International Journal of Scientific Research in Mathematical and Statistical Sciences
Volume-8, Issue-1, pp.52-55, February (2021)

Research Paper

E-ISSN: 2348-4519

Performance of Logistic Regression and Multilayer Perceptron Neural Network in Classification of Objects

E. Fajariyanto[1], M. Fajar[2*]

[1,2] Statistics Indonesia (Badan Pusat Statistik)

[*]*Corresponding Author: mfajar@bps.go.id, Tel.: +62-89622459926*

Available online at: www.isroset.org

Received: 31/Dec/2020, Accepted: 01/Feb/2021, Online: 28/Feb/2021

Abstract— Employment is a problem that always gets serious attention from the government. Problems with the quality of human resources, motivation, and work culture are influencing factors in labor force problems. Classification becomes important as a means of evaluation and concluding labor force problems. The classification method itself consists of the conventional method (require assumptions) and the robust method (does not require assumptions. This study uses several classification methods including the logistic regression, the backpropagation algorithm, and the backpropagation algorithm with the addition of momentum. The results of this study show that based on simulation, the backpropagation algorithm and the backpropagation with the addition of momentum results in a higher classification accuracy than the logistic regression, the existence of momentum can increase the accuracy of the backpropagation classification with the addition of momentum.

***Keywords* —** Performance, Logistic Regression, Neural Network, Classification

I. INTRODUCTION

The direction of Indonesian government policy in the field of employment 2014-2019 to solve the problems of the sector of employment among 2014-2019 by realizing vocational training as a national work in improving labor skills and attitude to become a reliable human capital; placing human resources to achieve the creation and expansion of job opportunities that adequate in number and decent in aspects of income and work standards both inside and outside the country, and improve the skills of Indonesian workers through the transfer of skills of foreign workers; achieve industrial relations and harmonious social security through the High Road Industrial Relations System that ensure labor market flexibility without undermining the supremacy of human values; and to realize the labor inspection function as guardian of the implementation of labor legislation by the labor inspection that based on the philosophy that a strong, competent, professional, dignified and worldwide [3].

The direction of government policy was done as anticipation of a free market where the flow of labor from abroad can freely enter Indonesia. In comparison, the regular publication of the ILO (International Labour Organization) showed productivity measured by output per worker, Chinese productivity has increased dramatically and has outperformed the average productivity of the ASEAN countries. The productivity of India is below the average productivity of the ASEAN countries. Compared with other ASEAN countries, Indonesia is relatively higher productivity than the Philippines, Vietnam, Cambodia, Burma; and lower than Thailand, Malaysia, and Singapore. Related to labor productivity, human resource potential is essentially one of the authorized capitals of national development. But this time they felt that the potential of human resources cannot be utilized optimally, given that most of the labor force skills and education levels are still low. This condition is still the most influence on the mental attitude of workers in their work environment resulting in low yields of works [13].

Therefore, the classification technique is important for the evaluation and conclusion of labor-force issues in Indonesia in particular. Individual determinations included in work or unemployed status is one example of classification. Classification is applied when an object needs to be classified in the class/groups that have been set based on the object's attributes [7].

II. RELATED WORK

One of the famous classification techniques is Logistic Regression. The Logistic Regression approach is considered more appropriate if most of the explanatory variables are categorical [4]. Modeling categorical variables with two categories typically used logistic regression [17].

In addition to logistic regression, a discriminant analysis that there is also an analysis by loading one response variable. Discriminant analysis is a multivariate analysis applied to modeling the relationship between one response variable that is dichotomous or multichotomous and qualitative with quantitative explanatory variables. In application, the discriminant function assumes that

explanatory variables data on each group have a multivariate normal distribution and have a similar variance-covariance matrix to any population in the response variable. However, the Logistic Regression analysis method can provide better classification results than the discriminant analysis [12]. Classification in large quantities of data and with a variety of features or attributes often makes the accuracy low. It required a method that has immunity in such diverse data types. The method that can deal with the issue is Multilayer Perceptron [11]. The ability of a neural network to assimilate a variety of different types of data without the need to recognize the distribution characteristics and weight of each data can be utilized to expand the application of the classifier [10].

The multilayer perceptron is usually trained using a backpropagation learning algorithm [14]. However, backpropagation has constraints in terms of data processing time, which is longer. To obtain better results, then the epoch number should be increased. However, this epoch accretion only reduces MSE (Mean Square Error) piecemeal. Therefore, achieving the target of MSE will require epoch accretion that is very large [1]. Backpropagation has advantages such as the addition of momentum and learning rate so the performance of ANN will be quick to recognize the same pattern as compared with no momentum and learning rate (standard backpropagation) [15]. The effects of momentum in the MSE caused the error will decrease [18].

Besides, there will be differences in the results obtained concerning the selection features used. The use of feature selection is very useful to prevent the occurrence of dimensionality effects, reducing the amount of time and memory required by the algorithms used, making it easier to visualize the data and helps to reduce the features that are not relevant (noise) without losing information of the original data [8].

Research to compare several classification methods such as bagging, classification via regression, logistic, logistic boost, logistic regression, multiclass classifier, multilayer perceptron, random forest, and simple logistic had been done in bio-science and bio-technology with the results of these studies that indicate the use of four biomarkers, logistic regression classification methods produce the highest degree of accuracy than other methods [9].On the other hand, reference [8] explains that the supply response model of the Soybean crop in terms of alternative specifications also implications economics. The model considers the availability of production lags concept, the existence of expected price, and gross revenue. The results showed that the existing lags were due mostly to the problems also quick adjustment expenditure rather than correcting the expected time.

III. METHODOLOGY

This paper compared the object classification using logistic regression and Multilayer Perceptron through simulation. Simulation data is carried out to understand the effect of the amount of data and the variation of the data to the accuracy, MSE, and APER obtained from Logistic Regression method[4],[6], Backpropagation [7], Backpropagation with momentum 0.01, and Backpropagation with momentum 0.1. The independent variables consist of two variables, namely x_1 and x_2 raised from the Bernoulli distribution with probability distribution is uniform so that the data $x_1$and x_2 is binary data (0 and 1). While the dependent variable (y) generated from Bernoulli trials with probability $\pi(x_i) = \frac{exp(\sum_{j=0}^{p}\beta_j x_{ij})}{1+exp(\sum_{j=0}^{p}\beta_j x_{ij})}$, which simulated $\beta_0 = -0.43043$, $\beta_1 = -0{,}88075$ and $\beta_2 = 0.92194$. Value of $\beta_0, \beta_1$and β_2 determined based on the results of a calculation using the data SAKERNAS. The simulation will be conducted according to the number of different observations using sample size n = 100, n = 300, and n = 500 with each simulation 100 times.

IV. RESULTS AND DISCUSSION

Simulation

Data were classified using logistic regression method [6], backpropagation [5], backpropagation with momentum [15] 0.01, and backpropagation with momentum 0.1 gives results like the following figure:

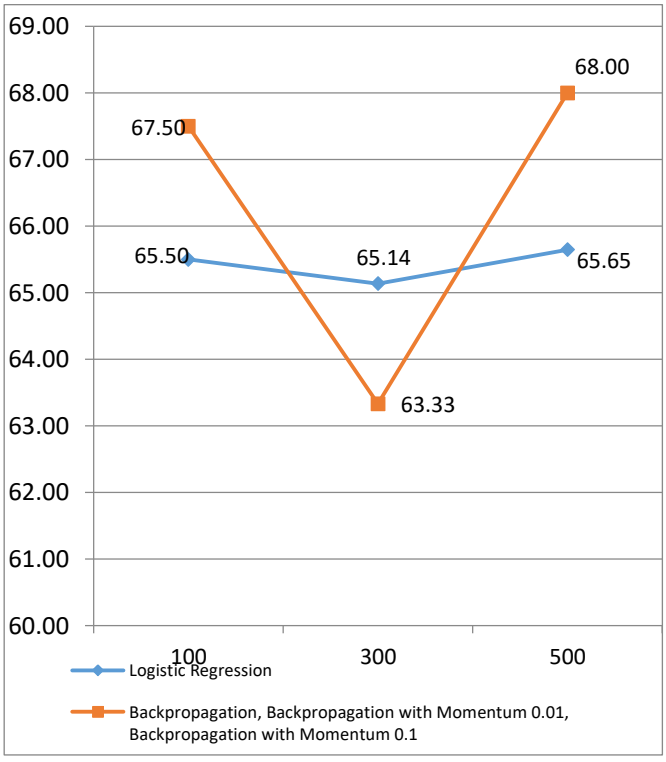

Figure 1. Accuracy of Logistic Regression, Backpropagation, Backpropagation with Momentum 0.01 and Backpropagation with Momentum 0.1

MSE of logistic regression, backpropagation, backpropagation with momentum 0.01, and backpropagation with momentum 0.1 seen in the following figure:

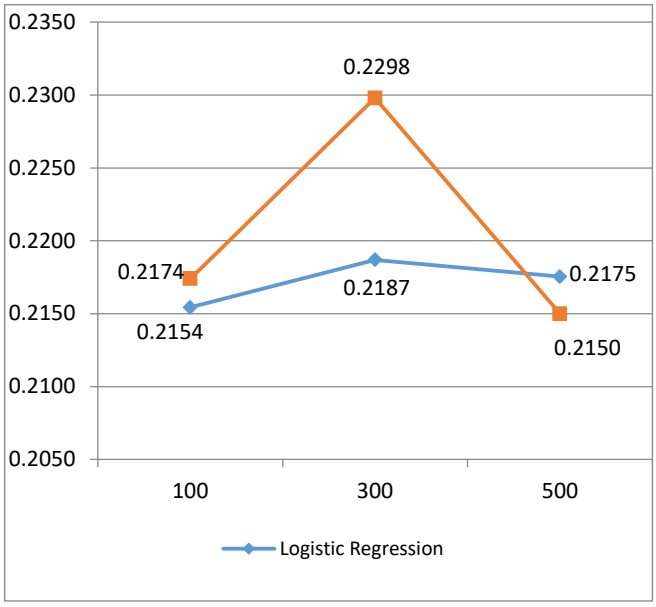

Figure 2. MSE of Logistic Regression, Backpropagation, Backpropagation with Momentum 0.01 and Backpropagation with Momentum 0.1

APER of logistic regression, backpropagation, backpropagation with momentum 0.01, and backpropagation with momentum 0.1 seen in the following figure:

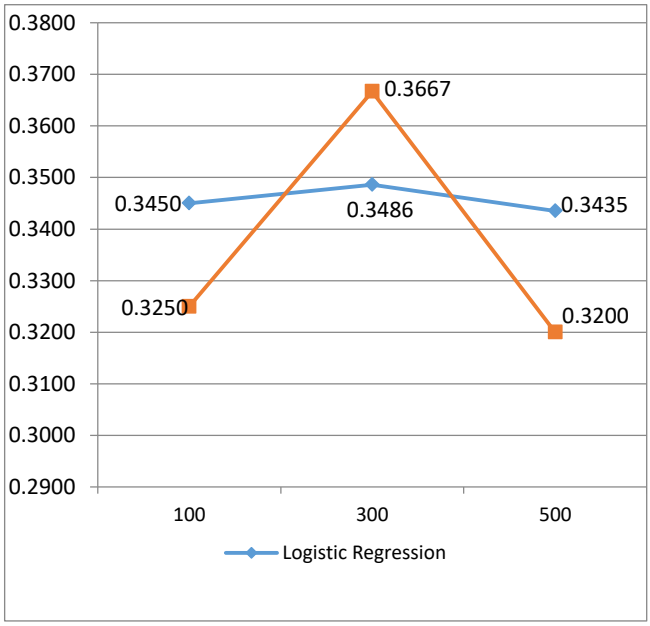

Figure 3. APER of Logistic Regression, Backpropagation, Backpropagation with Momentum 0.01 and Backpropagation with Momentum 0.1

Figure 1 shows the simulation result when the sample size n = 100 resulted in logistic regression accuracy of 65.5%, while the backpropagation algorithm, backpropagation with momentum 0.01, and backpropagation with momentum 0.1 (ANN) produces greater accuracy, respectively amounted to 67.5%. When sample size n = 300, logistic regression accuracy was 65.14%, higher than the ANN is only 63.33%. While the sample size n = 500, ANN accuracy of 65.65% higher than the logistic regression which only amounted to 68%.

Based on Figure 2, the MSE graph shows that the sample size n = 100, logistic regression MSE 0.2154, while the ANN produces MSE of 0.2174. Simulation with a sample size of n = 300 yield logistic regression MSE of 0.2187, while the ANN MSE amounted to 0.2298. Simulation with a sample size of n = 500 showed logistic regression MSE 0.2175 and ANN MSE 0.2150.

Figure 3 shows a sample size of n = 500 logistic regression APER amounted to 0.3450 while ANN APER 0.3250. Simulation with a sample size of n = 300 resulted in the logistic regression APER at 0.3486, while the ANN APER at 0.3667. Simulation with a sample size of n = 500 showed logistic regression APER 0.3435 and ANN APER 0.32.

Based on the previous table shows that the accuracy, MSE, and the resulting APER backpropagation algorithm, backpropagation with momentum 0.01 and backpropagation with momentum 0.1 will be the same in the classification with two independent variables. Besides the addition of the sample size and the more varied the data cause accuracy of logistic regression lower than backpropagation algorithm, backpropagation with momentum 0.01, and backpropagation with momentum 0.1. Increasing the sample size and variety of data also leads MSE and APER of the backpropagation algorithm, backpropagation with momentum 0.01, and backpropagation with momentum 0.1 better than the logistic regression.

Application

- **Data**

Based on the above results, the classification of the individual data of the National Labor Force Survey (SAKERNAS) 2015 BPS Kepahiang Regency Bengkulu Province will use a neural network that has better accuracy in the data with greater sample size and variety. SAKERNAS data consists of 973 people, consisting of 607 people included in the labor force [2] and 366 people including as not a labor force.

- **Compare**

Classification by Multilayer Perceptron Neural Network using Backpropagation algorithm and Backpropagation with momentum 0,1 gives the following results:

Table 1. Average Accuracy of Backpropagation Algorithm and Backpropagation with Momentum,
Epoch = 15,000, Learning Rate = 0.1

Number of Hidden Neuron	Backpropagation Algorithm	Backpropagation with Momentum μ = 0,1
2	90.40	90.53
3	91.47	91.52
4	91.93	91.89
5	92.30	92.42
6	92.67	92.75
7	92.63	92.67
8	92.71	92.59
9	93.12	93.20
10	92.92	92.80
11	93.29	93.16
12	93.25	93.37
13	92.88	93.49

The table above shows that the accuracy of the backpropagation algorithm with the number of hidden neuron 4, 8, 10, and 11 respectively by 91.93%, 92.71%, 92.92%, and 93.29%, higher than backpropagation with momentum 0,1 which produces accuracy respectively by 91.89%, 92.59%, 92.80%, and 93.16%. On the number of hidden neurons 2, 3, 5, 6, 7, 9, 12, and 13, the backpropagation with momentum 0.1 produces higher accuracy than the backpropagation algorithm. Meanwhile, if compared with logistic regression, the accuracy of the two other methods is higher. Logistic regression gives a classification accuracy of 89.51%.

V. CONCLUSION AND FUTURE SCOPE

Classification based on simulated data show that the addition of sample size and more varied data leads to the Logistic Regression accuracy lower than Neural Network. Besides, increasing the sample size and variety of data also leads MSE and APER of Neural Network better when compared with Logistic Regression. The number of independent variables that small can cause accuracy, MSE, and APER of backpropagation and backpropagation with momentum will be the same. On the use of SAKERNAS data, Neural Network provides better accuracy than Logistic Regression, this is in line with the results of research Liew et al (2007) [16]. Meanwhile, if both methods Neural Network compared, the backpropagation algorithm with the addition of momentum resulted in better accuracy than backpropagation.

For further research, the method used in this study was applied to the multi-category classification. Then, the comparison between this method and the support vector machine is also very interesting to study in terms of its performance.

ACKNOWLEDGMENT

Give thanks to Allah, and for our institutions, Statistics-Indonesia (Badan Pusat Statistik).

REFERENCES

[1] A.T.W. Utami, B.S.S. Ulama, "Penerapan Backpropagation untuk meningkatkan Efektivitas Waktu dan Akurasi pada Data Wall-Following Robot Navigation," *Jurnal Sains dan Seni ITS 4(2)*, pp. **2337-3520**, **2013**.

[2] Badan Pusat Statistik, "Pedoman Pencacah Survei Angkatan Kerja Nasional (SAKERNAS) Semesteran 2015," *Badan Pusat Statistik*, **Jakarta**, 2015.

[3] Badan Penelitian, Pengembangan dan Informasi Kementerian Tenaga Kerja dan Transmigrasi, "Naskah Akademik Arah Kebijakan Ketenagakerjaan 2014 – 2019," *Badan Penelitian, Pengembangan dan Informasi Kementerian Tenaga Kerja dan Transmigrasi*, **Jakarta**, 2013.

[4] B. Setyawam, "Pemodelan Regresi Logistik pada Kasus Berat Badan Lahir Rendah (BBLR) dan Pengaruh Agregasi Data terhadap Hasil Pendugaan ", *Institut Pertanian Bogor*, **Bogor**, **2015**.

[5] D. Puspitaningrum, "Pengantar Jaringan Saraf Tiruan," *Penerbit ANDI*, **Yogyakarta**, pp. 129-133, 2006.

[6] D. W. Hosmer, S. Lemeshow, "Applied Logistic Regression," *John Wiley and Sons*, **Canada,** pp. 6-14, 2000.

[7] E. Naibaho, "Perbandingan Backpropagation Neural Network dan Learning Vector Quantization (Studi Kasus: Klasifikasi Daerah Tertinggal di Indonesia Tahun 2015)." *Padjadjaran University*, **Bandung**, **2015**.

[8] F. A. Hermawanti, "Data Mining," *Penerbit ANDI*, **Yogyakarta**, 2007.

[9] H.J. Song, S.K. Ko, J.D. Kim, C.Y. Park, "Looking for the optimal machine learning algorithm for the Ovarian cancer screening," *International Journal of Bio-Science and Bio-Technology 5(2)*, pp. **41-48**, **2013**.

[10] J.D. Paola, R.A. Schowengerdt, "A Review and Analysis of Backpropagation Neural Networks for Classification of Remotely Sensed Multi-Spectral Imagery," *International Journal of Remote Sensing* 16(16), pp. **3033-3058**, **1995**.

[11] M. Mentari, "Klasifikasi Menggunakan Kombinasi Multilayer Perceptron dan Alignment Particle Swarm Optimization," *Seminar Nasional Teknologi Informasi dan Komputasi*, **2014**.

[12] M. Maulidya, "Perbandingan Analisis Diskriminan dan Regresi Logistik (Studi Kasus Klasifikasi Konsumen Berdasarkan Tempat Berbelanja di Wilayah Taman-Sidoarjo)," *Math UNESA* 3(1), **2014**.

[13] M. Sinungan," Produktivitas: Apa dan Bagaimana," *Penerbit Bumi Aksara*, **Jakarta**, **2005**.

[14]M. Zare, H.R. Pourghasemi, M. Vafakhah, B. Pradhan, "Landslide Susceptibility Mapping at Vaz Watershed (Iran) Using an Artificial Neural Network Model: a Comparison Between Multilayer Perceptron (MLP) and Radial Basic Function (RBF) Algorithms," *Arab Journal Geosciences 6 (8)*, pp. **1-16**, **2012**.

[15]N. Kamiyama, N. Iijima, A. Taguchi, H. Mitsui, Y. Yoshida, M. Sone, "Tuning of Learning Rate and Momentum on Back-Propagation," *Singapore ICCS/ISITA*, pp. **528-532**, **1992**.

[16] P.L. Liew, Y.C. Lee, Y.C. Lin, T.S. Lee, W.J Lee, W. Wang, C.W. Chien, "Comparison of Artificial Neural Networks with Logistic Regression in Prediction of Gallbladder Disease Among Obese Patients," *Digestive and Liver Disease 39*, pp. **356-362**, **2007**.

[17] R. Sastri, "Pemodelan Kejadian Kematian Bayi di Indonesia menggunakan Regresi Logistik Terboboti ", *Institut Pertanian Bogor*, **Bogor**, **2015**.

[18] S. Mishra, V.K. Singh, "Monthly Energy Consumption Forecasting Based On Windowed Momentum Neural Network. IFAC-PapersOnLine 48-30, pp. **433–438**, **2015**.

AUTHORS PROFILE

Eko Fajariyanto is currently working as a Statistician at BPS-Statistics Indonesia of Bengkulu Province. Previously, he served at BPS-Statistics Indonesia of South Bengkulu Regency, BPS-Statistics Indonesia of Rejang Lebong Regency, BPS-Statistics Indonesia of Kepahiang Regency, and BPS-Statistics Indonesia of Seluma Regency. He graduated from the Padjadjaran University in 2017. His research interests are economics, neural networks, and forecasting method comparisons.

Muhammad Fajar is currently the Statistician of the BPS-Statistics Indonesia Office. In March 2017, he published paper with title; Spectral analysis and markov switching model of Indonesia business; in AIP Conference Proceedings, in November 2017 he published paper with title; PEMODELAN KURVA ENGEL SEDERHANA INDONESIA (Pendekatan Bayesian Quantile Regression); in National Statistics Seminar of Padjadjaran University. In February 2019 he published paper with title; The Estimation of Production Function and Technical Efficiency Shallot Farming; in Jurnal Matematika Mantik of UIN Sunan Ampel. In October 2019 he published paper with title; An Application of Hybrid Forecasting Singular Spectrum Analysis – Extreme Learning Machine Method in Foreign Tourists Forecasting; in Jurnal Matematika Mantik of UIN Sunan Ampel.

