International Journal of Engineering & Technology, 7 (2.2) (2018) 90-93

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

Minimum wage prediction based on K-Mean clustering using neural based optimized Minkowski Distance Weighting

Marselina Endah Hiswati[1], Achmad Fanany Onnilita Gaffar[2], Rihartanto[2]*, Haviluddin[3]

[1]Faculty of Science and Technology, Respati University of Yogyakarta, Yogyakarta, Indonesia
[2]Department of Information Technology, State Polytechnic of Samarinda, East Kalimantan, Indonesia
[3]Faculty of Computer Science and Information Technology, Mulawarman University, Samarinda, East Kalimantan, Indonesia
**Corresponding author E-mail: rihart.c@gmail.com*

Abstract

Minimum Wage is a minimum standard used by employers to provide wages to workers in their business environment. The national minimum wage is the average of the provincial minimum wage. There are many factors need to be considered to set a minimum wage. The aim of this study is to predict the minimum wage based on K-Mean clustering concept. The Minkowski Distance Weighting (MDW) then used to estimate a value at the observation point in a cluster by using a linear combination of values of all cluster members around observation point mapped in 3-dimensional Cartesian coordinates. The prediction result by MDW then optimized by Artificial Neural Network Back Propagation (ANN-BP) to obtain a smaller Mean Absolute Percentage Error (MAPE). The net structure which already trained then used to predict the minimum wage for next year.

Keywords: *minimum wage; K-Mean clustering algorithm; MDW method; ANN-BP*

1. Introduction

Net wage/salary per month is wage received during last month, in the form of money or goods, paid by the company/agency/ employer to the employee for the major work done. The components of wage include salary and benefits, overtime pay, transportation allowance and meal allowance. The minimum wage is a minimum standard used by employers to provide wages to workers in their business environment. Since the fulfillment of decent needs in each province varies, it is called Provincial Minimum Wage. The national minimum wage is the average of the provincial minimum wage. The basic measure of the value added arising from economic activity is known as Gross Domestic Product (GDP) at a national level. In the regional level, it is called the Gross Regional Domestic Product (GRDP). GDP or GRDP is the sum of total value added produced by all economic activities and the way of using it. The poverty is measured by the inability to fulfill the basic needs of food and non-food which are measured by consumption or expenditure.

There are many factors need to be considered to set a minimum wage. This research uses a number of populations, the number of poor peoples, and GDP as an independent variable and minimum wage as a dependent variable to perform the minimum wage prediction based on data clustering concept by using K-Mean clustering algorithm. There are many methods that available to solve this problem, from the simple to the sophisticated one. The Minkowski Distance Weighting (MDW) method used to estimate a value at the observation point in a cluster. The Minkowski distance is one of the main distance measures because it generalizes a wide range of other distances[1], [2]. It uses a linear combination of values of all cluster members around the observation point mapped in 3-dimensional Cartesian coordinates (XYZ). The sampled data used are the number of populations as *X*-axis, nominal GDP as *Y*-axis, number of poor peoples as *Z*-axis, and the minimum wage as *V* value obtained from the period of 2000 to 2016 [3]–[7]. The prediction result of MDW then optimized using ANN-BP to obtained a smaller Mean Absolute Percentage Error (MAPE) [8]–[12] The net structure which already trained then used to predict the minimum wage for the next coming year.

2. Experimental detail

K-Means clustering algorithm is one of the best known and the most popular clustering algorithms that used in many domains. This algorithm works on the assumption that the initial centers are provided. These initial centers were used as the starting point in search of final clusters. *K* points from the dataset as the initial cluster center, putting the sample to the class where the nearest cluster center in. Then, the distance of all elements is calculated using a distance formula. *K*-Mean algorithm is shown in Figure 1.

K-Means clustering algorithm is one of the best known and the most popular clustering algorithms that used in many domains.

The k-means algorithm can be divided into two phases: the initialization phase and the iteration phase [13]. This algorithm works based on centroid initialization. Usually, a number of K points from the dataset are used as the initial cluster center. It each iteration, the distance between each element to its centroid is calculated using a certain distance function. The iteration will stop when there is no more change of it cluster members. The K-Mean algorithm is shown in Figure 1.

MDW method is similar to Inverse Distance Weighting (IDW) concept. This concept assumes that values closer to the unsampled location are more representative to estimate than that the farther [14]. The assumption concludes that nearby observations will have a higher weight. MDW interpolator declared:

$$V_P = \sum_{i=1}^{N} w_i . V_i$$

$$w_i = \frac{\left(\left|X_i - X_P\right|^r + \left|X_i - X_P\right|^r + \left|X_i - X_P\right|^r\right)^{1/r}}{\sum_{i=1}^{M}\left(\left|X_i - X_P\right|^r + \left|X_i - X_P\right|^r + \left|X_i - X_P\right|^r\right)^{1/r}} \qquad (1)$$

where $r \geq 1$, (X_i, Y_i, Z_i) represent the location of the sampled point and (X_P, Y_P, Z_P) is interpolation point in 3-dimensional Cartesian coordinates. The variables V_i and w_i are the measured values and their weight at sampled point (X_i, Y_i, Z_i). Thus, V_P and N represent the value at interpolation point (X_P, Y_P, Z_P) and the number of sampled point respectively.

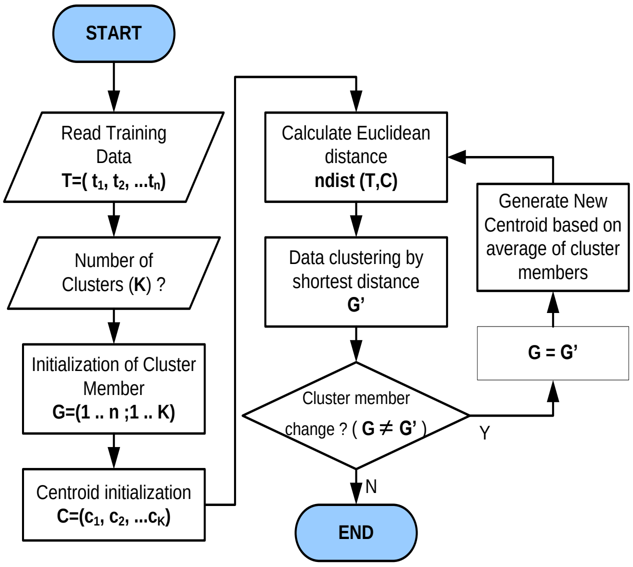
START
Read Training Data
T=(t1, t2, ...tn)
Number of Clusters (K) ?
Initialization of Cluster Member
G=(1 .. n ;1 .. K)
Centroid initialization
C=(c1, c2, ...cK)
Calculate Euclidean distance
ndist (T,C)
Data clustering by shortest distance
G'
Cluster member change ? (G ≠ G')
Y
N
END
G = G'
Generate New Centroid based on average of cluster members

Fig. 1: K-Mean clustering algorithm

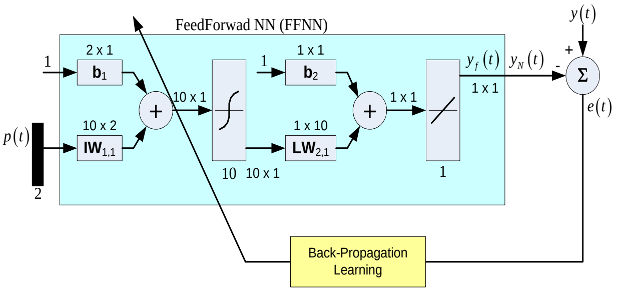

Fig. 2: Artificial Neural Network-Back Propagation

Neural networks are composed of simple elements operating in parallel which are adjusted, or trained so that a particular input leads to a specific target output. Backpropagation is a method used in the artificial neural network to calculate the contribution of each neurons error after a batch of data is processed. The input vectors and the corresponding target vector are used to train a network till it can approximate a certain target [15], [11-12]. Using Feed Forward NN (FFNN), the Artificial Neural Network-Back Propagation (ANN-BP) is shown in Figure 2.

Table 1: Raw data

Year	Number of Population (thousand)	GDP (trillion rupiahs)	Number of poor people (million)	Minimum wage (thousand)
	X	Y	Z	V
2000	205,843	1,389.77	38.70	408.09
2001	208,647	1,684.28	37.90	438.49
2002	212,003	1,863.27	38.40	465.77
2003	215,276	2,013.67	37.30	495.91
2004	217,854	2,295.83	36.10	529.92
2005	218,869	2,774.28	35.10	571.52
2006	222,192	3,339.48	37.30	612.06
2007	225,642	3,950.89	37.17	673.26
2008	228,523	4,951.36	34.96	743.17
2009	234,757	5,606.20	32.53	841.53
2010	237,641	6,446.85	31.02	908.82
2011	238,519	7,422.78	29.89	988.83
2012	245,425	8,672.95	28.59	1,088.90
2013	248,818	9,606.15	28.55	1,296.91
2014	252,165	10,681.77	27.73	1,584.39
2015	255,462	11,654.13	28.51	1,790.34
2016	258,705	12,658.17	30.00	1,997.82

This study uses the net structure of ANN-BP that has 10 hidden neurons and target error $e(t) = 5 \times 10^{-5}$. The performance of ANN-BP training is measured using MSE which is expressed as

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y(t_i) - y_{NN}(t_i)\right)^2 \qquad (2)$$

Where N is the number of the training data, $y(t_i)$ is the ith training target and $y_{NN}(t_i)$ is the ith ANN-BP output. Implementation of ANN-BP is done by using MATLAB programming tool. The performance of predicted results is measured using MAPE and expressed as

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\text{actual}(i) - \text{prediction}(i)\right|}{\text{actual}(i)} \times 100\% \qquad (3)$$

The variance of the distance between data in the cluster and the predicted value of the MDW method will be used as the training input data and the actual minimum wage as the training target. Distance variance expressed by

$$Var = \left(\sum_{i=1}^{n-1}(x_i - \bar{x})^2\right) / (n-1) \qquad (4)$$

Where x_i is the Euclidean distance between the pair of data $\overline{X}$ is the average of distance, n is the number of data. The steps used in this study are:

1. Mapping the sample data in 3-dimensional Cartesian coordinates.
2. Using K-Mean algorithm to cluster the sample data into 3 clusters.
3. The V value as a minimum wage prediction is estimated by implementing the MDW method in each cluster.
4. ANN-BP is used to improve the quality of predicted results.
5. The trained net structure is used to predict minimum wage in next coming year.

This study uses data in Table 1. The sampled data from 2000 to 2014 will be mapped to Cartesian coordinate while data in 2015 and 2016 will be used as test data. Data mapping and clustering are shown in Figure 4, Figure 5, Figure 6 and Figure 7. The minimum wage prediction results using the MDW methods is shown in Figure 8 while optimized MDW is shown in Figure 9.

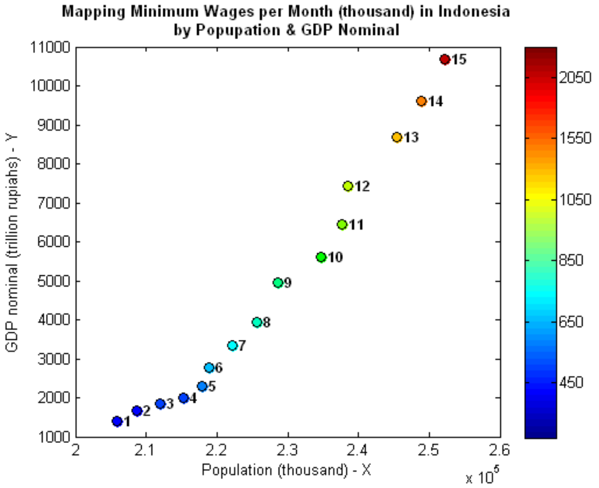

Fig. 4: The Minimum Wage mapping by (X-Y)

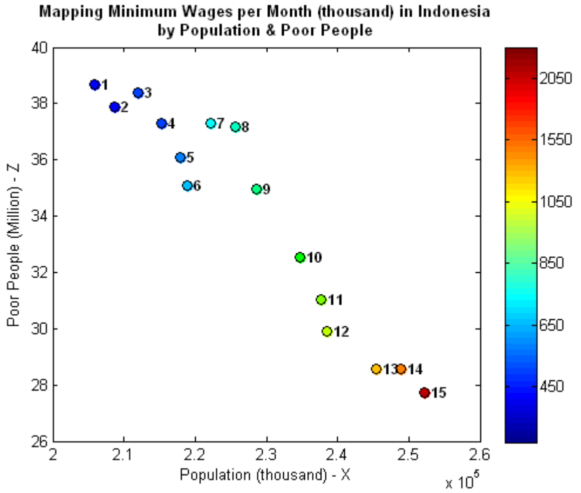

Fig. 5: The Minimum Wage mapping by (X-Z)

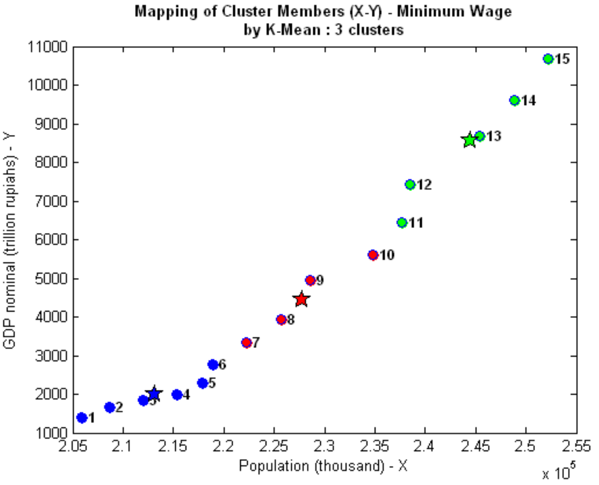

Fig 6: Mapping of cluster members by (X-Y)

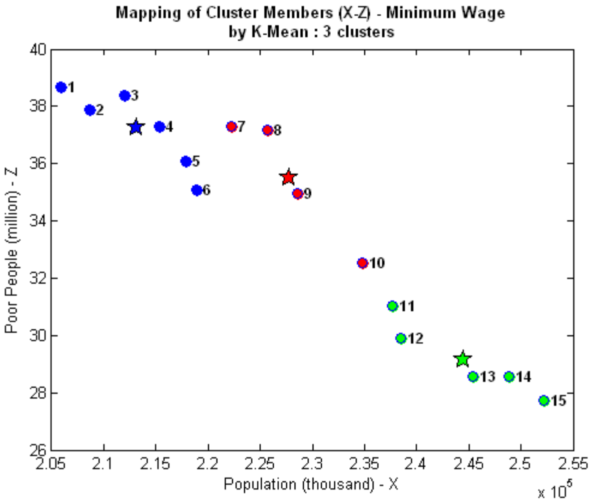

Fig. 7: Mapping of cluster members by (X-Z)

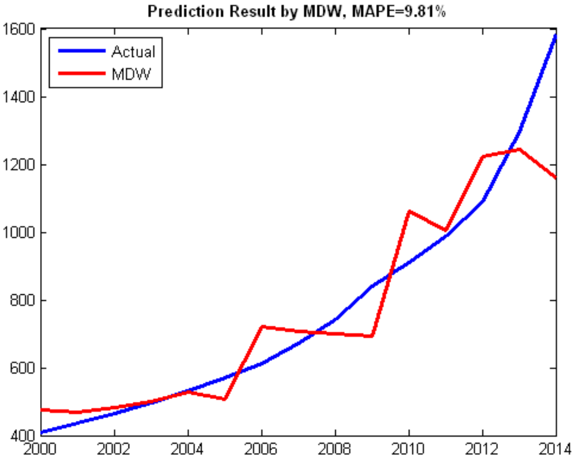

Fig. 8: Prediction result of MDW method

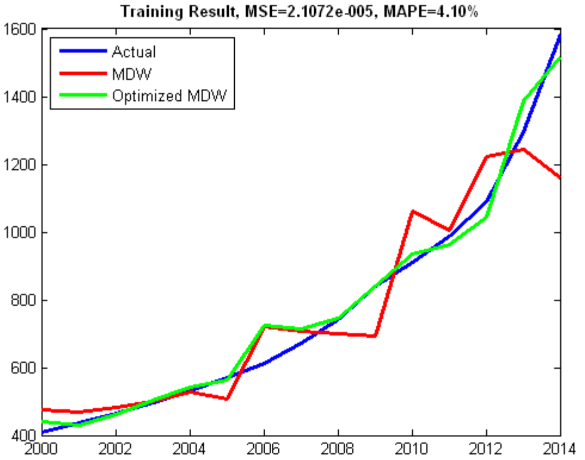

Fig. 9: Prediction result of optimized MDW

3. Result and discussion

From Figure 6, there was a decrease in MAPE value from 9.81% to 4.10% after optimization using ANN-BP, indicating that optimization was successful with performance improvement of:

$$|9.81 - 4.10|/9.81 \text{ x } 100\% = 58.21\%$$

Prior to predict the minimum wage in 2015, it is necessary to calculate the shortest distance between the test data and the final centroid of the clustering result. The goal is to know the cluster where the test data resides, as follows:

$$XY_{2015,dist} = dist\left(W_{K-Mean}, \begin{bmatrix} X \\ Y \end{bmatrix}_{2015}\right)$$

$$= dist\left(\begin{bmatrix} 227{,}779 & 4{,}462 & 36 \\ 244{,}514 & 8{,}566 & 29 \\ 213{,}082 & 2{,}004 & 37 \end{bmatrix}\begin{bmatrix} 255{,}462 \\ 11{,}654.13 \\ 28.51 \end{bmatrix}\right) = \begin{bmatrix} 28{,}603 \\ \boxed{11{,}376} \\ 43{,}465 \end{bmatrix}$$

From the calculation obtained that the test data resides in the second cluster. The distance variance between the test data with all the members of the second cluster and the predicted minimum wage of all second cluster members by the MDW are used as inputs to predict using a trained net structure. The prediction result is $V_{2015} = 1{,}681.45$. The performance of predicted results calculating using Absolute Percentage Error (APE) and obtained:

APE$_{2015}$ = (abs(1,790.34 – 1,681.45) / 1,790.34) x 100% = 6.08%

In the same way for the 2016 prediction results obtained:

APE$_{2016}$ = (abs(1,997.82 – 1,821.10) / 1,997.82) x 100% = 8.85%

4. Conclusion

The performance improvement of 58.21% was gain by MDW optimization using ANN-BP with a final MAPE 0f 4.10%. The minimum wage prediction for 2015 is 1,681,450 rupiahs with APE of 6.08% and 1,821,100 rupiahs with APE of 8.85% for the year of 2016. From the results can be stated that the proposed method is good enough to be applied in the case of minimum wage prediction based on data clustering method with input in the form of the number of population, nominal GDP, and the number of poor people. For the long-term period prediction, each next year's prediction results should be input as sample data. Then the process is repeated from the beginning, and so on until the prediction of the target year is reached.

References

[1] J. M. Merigó and M. Casanovas, "A new minkowski distance based on induced aggregation operators," *Int. J. Comput. Intell. Syst.*, vol. 4, no. 2, pp. 123–133, 2011.

[2] J. M. Merigó and A. M. Gil-lafuente, "Using the OWA Operator in the Minkowski Distance," *Int. J. Econ. Manag. Eng.*, vol. 2, no. 9, pp. 1032–1040, 2008.

[3] BPS - Statistics Indonesia, *Statistical Yearbook of Indonesia 2005/2006*. Jakarta: BPS - Statistics Indonesia, 2006.

[4] BPS - Statistics Indonesia, *Statistical Yearbook of Indonesia 2008*. Jakarta: BPS - Statistics Indonesia, 2008.

[5] BPS - Statistics Indonesia, *Statistical Yearbook of Indonesia 2011*. BPS - Statistics Indonesia, 2011.

[6] BPS - Statistics Indonesia, *Statistical Yearbook of Indonesia 2014*. Jakarta: BPS - Statistics Indonesia, 2014.

[7] BPS - Statistics Indonesia, *Statistical Yearbook of Indonesia 2017*. Jakarta: BPS-Statistics Indonesia, 2017.

[8] A. S. Ahmar, "Sutte Indicator : an Approach to Predicting the Direction of Stock Market Movement," *Songklanakarin J. Sci. Technol.*, 2017.

[9] A. S. Ahmar, "International Journal of Economics and Financial Issues Sutte Indicator: A Technical Indicator in Stock Market," *Int. J. Econ. Financ. Issues*, vol. 7, no. 2, pp. 1–4, 2017.

[10] A. Rahman and A. S. Ahmar, "Forecasting of primary energy consumption data in the United States: A comparison between ARIMA and Holter-Winters models," in *AIP Conference Proceedings*, 2017, vol. 1885.

[11] A. S. Ahmar *et al.*, "Modeling Data Containing Outliers using ARIMA Additive Outlier (ARIMA-AO)," *J. Phys. Conf. Ser.*, vol. 954, 2018.

[12] A. S. Ahmar, "A Comparison of α-Sutte Indicator and ARIMA Methods in Renewable Energy Forecasting in Indonesia," *Int. J. Eng. Technol.*, vol. 7, no. 1.6, pp. 9–11, 2018.

[13] E. Rama Kalaivani, E. R. Marivendhan, and N. Suma, "Prediction of diabetes with hybrid prediction model using big data in health care," *Int. J. Eng. Technol.*, vol. 7, no. 13, pp. 21–23, 2018.

[14] E. Oktavia, Widyawan, and I. W. Mustika, "Inverse distance weighting and kriging spatial interpolation for data center thermal monitoring," *Proc. - 2016 1st Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2016*, pp. 69–74, 2016.

[15] M. H. Beale, M. T. Hagan, and H. B. Demuth, *Neural Network ToolboxTM MATLAB R2015a – User's Guide*. The MathWorks, Inc., 2015.