

PAPER • OPEN ACCESS

Predicting GDP of Indonesia Using K-Nearest Neighbour Regression

To cite this article: Bagus Priambodo *et al* 2019 *J. Phys.: Conf. Ser.* **1339** 012040

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Predicting GDP of Indonesia Using K-Nearest Neighbour Regression

Bagus Priambodo^{1*}, Sarwati Rahayu², Al Hamidy Hazidar³, Emil Naf'an⁴, Mardhiah masrif⁵, Inge Handriani⁶, Zico Pratama Putra⁷, Asama Kudr Nseaf⁸, Deni Setiawan⁹, Yuwan Jumaryadi¹⁰

^{1,2,6,10}Faculty of Computer Science, Universitas Mercu Buana, Kembangan, Jakarta, Indonesia

^{3,8}Institute of Visual Informatics, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

^{4,5}Faculty of Computer Science, Universitas Putra Indonesia YPTK, Sumatera Barat, Indonesia

^{7,9}School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK

*bagus.priambodo@mercubuana.ac.id

Abstract. The impact of the global recession in 1998 that originated from the recession in the US will affect the projected economies in Asia, including Indonesia, both direct and indirect nature. In this study, we predicted Indonesia's GDP in the event of the economic crisis that hit Indonesia starting in 1998. Instead of using the famous prediction algorithm as a neural network and linear regression. K-Nearest Neighbour is selected because it is easy and fast to use in the small dataset. We use a dataset from 1980-2002, consisting of rice prices, premium prices, GDP of Japanese country, American GDP, currency exchange rates, Indonesian government consumption, and the value of Indonesia's oil exports. For evaluation, we compare k-NN regression prediction result with prediction result using back propagation neural network and multiple linear regression algorithm. Result show, k-NN regression is able to predict Indonesia's GDP using small dataset better than the neural network, and multiple linear regression method.

1. Introduction

¹ bagus.priambodo@mercubuana.ac.id

² sarwati.rahayu@mercubuana.ac.id

³ al_didy@yahoo.com

⁴ emilnafan@upiypk.ac.id

⁵ mardhiah_m@upiypk.ac.id

⁶ inge.handriani@mercubuana.ac.id

⁷ z.putra@qmul.ac.uk

⁸ asama@ukm.edu.my

⁹ d.setiawan@se17.qmul.ac.uk

¹⁰ yuwan.jumaryadi@mercubuana.ac.id



The understanding of the phenomenon of Indonesia's economy and its relationship with the possibilities that occur in the future are very helpful in determining attitudes or policies that lead to hope or expectation to be achieved. To understand the phenomenon of the Indonesian economy both now and in the future, we need to develop quantitative methods to be able to do the analysis of the phenomenon of the Indonesian economy. BPS has developed a model for forecasting economic growth of Indonesia based on the Keynesian Model [1], [2]. Neural networks and linear regression have been widely used for prediction methods, from the economic field (macro economy) [3], [4], also the problem of predicting traffic flow[5] [6]. The neural network is more accurate than linear regression but they need a lengthy amount of time required to train the network[7]. In this study, we use non-parametric k-Nearest neighbour regression for predicting Indonesian GDP from 1998 to 2002 when the economic crisis hit Indonesia in 1998. K-nearest neighbour non-parametric regression method is a reliable method for short-term prediction, which can reflect the nonlinear characteristics between parameters. We use the same parameters that used in Keynesian Model [1], [2], the result of prediction using the k-NN method then compare with the result of prediction using backpropagation neural network and multiple linear regression method.

2. Related Work and Problem

Many studies have applied K-NN to solve the problem in many fields for classification and regression [8]. Some studies use k-NN for classification, predict heart disease [9] K-NN prediction result show better accuracy compare with naïve Bayes and decision tree, classification of the poor for health cards distribution[10] the accuracy of the results of determining feasibility using a combination of K- Nearest Neighbour-Naive Bayes Classifier algorithms is better than just the K-Nearest Neighbour algorithm. While some studies use K-NN for regression, in this study we predict traffic speed using K-NN[7] result of prediction using similar data show better accuracy than using all data, other study use k-NN for predicting stock price[11], also there is study predicting using multivariate k-NN model for dependent variable [12]

In economic field many studies have been done in predicting GDP and inflation, using various method and algorithm, most of them use the neural network and linear regression [3], [4], [11], [13], [14], there was also study in predicting GDP using k-NN. Indonesia was hit by the economic crisis in 1998, seen in the graph in figure 1 below, Indonesia's economic growth declined sharply. We want to find out whether the K-NN regression is able to predict Indonesia's GDP after experiencing a crisis. Next, we will evaluate whether the K-NN prediction results are better than using neural networks and multiple linear regression for short time prediction.



Figure 1. Line chart of Indonesia GDP from 1980 to 2002 [1], [2]

3. Dataset and Method

3.1. Dataset

Our dataset is collected from BPS, are data from 1992 until 2002, follow [1], [2]. Component inputs are variable based on the model Keynesian macroeconomics.

Table 1. Example of GDP dataset [1], [2].

Year	P_RICE	P_GASOL	YUSA	YJPN	NGC	ER	GDP
1980	218	150	5162	313140	49918.76863	631.78	510343.7318
1981	243	150	5292	322325	54992.65957	636.58	546392.529
1982	263	240	5189	331236	59520.0004	666.4	548704.6318
1983	294	320	5424	336575	58943.78765	906.15	571112.8582
1984	310	350	5814	347072	61006.2885	1026.6	616038.9267
1985	307	385	6054	364712	68654.20321	1110.1	638044.8962
1986	335	385	6264	375502	68691.4393	1288.8	685104.8487
1987	369	385	6475	389753	67679.7536	1644.3	729413.662
1988	458	385	6743	416119	72831.27233	1687.3	778293.2381
1989	476	385	6981	438135	80335.29163	1771.7	849093.7644
1990	515	450	7113	460925	84219.20541	1848.3	925909.2462

3.2. Methods

The KNN algorithm is a method for classifying objects based on nearest training examples in the feature space. While testing the data, it will find the value of K closest to the test data. The nearest neighbours are defined in the Euclidean range between two points $X=(x_1, x_2, \dots, x_n)$ dan $Y=(y_1, y_2, \dots, y_n)$ [15].

3.2.1. Minkowski distance.

Our study uses K-NN following R library kkn [16]. The Minkowski distance. It is expressed as [17]:

$$d = \left(\sum_{i=1}^m |u_i - v_i|^p \right)^{1/p}$$

where $(u_1, u_2, u_3, \dots, u_m)$ and $(v_1, v_2, v_3, \dots, v_m)$ are two vectors in m-dimension Euclidean space, and p is a positive real number. It is a generalization of all commonly used metrics on a Euclidean space.

3.2.2. Multivariate prediction

Based on the Keynesian model type following [1], [2], there are some functions that will be used as a forecast model is

$$GDP = PC + PCF + GC + GCF + EX-IM + S$$

$$PC = f(GDPV / PGDP, PC (-1), PPC)$$

$$CF = f((LO + IMEQ) / PCF, GDP (-1), FI, ER)$$

$$IM = f(GDP, IM (-1), PIM / PGDP)$$

$$PGDP = f(M1, M2, PEX, P_RICE, P_GASOL, ER)$$

$$XNO = f(GDP-USA, GDP-Japan, PXNO, ER, IM)$$

$$EX = XNO + XO$$

$$PPC = f(PGDP)$$

From the econometric equation that is formed will be seen that the Exogenous Variable used as assumptions macroeconomic variables are:

- a. LO (Outstanding Bank Credits)
- b. M1
- c. P_RICE (Rice Price)
- d. P_GASOL (Premium)
- e. Japan GDP
- f. GDP-USA
- g. ER
- h. GC
- i. XO (Export of Oil)

Only 7 components to be used in the manufacture of artificial neural network models, namely:

- a. P_RICE (Rice Price)
- b. P_GASOL (Premium)
- c. Japan GDP
- d. GDP-USA
- e. ER (Exchange Rate)
- f. GC (Government Consume)
- g. XO (Export of Oil).

Where P_RICE is the price of rice, P_GASOL is a premium price, JAPAN GDP is the value of the GDP of the country of Japan, GDP USA GDP is the value of the American States, the ER is the currency exchange rate, GC is the Indonesian government consumption, and XO is Indonesia's oil export.

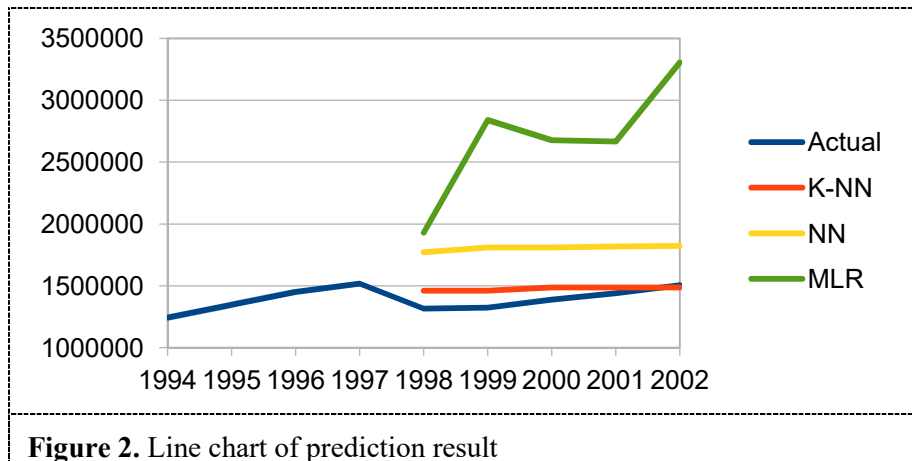
The dependent variable is GDP Indonesia (Y). For predicting GDP Indonesia, the independent variables are P_RICE, P_GASOL, Japan GDP, GDP USA, ER, GC, and XO.

4. Result and Discussion

We use data 1980 to 1997 as training data. Then we predict GDP 1998 to 2002, and for evaluation, we calculate the error of the result of prediction with actual data. Our result of prediction is shown in table 2 below and visualise in figure 2 below. As seen in figure 2 below, prediction GDP using K-NN with K=3 show better result compare with neural network and multiple linear regression. Line chart of K-NN result is the nearest with actual data.

Table 2. Prediction Result

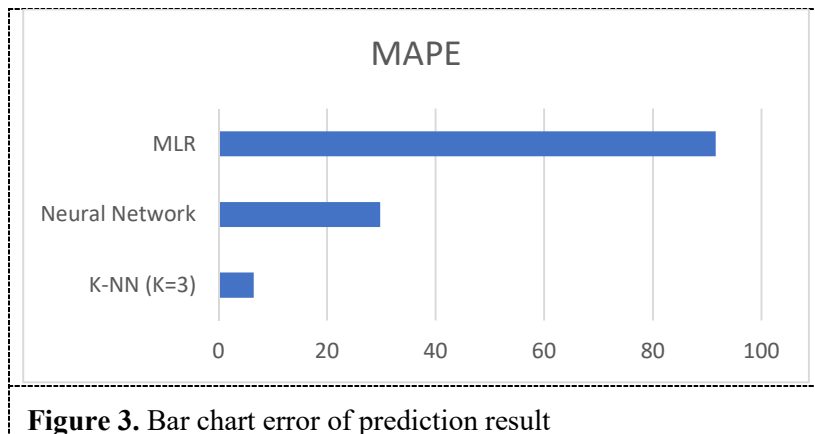
Year	Actual	K-NN (K=3)	NN	MLR
1998	1317281	1461908	1772240.097	1928309
1999	1325352	1461908	1811397.378	2838850
2000	1389770	1485901	1810886.817	2676212
2001	1440406	1485901	1819270.103	2667334
2002	1505216	1485901	1823764.3	3305946



To evaluate our prediction result we calculate the error using mean absolute percentage error (MAPE). The result is shown in table 3 and visualizes in figure 3 below. As seen in table 3 and figure 3, K-NN with $k=3$ show lowest MAPE value compares to the neural network and multiple linear regression. This show that the K-NN prediction result has better accuracy than the neural network and multiple linear regression.

Table 3. MAPE error result

Name	MAPE
K-NN (K=3)	6.53
Neural Network	29.80
MLR	91.59



5. Conclusions

Generally, our research objective is to predict Indonesia GDP when Indonesia hit by the economic crisis in 1998 using only 17-row training data (1980-1997). Experiment result shows that multivariate prediction using K-NN show better result compare with neural network and multiple linear regression for short time prediction using little data training.

6. Acknowledgement

The authors gratefully acknowledge the use of service and facilities funded by Universitas Mercu Buana.

7. References

- [1] S. Baidowi, "Indonesia's Economic Forecast Model," *Cent. Bur. Stat.*, 2006.
- [2] S. Baidowi, "Indonesian Economic Performance And Forecasting Model," *BPS Stat. Indonesia.*, 2004.
- [3] E. Giovanis, "ARIMA and Neural Networks: An Application to the Real GNP Growth Rate and the Unemployment Rate of U.S.A.," *Ssrn*, pp. 1-18, 2009.
- [4] M. A. Choudhary and A. Haider, "Neural network models for inflation forecasting: an appraisal," *Appl. Econ.*, vol. 44, no. 20, pp. 2631-2635, 2012.
- [5] B. Priambodo and A. Ahmad, "Predicting traffic flow based on average speed of neighbouring road using multiple regression," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10645 LNCS, pp. 309-318.
- [6] B. Priambodo and A. Ahmad, "Traffic flow prediction model based on neighbouring roads using neural network and multiple regression," *J. Inf. Commun. Technol.*, vol. 17, no. 4, pp. 513-535, 2018.
- [7] B. Priambodo and Y. Jumaryadi, "Time Series Traffic Speed Prediction Using k-Nearest Neighbour Based on Similar Traffic Data," *MATEC Web Conf.*, vol. 218, p. 03021, 2018.
- [8] S. B. Imandoust and M. Bolandraftar, "Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background," *Int. J. Eng. Res. Appl.*, vol. 3, no. 5, pp. 605-610, 2013.
- [9] I. K. A. Enriko, M. Suryanegara, and D. Gunawan, "Heart disease prediction system using k-Nearest neighbor algorithm with simplified patient's health parameters," *J. Telecommun. Electron. Comput. Eng.*, vol. 8, no. 12, pp. 59-65, 2016.
- [10] Y. F. Safri, R. Arifudin, and M. A. Muslim, "K-Nearest Neighbor and Naive Bayes Classifier Algorithm in Determining The Classification of Healthy Card Indonesia Giving to The Poor," *Sci. J. Informatics*, vol. 5, no. 1, p. 18, 2019.
- [11] K. Alkhatib, H. Najadat, I. Hmeidi, and M. K. A. Shatnawi, "Stock Price Prediction Using K-Nearest Neighbor Algorithm," *Int. J. Business, Humanity. Technol.*, vol. 3, no. 3, pp. 32-44, 2013.
- [12] D. G. Uegan and P. R. Akotomarolahy, "The Multivariate k-Nearest Neighbor Model for Dependent Variables: One-Sided Estimation and Forecasting," *Doc. Trav. du Cent. d'Economie la Sorbonne*, vol. 50, pp. 106-112, 2009.
- [13] M. Pasarica, Armand Eugen; Popescu, "Academy of Economic Studies, Bucharest, Romania PREDICTION OF THE GDP IN ROMANIA," in *The 7th SEA International Conference*, 2017, no. April 2015.
- [14] M. Jahn, "Artificial neural network regression models: Predicting GDP growth," *Hambg. Inst. Int. Econ.*, 2018.
- [15] Mustakim, "Effectiveness of K-means clustering to distribute training data and testing data on K-nearest neighbour classification," *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 21, pp. 5693-5700, 2017.
- [16] K. Schliep and K. Hechenbichler, "kkn: Weighted k-Nearest

Neighbors." 2016.

- [17] B. Lu, M. Charlton, C. Brunson, and P. Harris, "The Minkowski approach for choosing the distance metric in geographically weighted regression," *Int. J. Geogr. Inf. Sci.*, vol. 30, no. 2, pp. 351–368, 2016.