 passengers a day. The biggest increase occurred in October to November 2018 in the amount of 149 passengers while the biggest decrease occurred in March 2019 to April 2019 in the amount of 208 passengers.

In general, the data obtained, whether from an agency or company database have different and imperfect entries, such as missing data, invalid data, or just typos. In addition, there are also data attributes that are not relevant to the data we want. Irrelevant data is better discarded because its existence can reduce the quality of data mining accuracy by completing deficiencies (missing values), eliminating confusion, and correcting data that is not appropriate.

3. Results and discussion

There are several steps for the passenger forecasting model. This work using fuzzy times series to determine the next day number of the passenger as shown in figure 3.

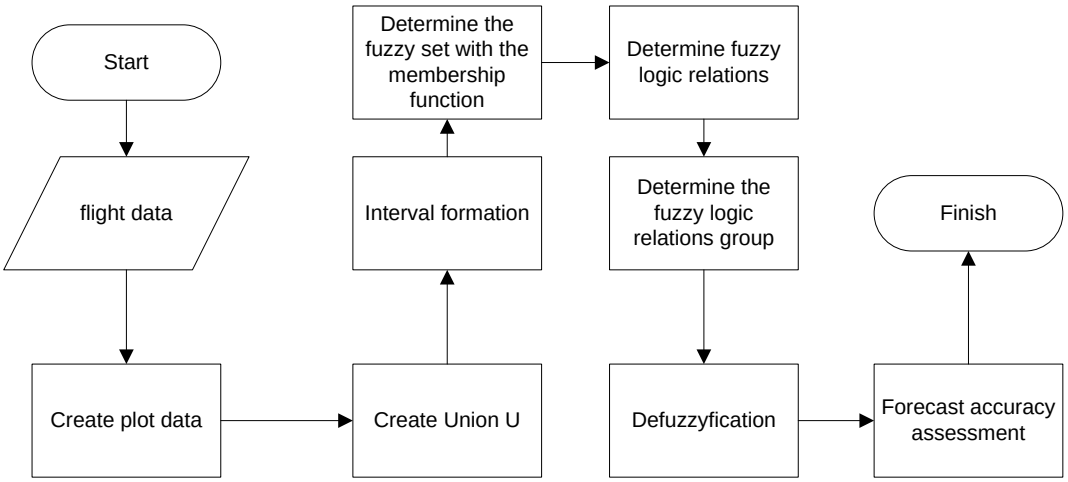

Figure 3. The flow of the proposed model.

The following are the stages of forecasting in time series data using fuzzy times series for example to determine the prediction of August 1 and August 2:

- Defining a universal set of (U) data using (1) where d_{max} is the biggest value and d_{min} is the smallest value.

$$U = [d_{min}, d_{max}] \tag{1}$$

- Determination of the width of the interval using frequency distribution, with the following steps. First, determine the range (range) to use (2):

$$R = d_{max} - d_{min} \tag{2}$$

$R = 43 - 3$
$R = 41$

- Next, determine the number of class intervals using the Sturges Equation as in formula (3).

$$K = 1 + 3.322\, x \log n \tag{3}$$

$K = 1 + 3.322\, x \log 304$
$K = 9$

- Determines interval width according to formula (4):

$$I = \frac{R}{K} \tag{4}$$

$I = \frac{41}{9} = 4.555 = 5$

- The next step is to determine the median using (5) where *i* is the number of fuzzy sets

$$m_i = \frac{(lower\ value + upper\ value)}{2} \tag{5}$$

- The next step is fuzzification by using the lower limit and based on the number of intervals formed as many as 9 then A1, A2, A3, A4, A5, A6, A7, A8, A9 as presented on table 1.

Table 1. List of fuzzification.

	Lower	Upper	Median	Fuzzification
1	3	8	5,5	A1
2	8	13	10,5	A2
3	13	18	15,5	A3
4	18	23	20,5	A4
5	23	28	25,5	A5
6	28	33	30,5	A6
7	33	38	35,5	A7
8	38	43	40,5	A8
9	43	48	45,5	A9

- The Fuzzification stage is based on the number of intervals formed. The results of fuzzification are notated into linguistic numbers are presented in table 2.

Table 2. Linguistic value.

Fuzzification	Linguistic Value
A1	Very reduced
A2	reduced
A3	Quite reduced
A4	Slightly reduced
A5	Constant
A6	A little increase
A7	Enough to increase
A8	increase
A9	Greatly increased

- The defuzzification process has two stages. First, look for the middle value at each interval. Second, calculate the forecast value. The method of calculation is, for example, A1, A2 so that A1 the middle value of U1 and A2 is the result of the middle value at U2, then both are added together and divided by the number of relations. Thus obtained defuzzification results from FLRG in table 3:

Table 3. Defuzzification process and forecasting.

Class	The Median	Current State	Next State	Total Current State	Predicted
1	5,5	A1	A1, A2, A3, A4	31	13
2	10,5	A2	A1, A2, A3, A4, A5, A7	79	18,83
3	15,5	A3	A1, A2, A3, A4, A5, A6	94	18
4	20,5	A4	A1, A2, A3, A4, A5, A6	65	18
5	25,5	A5	A2, A3, A4, A5, A6, A7	27	23
6	30,5	A6	A3, A4, A5	5	20,5
7	35,5	A7	A9, A1	2	25,5
8	40,5	A8	-	0	-
9	45,5	A9	A5	1	25,5

Table 3 uses the formula for the number of next state values divided by the number of next. And to find a prediction value for example on August 2, it can be seen from the number of passengers on the previous day, if on August 1 there were 17 passengers then it was included in class 3 and the prediction for August

2 was 18 passengers. Likewise on August 3, if on August 2 there were 18 passengers, it was included in class 3 and the prediction for August 3 was 18 passengers.

The modelling phase has been carried out, the next step is to conduct a trial run. In this work, the tested data are travel data from 1 August 2019 to 31 August 2019. Based on the test results, the highest number of passengers in a day occurred on August 25, 2019, with 31 passengers and the smallest occurred on August 17, 2019, with 6 passengers. The highest amount of accuracy occurred on the 2nd, 9th, 13th, 29th of August 2019 with a difference of 0, and the largest occurred on the 25th of August 2019 which was 13. If on average the accuracy or difference in prediction with the original data using PE as many as 4.4 people or 34.6%.

4. Conclusion

Fuzzy time series can work well in predicting the number of passengers. Based on the test results, the proposed model has low average accuracy. This is caused by an uncertain number of passengers. For example, the same date in different months even the same date in different years will certainly have different moments such as moments of vacation, weekends, holidays, and others. Further work, we suggest improving this model with these determinant factors.

References

[1] Jian T, Zhi-qiang L, Jian-feng X and Hong-yu G 2019 Cause Analysis and Countermeasures of fatal Traffic Accidents on Road Passenger Transportation Based on Typical Cases *2019 5th International Conference on Transportation Information and Safety (ICTIS)* (IEEE) pp 951-955

[2] Meiying J, Hainana L and Led N 2015 The evaluation studies of regional transportation accessibility based on intelligent transportation system: Take the example in yunnan province of china *2015 International Conference on Intelligent Transportation, Big Data and Smart City* (IEEE) pp 862-865

[3] Bao Y and Jiang X 2010 Modeling the Operation of Passenger Flow on Railway Passenger Dedicated Line Based on Multi-agents *2010 International Conference on Artificial Intelligence and Computational Intelligence* vol 3(IEEE) pp 37-41

[4] Chen S and Liu D 2011 Assessing the influence of public transit headway irregularity on passenger's distribution *7th Advanced Forum on Transportation of China (AFTC 2011)* (IET Conference Publications) pp 130–5

[5] Feng X, Liu J, Li M, Sun Q and Chen S 2010 Prediction of intergraded transportation demand based on development strategy *6th Advanced Forum on Transportation of China (AFTC 2010)* (IET Conference Publications) pp 127–31

[6] Niculescu A I, Wadhwa B and Quek E 2016 Smart City Technologies: Design and Evaluation of An Intelligent Driving Assistant for Smart Parking I*nt J Adv Sci Eng Inf Technol* **6**(6) 1096–102

[7] Ya-jun W, Zhi-jun Z, Song-bai H, Zhao-cheng Y and Pei L 2009 Research on the security efficiency evaluation of urban transportation system based on fuzzy synthetic evaluation *2009 2nd International Conference on Power Electronics and Intelligent Transportation System (PEITS)* vol 2 (IEEE) pp 71-74

[8] Amaliah B, Zeinita A and Suryani E 2016 Dynamics simulation of air passenger forecasting and passenger terminal capacity expansion scenario in Yogyakarta Airport *2016 International Conference on Information & Communication Technology and Systems (ICTS)* (IEEE) pp 187-192

[9] Lin Y, Wan H, Jiang R, Wu Z and Jia X 2014 Inferring the travel purposes of passenger groups for better understanding of passengers *IEEE Transactions on Intelligent Transportation Systems* **16**(1) 235-243

[10] Lin J H and Coyle E J 1988 Generalized stack filters and minimum mean absolute error estimation *1988 IEEE International Symposium on Circuits and Systems* (IEEE) pp 2799-2802

[11] He L, Zhang N and Yin L 2016 Research on the Evaluation of Product Quality Perceived Value Based on Text Mining and Fuzzy Comprehensive Evaluation *2016 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI)* (IEEE) pp 563-566
[12] Zhou N and Chen W 2009 Improving high range resolution ATR performance with average absolute error metric *2009 4th IEEE Conference on Industrial Electronics and Applications* (IEEE) pp 3859-3863
[13] Jonathan B, Rahim Z, Barzani A and Oktavega W 2019 Evaluation of Mean Absolute Error in Collaborative Filtering for Sparsity Users and Items on Female Daily Network *2019 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)* (IEEE) pp 41-44
[14] Mendo L 2009 Estimation of a probability with guaranteed normalized mean absolute error *IEEE communications letters* **13**(11) 817-819
[15] Kim I and Lee S R 1999 A fuzzy time series prediction method based on consecutive values *FUZZ-IEEE'99. 1999 IEEE International Fuzzy Systems. Conference Proceedings (Cat. No. 99CH36315)* vol 2 (IEEE) pp 703-707
[16] Tsai C C and Wu S J 1999 A study for second-order modeling of fuzzy time series *FUZZ-IEEE'99. 1999 IEEE International Fuzzy Systems. Conference Proceedings (Cat. No. 99CH36315)* vol 2 (IEEE) pp 719-725
[17] Hirota K, Nakagawa Y, Sakakibara K, Tanaka K, Kusanagi K and Yoshida Y 1994 A real time fuzzy dynamic image understanding system on general roads *Proceedings of 1994 IEEE 3rd International Fuzzy Systems Conference* (IEEE) pp 106-111
[18] Palit A K and Popovic D 1999 Fuzzy logic based automatic rule generation and forecasting of time series *FUZZ-IEEE'99. 1999 IEEE International Fuzzy Systems. Conference Proceedings (Cat. No. 99CH36315)* vol 1 (IEEE) pp 360-365
[19] Soma H, Hori M and Sogou T 1995 Schedule optimization using fuzzy inference *Proceedings of 1995 IEEE International Conference on Fuzzy Systems* vol. 3 (IEEE) pp 1171-1176
[20] Zardecki A 1994 Fuzzy control for forecasting and pattern recognition in a time series *Proceedings of 1994 IEEE 3rd International Fuzzy Systems Conference* (IEEE) pp 1815-1819
[21] Cui X and Zhao W 2011 Prediction method of transfer volume in Urban Comprehensive Passenger Transportation Hubs based on the categorization of passengers *2011 International Conference on Remote Sensing, Environment and Transportation Engineering* (IEEE) pp 2310-2312