Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia
Vol. 9, 2020, page 298-307

Fuzzy Time Series Forecasting Techniques for Indonesia's Seaweed Production

Nur Afifah Syahirah Shaiful Bahri, Normah Maan
Department of Mathematical Sciences, Faculty of Science
Universiti Teknologi Malaysia, 81310 Johor Bahru, Malaysia

Abstract
Various forecasting methods have been developed based on fuzzy time-series data. Fuzzy time series is a special dynamic process that is linked with linguistic values and its observation is defined and studied. One of the applications of fuzzy time series is to forecast a fuzzy environment in which the data is numerical. The study aimed to forecast Indonesia's seaweed exportation using fuzzy time series models. The data used in this work is from Indonesia's seaweed exportation data from 1989 to 2018. The method consists of three phases: fuzzification, fuzzy inference, and defuzzification. A fuzzy inference system is widely used to process simulation or control. The definition of fuzzy time series is provided, and some properties and procedures to develop fuzzy time series method have been discussed. The forecasted seaweed production has been obtained and analyzed. The results of forecasting Indonesia's seaweed exportation are computed by using Microsoft Excel. The universe of discourse is defined and there are divided into 6 intervals. The data for 1989 cannot be forecast since the previous data did not exist instead of the forecast value for 2019.

Keywords: Fuzzy Time Series; Fuzzy Relation; Linguistic Variable; Fuzzy Logic; Fuzzy Relation

1. Introduction

Indonesia is the largest producer of seaweeds globally. It is because the seaweed can attract the farmers in the rural area as the operating systems are low than the others commodity fishery. Fuzzy time series forecasting can contribute to predicting the future production of seaweed. Chapter 1 includes five main subsections which are the background of the problem, problem statement that needs to be solved. This research also aims to satisfy some objectives that are also included in this chapter. Hence, the scope of the research will also be discussed.

Seaweed is one of the most commodities of the fishery. In Indonesia, this commodity has a great potential to be developed as it is supported by increasing market demand. According to the Ministry of Marine and Fisheries [3], Indonesian seaweed is the most significant international trade and dominates the world market share. The most important seaweed export is dried seaweed. Dried seaweed, gelatine, and carrageenan are seaweed products that are exported by Indonesia that experience fluctuations per annum.

Marketing of dried seaweed from farmer to the exporter is dispensed through four channels: farmer sell to wholesaler and wholesaler to the exporter, wholesaler to factories or farmer sell to retail and retail deal with wholesaler and exporter and lastly retail sell to the domestic market. Indonesia is in the second rank after Korea to export fresh, chilled, and dried seaweed [1]. It implies that Indonesia is one of the biggest suppliers of dried seaweed globally. The importance of seaweed commodities for the Indonesian people is collectively the most reliable source of income. The upstream seaweed sector contributed to exchange reaching US $ 160, 408, 809.

The choice of forecasting methods depends on the price of preparing forecasting and benefit resulting from its use, the fundamental measure in making a call, forecasting period, desired level of accuracy, data quality, and availability, and therefore the level of complexity of the relationships that may be predicted [4]. The fuzzy statistic model is applied as a sound approach for forecasting the longer-term value during a situation where neither a trend is viewed nor a pattern in variations of the time series is visualized, and the knowledge is incomplete and ambiguous. Various forecasting methods have been developed based on fuzzy time series data, but accuracy has been a concern in these forecasts. But classical time series methods cannot deal with forecasting problems in which the time series values are linguistic terms.

This research aims to (1) to investigate the effects of interval length and information on models' forecasting ability, (2) compare the forecasted value with the actual data and (3) manage and predict the exportation of Indonesia's seaweed performance in small-scale areas.

2. Literature Review

2.1. Introduction

The fuzzy time series model deals with the situation where the data are linguistic values, unlike the basic time series approaches that typically manipulate numerical data. The main point in time series analysis is played by processes, the order of observations provides a source of additional information that should be analyzed and used in the prediction process. Time series are typically assumed to be generated at regularly spaced interval of time and so are called regular time series. If data are available in a crisp set, it must be fuzzified before the fuzzy time series methodology may be applied [5]. To form predictions, it is needed to assume that those data do not vary with time. It is typical to assume that either the function itself or one amongst its derivatives is constant in extrapolating functions.

Based on Arumugam [6], the fuzzy sets theory is defined as a mathematical formalization that allows us to eliminate indefiniteness and cope with incomplete, inaccurate information of both qualitative and quantitative naturally. Zadeh [18] introduced the fuzzy set theory and released the idea of classical sets into the fuzzy set so that the classic set which is a crisp set can be a special event of a fuzzy set. Using the membership value function at closed intervals [0, 1] has been approved by Zadeh.

2.2. The Universe of Discourse

All elements of a set are taken from a discourse universe or set of universes that contains all the elements that can be considered in forming the set. In fact, there is no such thing as a set or a fuzzy set, since all sets are subsets of the universal set, although the word set is widely used. In the fuzzy case, every element in the cosmic set is in some way a member of the fuzzy set, even zero. A collection of elements with non- zero members is called as support. The notation $\cup$, is used to denote the universe set.

2.3. Linguistic Variable

According to Zadeh [11], a variable whose values are words or sentences in a natural or artificial language is defined as a linguistic variable. For instance, age is a linguistic variable since its values are linguistic like young, not young, very young, quite young, very old, or not very young besides 19, 35, or 55.

Fuzzy sets are like sets whose elements have degrees of membership. Based on Zimmermann [12], Zadeh wrote that the notion of a fuzzy set provides a convenient point of departure for the development of a conceptual framework which parallels in many respects the framework employed in the case of ordinary sets but in a more general way and hopefully able to influence a broader scope of applicability, within the fields of pattern classification and knowledge processing.

Linguistic variables are variables whose values are words or sentences in a natural or artificial language. Each of those linguistic variables may be assigned one or more linguistic values, or in the other words are in turn connected to a numeric value through the mechanism of membership function.

It can have the other definition which is a variable that can take words in natural language as its values. The words are characterized by fuzzy sets defined in the universe of discourse in which the variable is defined.

2.4. Defuzzification

In most of the real-world application problems, a control command is given as a numerical value. Therefore, it is required to defuzzify the result of the fuzzy inference. The process of attaining a single numeric value or a vector of crisp value from a fuzzy set is called defuzzification, which in some way is the best representation of the output fuzzy set viewed as an isolated entirety.

2.5. Fuzzy Time Series

A forecasting data method that uses the concept of the fuzzy set as the basis of calculations defined by fuzzy time series. The forecasting system with this method works by returning patterns from the actual data and it will be used to project future data. The method also does not require a learning system from a complicated system because the case with genetic algorithms and neural networks so it is easy to use and develop.

3. Methodology

3.1. Research Data

For this research, quantitative research methods is used to analyse the data. Indonesia seaweed production forecasting has been explored by the previous researcher and there is plenty of research paper that can guide this research. This kind of research methodology helps to enrich and understand the fuzzy time series more. Existing data also can help to complete this research. Based on McCombes, [25], this approach can be explained by gathering and selecting the material for inclusion in this research. The data used in this work is from Indonesia's seaweed exportation data from 1989 to 2018 obtained from the Central Statistics Agency (BPS).

3.2. Proposed Method

The stages in the analysis of fuzzy time series using Chen method are as follows:

1) Specify the universe of discourse, U.

$$U=\left[D_{min}-D_1, D_{max}+D_2\right] \qquad (1)$$

where D_1 and D_2 are two positive number between the maximum and minimum value in historical data.

2) Determine the length and number of intervals. In this step, the frequency distribution, F is needed to partitioned the data into several parts using the formula

$$k = 1 + 3.3 \log N \qquad (2)$$

where N is the number of historical data. Then, determine the of interval by using

$$l=\left[\left(D_{max}+D2\right)-\left(D_{min}-D1\right)\right]/n \qquad (3)$$

where L is the length of the interval and k is the number of intervals. Each of the interval can be find by using

$$U_n = \left[D_{min} - D + (k-1)L,\ B_{max} + B + kL\right] \qquad (4)$$

3) Define the fuzzy sets for each linguistic interval. The fuzzy set can be determine by

$$A_i = \sum_{j=1}^{n} \frac{\mu_{ij}}{U_{ij}} \qquad (5)$$

1) Perform the fuzzification of the observable data. The fuzzification process is the process of changing numerical values into fuzzy variables. The method is used to define a cut set for each $A_i(i=1\ldots6)$ and if the collected time series data belongs to an interval u_i, then it is fuzzy to the fuzzy set A_i.

4) Determine Fuzzy Logistic Relation (FLR) and Fuzzy Relation Group (FLRG). Relationship are identified from the fuzzified data. If the time series variable $F(t-1)$ is fuzzified as A_k and $F(t)$ as A_m, then they are related to each other. It is denoted by $A_k \rightarrow A_m$, where A_k is the current state of exportation and A_m is the next state of exportation. Obtain the fuzzy

relationship groups which is can easily get from the state $A2$, then a transition is made to another state $A_i, i = 1, 2, \ldots, n$, as $A2 \rightarrow A3, A2 \rightarrow A2, \ldots, A2 \rightarrow A1$.

5) Determine the forecast value. If $F(t-1) = A_i$, the forecasting $F(t)$ is conducted based on the following rules.
    1. If the membership of output has only one maximum, then select the midpoint of that corresponding interval as the forecasted value.
    2. If the membership of output has two or more consecutive maximum, then select the midpoint of the corresponding conjunct intervals as forecasted value.
    3. On the contrary, standardize the fuzzy output and use the midpoint of each interval to calculate the centroid of the fuzzy set as the forecasted value.
6) Perform the defuzzification process. The result of the defuzzification process in the form of the real numbers is the forecast value. For instance, $F(t) = \left(A_{j1}, A_{j2}, \ldots, A_{jk}\right)$, the defuzzification for F(t) is

$$\widehat{x}_t = \frac{\sum_{p=1}^{k} m_{jp}}{k} \tag{6}$$

where $\widehat{x}_t$ is the defuzzification and m_{jp} is the middle of A_{jp} .

7) Perform Mean Absolute Percentage Error (MAPE). The forecasted values need to be calculate with the actual data to compare the forecasting results by using

$$MAPE = \sum_{i=1}^{n} \frac{\frac{\left|D_i - E_i\right|}{D_i}}{n} \times 100\% \tag{7}$$

where D_i indicate the actual data and E_i represent the forecasting data for year i.

4. Results and discussion

The minimum and maximum data given by actual data are obtained as $D_{min} = 5213188$ and $D_{max} = 159075454$ when defining the universe of discourse. It is suggested that the universe of discourse can be defined as $U = \left[D_{min} - D_1, D_{max} + D_2\right] = [5191870,\ 159150908]$ where $D_1 = 21318$ and $D_2 = 75454$ and divided into 6 intervals. The frequency of the seaweed exportation that happened within every interval is shown in Table 4.2.

Table 4.2. The frequency of seaweed exportation

D MAX	159,075,454
D MIN	5,213,188
Frequency	5.87450014
Range	153,862,266
Interval	25643711

In table 4.2, the frequency can be obtained by finding D_{max} and D_{min}. The range can be obtained by calculating the subtraction between D_{max} and D_{min}. As a result, the interval can be computed by dividing the range by the frequency. Table 4.2 and 4.3 can be referred from Appendix A.

Meanwhile, the lower and upper limits also can be computed by using equation (4). The middle limit can be calculated by adding the lower and upper limits and divide it by two. Table 4.3 summarized the fuzzy interval based on the frequency distribution.

Table 4.3. Fuzzy interval based on the frequency distribution

Interval			
Lower Limit	Middle Limit	Upper Limit	
NA	18035043.5	5,213,188	A1
5,213,188	43678754.5	30,856,899	A2
30,856,899	69322465.5	56,500,610	A3
56,500,610	94966176.5	82,144,321	A4
82,144,321	120609888	107,788,032	A5
107,788,032	146253599	133,431,743	A6
133,431,743	0	159,075,454	A6

According to Table 4.1, the data in 1989 was 12085220 which lies within the boundaries of interval u_1. Since the highest membership degree of u_1 occurs at A_1, the data variable F(1989) is fuzzified as A_1. A complete overview of fuzzified seaweed exportation is show in the Table 4.4.

Table 4.4. Fuzzified seaweed exportation data

Period	Net weight (Kg)	Fuzzification
1989	12,085,220	A_1
1990	11,787,614	A_1
1991	11,304,509	A_1
1992	12,046,665	A_1
1993	16,561,598	A_1
1994	18,688,648	A_1
1995	24,957,652	A_1
1996	22,310,072	A_1
1997	12,698,516	A_1
1998	5,213,188	A_1
1999	25,084,399	A_1
2000	23,073,441	A_1
2001	27,874,058	A_1
2002	28,559,855	A_1
2003	40,162,037	A_2
2004	51,010,828	A_2
2005	69,264,256	A_3
2006	95,588,055	A_4
2007	94,073,398	A_4
2008	99,948,576	A_4
2009	94,002,964	A_4
2010	123,074,961	A_5
2011	159,075,454	A_6
2012	81,462,769	A_3

2013	74,563,691	A_3
2014	81,946,882	A_3
2015	49,914,826	A_2
2016	81,399,417	A_3
2017	38,019,409	A_2
2018	12,056,862	A_1

Next, in order to calculate the fuzzy logistic relation (FLR) in Table 4.5, the defuzzification by equation (6) is needed. The main data of Indonesia's seaweed exportation is a must to calculate FLR. All those values can be optimized using Excel.

Table 4.5. Fuzzy Logistic Relation (FLR)

Defuzzification	FLR		
A1	NA	>	A1
A1	A1	>	A1
A1	A1	>	A1
A1	A1	>	A1
A1	A1	>	A1
A1	A1	>	A1
A1	A1	>	A1
A1	A1	>	A1
A1	A1	>	A1
A1	A1	>	A1
A1	A1	>	A1
A1	A1	>	A1
A1	A1	>	A1
A1	A1	>	A1
A2	A1	>	A2
A2	A2	>	A2
A3	A2	>	A3
A4	A3	>	A4
A4	A4	>	A4
A4	A4	>	A4
A4	A4	>	A4
A5	A4	>	A5
A6	A5	>	A6
A3	A6	>	A3
A3	A3	>	A3

A3	A3	>	A3
A2	A3	>	A2
A3	A2	>	A3
A2	A3	>	A2
A1	A2	>	A1

Hence, the fuzzy logical relationships are grouped into a fuzzy logical relationship group $A2 \rightarrow A1, A2, A3$ which is calculated by equation (5). A complete overview of the relationship groups obtained from Table 4.5 is shown in Table 4.6. The function Microsoft Excel for Table 4.4 and Table 4.5 can be referred to Appendix B. Meanwhile, Appendix C is just formed for Table 4.6.

Table 4.6. Fuzzy Logistic Relation Group (FLRG)

A1	A_1, A_2	30856899
A2	A_1, A_2, A_3	43678754.5
A3	A_2, A_3, A_4	69322465.5
A4	A_4, A_5	107788032
A5	A_6	146253599
A6	A_3	69322465.5

Once the fuzzy logistic relation group is obtained, the data can be forecasted like in Table 4.7. It is clearly showed that in year 1989, the data cannot be forecast as the previous data did not exist. On the contrary, the forecast value for 2019 can be obtained directly since the data for 2018 is there.

Table 4.7. Forecasted seaweed exportation for the period 1989- 2018

Year	Actual data	Forecasted Value	Year	Actual data	Forecasted Value
1989	12,085,220	0	2004	51,010,828	43,678,754.50
1990	11,787,614	30,856,899	2005	69,264,256	43,678,754.50
1991	11,304,509	30,856,899	2006	95,588,055	69,322,465.50
1992	12,046,665	30,856,899	2007	94,073,398	1,077,880,32
1993	16,561,598	30,856,899	2008	99,948,576	1,077,880,32
1994	18,688,648	30,856,899	2009	94,002,964	1,077,880,32
1995	24,957,652	30,856,899	2010	123,074,961	1,077,880,32
1996	22,310,072	30,856,899	2011	159,075,454	1,462,535,99
1997	12,698,516	30,856,899	2012	81,462,769	69,322,465.50
1998	5,213,188	30,856,899	2013	74,563,691	69,322,465.50
1999	25,084,399	30,856,899	2014	81,946,882	69,322,465.50
2000	23,073,441	30,856,899	2015	49,914,826	69,322,465.50
2001	27,874,058	30,856,899	2016	81,399,417	43,678,754.50
2002	28,559,855	30,856,899	2017	38,019,409	69,322,465.50
2003	40,162,037	30,856,899	2018	12,056,862	43,678,754.50

From the above results, it is shown that the data already forecasted. For example, year 1990 is forecasted using the fuzzified exportation of 1989. According to table 4.4, the fuzzified exportation of year 1990 is A_1. From table 4.6, it can be seen that A_1 is related to A_1 and A_2. The highest degree of belongingness of A_1 and A_2 are the sets of u_1 and u_2, where $u_1 = [5213188;\ 30856899]$ and $u_2 = [30856899;\ 56500510]$. The midpoints of the intervals, u_1 and u_2, are 18035043.5 and 43678754.5, respectively. Based on rule 3, the forecasted exportation of 1990 is computed as (18035043.5+43678754.5)/2 = 30856899.

Year 2004 is forecasted using the fuzzified exportation of 2003 as basis. Since the fuzzified exportation of 2003 is A_2, it will be the following FLRG: $A_2 \rightarrow A_1, A_2, A_3$ .The highest degrees of belongingness for the fuzzy sets A_1, A_2 and A_3 are at intervals $u_1 = [5213188;\ 30856899]$, $u_2 = [30856899;\ 56500510]$ and $u3 = [56500610;\ 82144321]$ respectively, and the midpoints of u_1, u_2 and u_3 are 18035043.5, 43678754.5 and 69322465.5, respectively. Hence, the forecasted output is calculated as (18035043.5+43678754.5+69322465.5)/3 = 43678754.5. Since there are no empty relationship groups, rule 2 is not been applied in this research. Table 4.7 shown the complete forecasted exportation of Indonesia's seaweed and also can be referred to Appendix D.

To compare the accuracy of the forecasting results, the mean absolute percentage error is utilized in Table 4.8 below as specified in the research methodology. In addition, the suggested method gives the minimum MAPE values in fuzzy time series forecasting when the frequency of intervals inside the universe of discourse is subdivided depending on the frequency density. These numbers are not specific to each circumstance, but this paper attempted to improve the model's accuracy. The mean absolute percentage error for the derived corrected values is 70.390788%. Table 4.8 summarizes the forecasting results of the Chen model in MAPE.

Table 4.8. Forecasting results of the Chen model in MAPE

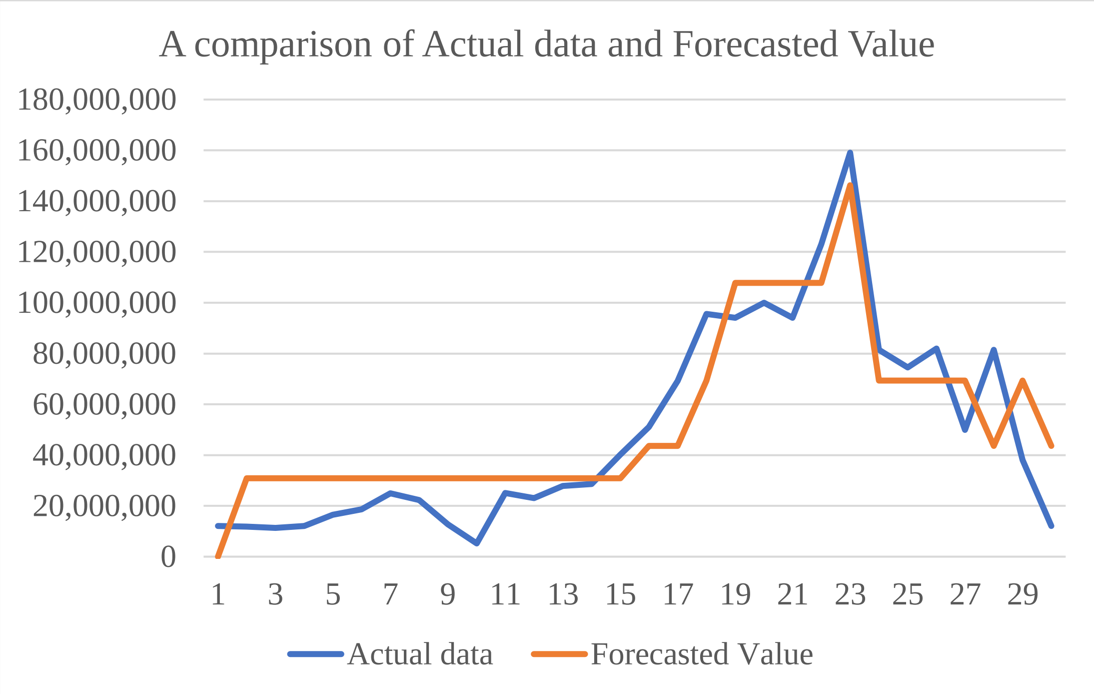
Year	Actual data	Forecasted Value	Absolute Percent Error
1989	12,085,220	0	NA
1990	11,787,614	30,856,899	161.773918
1991	11,304,509	30,856,899	172.960984
1992	12,046,665	30,856,899	156.144742
1993	16,561,598	30,856,899	86.3159521
1994	18,688,648	30,856,899	65.1103868
1995	24,957,652	30,856,899	23.6370272
1996	22,310,072	30,856,899	38.3092757
1997	12,698,516	30,856,899	142.996103
1998	5,213,188	30,856,899	491.900752
1999	25,084,399	30,856,899	23.0123114
2000	23,073,441	30,856,899	33.733408
2001	27,874,058	30,856,899	10.7011365
2002	28,559,855	30,856,899	8.04291198
2003	40,162,037	30,856,899	23.1689892
2004	51,010,828	43,678,754.50	14.373563
2005	69,264,256	43,678,754.50	36.9389682
2006	95,588,055	69,322,465.50	27.4778993
2007	94,073,398	107,788,032	14.5786527
2008	99,948,576	107,788,032	7.84348944
2009	94,002,964	107,788,032	14.6645036

2010	123,074,961	107,788,032	12.4208278
2011	159,075,454	146,253,599	8.06023474
2012	81,462,769	69,322,465.50	14.9028859
2013	74,563,691	69,322,465.50	7.02919267
2014	81,946,882	69,322,465.50	15.4056093
2015	49,914,826	69,322,465.50	38.8815129
2016	81,399,417	43,678,754.50	46.3402121
2017	38,019,409	69,322,465.50	82.3344111
2018	12,056,862	43,678,754.50	262.27299
		MAPE	70.390788

Figure 4.1 shows the graph of comparison between the actual data and forecasted value against year. Appendix E shown that the results of Table 4.8 and how Figure 4.1 is formed.

The results of the analysis showed that the proposed algorithm is rather simple, and it is crucial to focus more on the other efficient method to create a better performance of forecasting time series using the Chen model. Using a higher order of fuzzy logistic relation group will produce more effective results.

Figure 4.1 A comparison of Actual data and Forecasted value



Conclusion

The results of the analysis showed that the proposed algorithm is rather simple, and it is crucial to focus more on the other efficient method to create a better performance of forecasting time series using the Chen model. Using a higher order of fuzzy logistic relation group will produce more effective results.

Acknowledgement

In preparing this report, I was in contact with many people, researchers and academicians. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main final year project supervisor, Assoc. Prof. Dr. Normah Bte Maan, for encouragement, guidance, critics and friendship. I am also very thankful to my academic advisor, Assoc. Prof. Dr. Yeak Su Hoe for his guidance, advices and motivation. Without their continued support and interest, this report would not have been the same as presented here.

References

[1] Bintang, A. S., Huang, W. & Asmara, R. (2019). *Forecasting of Indonesia Seaweed Export: A Comparison of Fuzzy Time Series with and without Markov Chain.* Agricultural Socio-Economics Journal, 19(3), 155-164. doi: http://dx.doi.org/10.21776/ub.agrise.2019.019.3.4

[2] Ministry Marine and Fisheries (2019). Retrieved November 25, 2021, from https://kkp.go.id/djpdspkp/artikel/7947-kinerja-ekspor-produk-perikanan-indonesia-tahun-2018

[3] Brockwell, P.J., R.A. Davis, M.V. Calder. (2002). *Introduction to time series and forecasting.* Vol. 2: Springer.

[4] Song, Q., Chissom, B.S.J.F.s. and systems (1993). *Forecasting enrollments with fuzzy time series-part I.* 54(1): p. 1-9

[5] Arumugam, P., Anithakumari, V.J.I.J.o.E.T. & T.-V. (2013). *Fuzzy Time Series Method for Forecasting Taiwan Export Data.* Issue8-August 2013.

[6] Zadeh, L. A. 1975. The concept of a linguistic variable and its application to approximate reasoning - Part 1. *Information Sciences*, 8:199-249.

[7] Zimmermann, H.J. (2010). *Fuzzy set theory*. Retrieved December 15, 2022, from https://www.mv.helsinki.fi/home/niskanen/zimmermann_review.pdf

[8] Zadeh, L.A., G.J. Klir, & B. Yuan. (1996). *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers*. Vol. 6. 1996: World Scientific. Retrieved December 20, 2022, from https://www.worldscientific.com/doi/pdf/10.1142/9789814261302_fmatter

[9] McCombes, S. (2019). Descriptive Research Design | Definition, Methods & Examples. Retrieved April 25, 2022 from https://www.scribbr.com/methodology/descriptive-research/