Implementation of bayesian model averaging on the weather data forecasting applications utilizing open weather map

To cite this article: R F Rahmat et al 2018 IOP Conf. Ser.: Mater. Sci. Eng. 309 012054

View the article online for updates and enhancements.
Implementation of bayesian model averaging on the weather data forecasting applications utilizing open weather map

R F Rahmat*, F R Nasution, Seniman, M F Syahputra and O S Sitompul
Department of Information Technology, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia

*E-mail: romi.fadillah@usu.ac.id

Abstract. Weather is condition of air in a certain region at a relatively short period of time, measured with various parameters such as; temperature, air pressure, wind velocity, humidity and another phenomenons in the atmosphere. In fact, extreme weather due to global warming would lead to drought, flood, hurricane and other forms of weather occasion, which directly affects social and economic activities. Hence, a forecasting technique is to predict weather with distinctive output, particularly mapping process based on GIS with information about current weather status in certain coordinates of each region with capability to forecast for seven days afterward. Data used in this research are retrieved in real time from the server openweathermap and BMKG. In order to obtain a low error rate and high accuracy of forecasting, the authors use Bayesian Model Averaging (BMA) method. The result shows that the BMA method has good accuracy. Forecasting error value is calculated by mean square error shows (MSE). The error value emerges at minimum temperature rated at 0.28 and maximum temperature rated at 0.15. Meanwhile, the error value of minimum humidity rates at 0.38 and the error value of maximum humidity rates at 0.04. Afterall, the forecasting error rate of wind speed is at 0.076. The lower the forecasting error rate, the more optimized the accuracy is.

1. Introduction
Global warming is the uprising of greenhouse gas in the atmosphere factored by human activities leading mainly to weather inconsistency. It also allows extreme weather that will bring more severe storm, drought, floods, hurricanes and more weather phenomenon which not only impact the social and economic life directly, but also bring difficulties for mankind on agriculture, water supply and forestry [1]. The weather is a condition of air in a relatively short time at a specific area [9], which is depicted by the rates of various parameters such as; temperature, air pressure, wind speed, humidity, and other atmospheric phenomena. The selected method to properly determine the weather conditions is with the current activities frequently performed by atmospheric or weather researchers [2]. The current activities frequently performed by previous researchers are critical since there are greater demands from various stakeholders for a more rapid, accurate, and detailed information about the atmospheric conditions. Even some other parties demand the availability of forecasting over the atmospheric conditions with a fairly short time span within daily, hourly, or even in minutes. Given this need, researchers are encouraged to continuously conduct atmospheric studies related to atmospheric weather and develop even more sophisticated methods to determine the weather conditions for a better and more accurate results [3].

There are several methods to do forecasting, for example Evolving Connectionist System [7] or Exponential Smoothing Methods [8], and Bayesian Model Averaging to predict cases of pneumonia in
infants [4]. Since the data pre-processing in this study is a long process, we prefer to use BMA as our proposed method. In this study, Bayesian Model Averaging (BMA) method is performed to forecast the weather data from three to seven days onwards. BMA is a method that is capable to predict the best model based on the weighted average of all models. BMA evenly averages the posterior distribution of all the possibly formed models. The purpose of the BMA is to combine uncertain models in order to get the best model. The estimation result includes all models that may have formed so they can get a better estimation [5].

Impacts of global warming on the climate and weather conditions that occur at these days are unstable and fluctuate. This causes many environmental problems and social issues which directly impact livings. Therefore, forecasting techniques to predict weather with diverging outputs are necessary; such method which output is mapped into each coordinate of a region presenting detailed information on the status of the weather. Weather forecast within three to seven days ahead in a particular area would also be provided, so that the environmental and social problems could be anticipated.

Previous research forecasted the monthly rainfall in Medan, Northern Sumatera, using box-Jenkins method. This study was conducted to analyse, implement and apply the Box-Jenkins method in order to predict the rate of monthly rainfall in Medan. The data used is the historical data of relative humidity and rainfall in Medan from 2005 to 2009. The forecasting model used for this data is according to the ARIMA model identification by using the order (111; 111). In the evaluation process the estimated value of the constant is tested against the value zero by using statistical t-test. [6].

Following research about the weather forecast was based on Time Series Data using Adaptive Neuro Fuzzy Inference System (ANFIS). The data used in this research sourced from secondary observational data of Badan Meteorologi, Klimatologi dan Geofisika (BMKG) from Karangploso sub district, Malang regency, from January 2011 to May 2012 and was obtained with the accuracy ranging from 48.39% to 76.67 % [3]. Bayesian Model Averaging method has also been employed for other researches. The method is used to predict cases of pneumonia in infants. This method is used to predict the cases of pneumonia in children under the age of five in Situbondo [4].

2. Methodology

Weather forecasting system using Bayesian Model Averaging is a system that serves to generate forecasting from three to seven days. The employed data are weather data obtained from combination of BMKG OpenWeatherMap API and API. The collected data are passed as preprocessing data. Preprocessing consists of three stages; namely data cleaning, data integration, and data selection. Then subsequently these data are next entered into the forecasting process consisting of several stages of training data normalization, determining activation function process, calculating on each weight differences and calculating on net weight and output. The testing process is then performed. Forecasting results on testing process are analyzed using the forecasting error rate Mean Square Error (MSE) and the analysis of forecasting results. Furthermore, the forecasting results are sent to the web server.

2.1. Data acquisitions

In early stage prior to data processing, collection of data from the openweathermap server and BMKG are implemented. In data collection process, historical data and classification data are required.

2.1.1. Historical data. Historical data acquired is the weather data at some geographical points of the city’s districts in North Sumatra. The collection comes from raw data list retrieved from the site in openweathermap and data.bmkg.go.id. The next data sources are sorted and stored into a database application. List of historical raw data used in this study is shown in figure 1 below.
Data sources from bmkg.go.id are updated daily that triggered by values of weather parameters which is later processed in this research. The sample of the data are shown in figure 3 below. Then the web server sends data to the web client forecasting view. Eventually, the output is informed in PDF-based reports and graphs. The general architecture of this research can be seen in figure 2.

**Figure 1.** List of raw data from Openweathermap

**Figure 2.** General architecture

**Figure 3.** List of raw data from bmkg.go.id
2.1.2. *Data classification.* In this stage, the data is identified. Such identification can be accomplished for a group or several groups of data so that later becomes characteristic of the data concerned. In this study, there are several categories or types of data from data sources, such as satellite imagery, world-weather data, list of locations or areas with high-potential waves, and so forth. For this study, the data used is particularly categorized on weather in North Sumatra.

2.2. *Pre-processing*

On this stage, the process has been focusing to pre-processing steps in the application. The collected data are processed to enhance information from these data using Bayesian Model Averaging. The techniques are staged as follows:

2.2.1. *Data cleaning.* This stage is to eliminate data noise or unintelligible data which lead to inconsistency on the ultimate goal of forecasting. Data from the data source is composed using numerous attributes. Some unnecessary attributes in the forecasting process are removed. Several attributes are taken, such as; weather status, minimum temperature, maximum temperature, id city, latitude and longitude, minimum humidity and maximum humidity, and wind speed. Raw data obtained from the data source is formatted in XML, as shown in figure 4 below.

2.2.2. *Data integration.* This stage is to merge the collected data from the openweathermap server and BMKG to avoid dataduplication.

2.2.3. *Data selection.* This stage is to select relevant data, such as; air temperature, humidity, wind speed, wind direction, and weather status which are essential for weather forecasting necessity. At this stage, the data collected from data source will be reformatted with defined attributes and converted into XML format as shown in figure 4 below.

2.3. *Bayesian model averaging forecasting*

Once the data has undergone a series of pre-processing, the data is then stored as initial data to be used in the forecasting process. There is a chronology of the forecasting process using Bayesian Model Averaging (BMA). Dataset used as training data on this flow is the weather data obtained from BMKG and OpenWeatherMap consistingly; minimum temperature, maximum temperature, minumum humidity, maximum humidity, wind speed and weather status. The weather data is normalized to facilitate the calculation process. In the process, max_forecast_day variable, also specified as the maximum value, is set to determine the results of forecasting. In the calculation of forecasting the weather data, the value of the max_forecast_day variable is one week since the results of forecasting the weather data are short-term.

2.3.1. *Data normalization.* Normalization allows to obtain values with a certain range tailored to the output of the activation function. The aim of data normalization is to reduce data complexity, eliminate duplicated data, and ease in modifying data. In this study, normalization is set with value ranging 0 to 1. To obtain the value of the normalization used the following equation:

\[
Y_n = \frac{X_n - X_{min}}{X_{max} - X_{min}}
\]

Where: 
- \(Y_n\) = Value of data normalization 
- \(X_0\) = The Actual Data 
- \(X_{min}\) = The minimum value of the overall actual data 
- \(X_{max}\) = The maximum value of the overall actual data
2.3.2. Determination of activation function. The type of activation function for forecasting process is bipolar threshold activation function. The modification of the threshold function has three output values, namely -1, 0 or 1. Calculating the value of each node activation function is by setting with random value that outputs the value of -1, 0 or 1.

2.3.3. Determination of probability values, weights, and difference. The next process is probability calculation of each node stored in an array or a list. In addition, the outcome error value by each node and the averaging data value in entire nodes are also determined using the MSE. Both of these values are to update weights in order to obtain forecasting value for the next few days in a row. Weight results saved into a list are able to be stored into a database and to generate forecasting process report.

In determining the difference of each node, actual data value of the previous day is trimmed with current data value. This determination occurs in each category list of parameters being calculated. The process of calculating the difference is then accumulated and is then divided by the number of nodes initiated (in this study, 26 nodes are initiated with 26 locations, also known as id based on the city, in the province of North Sumatra). Thus, averaging the value is essential to be an addition or deduction factor with the initial weight value. Each value determines whether the final weight will be added or subtracted depending on the output value of activation function f(a). If the activation value f(a) is greater than 0 (f(a) > 0), the value will be added by a factor enhancer. Otherwise, if the activation value f(a) is smaller than 0, it will be reduced by adder value. So the final weights are obtained with the following formulas.

2.4. Web service, web server, and client...
The Results of statistical information and charts formerly drawn from the forecast are then compressed into JSON (Java Script Object Notation) and XML files. JSON and XML files are proceeding sent to web service. Next, the web service receives JSON and XML files from the Open Weather Map and sends it to a web server.

Web server receives JSON and XML files which are to be delivered by the Web Service. The data are then processed and made into a web-based application Dashboard which is accessible for clients online. Graphs and charts are displayed according to client request. Clients select the forecast from a dropdown menu that contains the forecast of previous or next few days. After the client chooses request, the Web Service sends the requested data to the web server and displays it. Graphs and charts are steadily updated as the client chooses new option.

Clients are allowed to access web pages on a web server dashboard. This page specifically contains graphs and charts of forecasting results in real time. When the clients select their desired forecasting data, the graphs and charts will be updated and the data obtained is available to be reported as a printable PDF file format.

### 3. Result and Discussion

In the testing process, data used for testing in system is the weather data retrieved from the OpenWeatherMap server and BMKG which has been stored in the database earlier. Forecasting the weather data begins by selecting forecasting the weather menu on the main page. Then on the same page, system requires users to choose amount of forwarding days that forecasting will be performed. Results of forecasting the weather data are shown in table 1.

#### Table 1. Weather forecasting result.

| Day       | City ID | Min. Temperature | Max. Temperature | Min. Humidity | Max. Humidity | Wind Speed |
|-----------|---------|------------------|------------------|---------------|---------------|------------|
| Day 1     | 24      | 23.02            | 32.15            | 50.38         | 92.04         | 20.35      |
| Day 1     | 25      | 23               | 32.3             | 50.76         | 92.08         | 20         |
| Day 1     | 26      | 23.02            | 32.45            | 51.14         | 92.04         | 19.65      |
| Day 1     | 27      | 23.04            | 32.6             | 50.76         | 92.08         | 20         |
| Day 1     | 28      | 23.06            | 32.45            | 50.38         | 92.04         | 19.65      |
| Day 1     | 29      | 23.04            | 32.3             | 50.76         | 92.08         | 20         |
| Day 1     | 30      | 23.02            | 32.45            | 50.38         | 92.12         | 19.65      |
| Day 1     | 31      | 23               | 32.3             | 50.76         | 92.16         | 19.3       |
| Day 1     | 32      | 23.02            | 32.45            | 50.38         | 92.12         | 19.65      |
| Day 1     | 33      | 23.04            | 32.6             | 50            | 92.16         | 19.3       |
| .....     | .....   | .....            | .....            | .....         | .....         | .....      |
| Day 3     | 24      | 22.86            | 33.35            | 48.86         | 92.12         | 21.75      |
| Day 3     | 25      | 22.88            | 33.5             | 49.24         | 92.16         | 22.1       |
| Day 3     | 26      | 22.86            | 33.65            | 48.86         | 92.2          | 22.45      |
| Day 3     | 27      | 22.84            | 33.8             | 49.24         | 92.16         | 22.8       |
| Day 3     | 28      | 22.86            | 33.65            | 49.62         | 92.12         | 22.45      |
| Day 3     | 29      | 22.84            | 33.8             | 50            | 92.16         | 22.8       |
| Day 3     | 30      | 22.86            | 33.65            | 49.62         | 92.2          | 22.45      |
| Day 3     | 31      | 22.84            | 33.8             | 50            | 92.16         | 22.1       |

In table 1, It is shown that the BMA method performs forecasting the weather data which has been recorded into database for three days. Forecasting error is calculated by MSE divergingly at 0.2% of minimum temperature parameter, 0.15% of maximum temperature parameter, 0.38% of minimum humidity parameter, 0.04% of maximum humidity parameter, 0.076% of wind speed parameter. MSE is another method for measuring forecasting error. In this method, each error or residual is squared, then summed or divided by the number of observations. Thus, MSE calculations also show averaged absolute forecast error against its actual data. Error calculation results using the MSE is are provided in table 2.
Table 2. Error results using MSE.

| Node | Min Temp | Max Temp | Min Humidity | Max Humidity | Wind Speed |
|------|----------|----------|--------------|--------------|-----------|
| 1    | 0        | 0        | 0            | 0            | -1        |
| 2    | -5       | -2       | 0            | 3            | -2        |
| 3    | 5        | 0        | 15           | -5           | -2        |
| 4    | 0        | 1        | -10          | 2            | 4         |
| 5    | 0        | 2        | -5           | 0            | 0         |
| 6    | -6       | -4       | 20           | 4            | 0         |
| 7    | 6        | 4        | -20          | -6           | 0         |
| 8    | -4       | -2       | 18           | 4            | -5        |
| 9    | 4        | 2        | -13          | -4           | 5         |
| 10   | 0        | 1        | 0            | 2            | 0         |
| 11   | -5       | -4       | 15           | 4            | 0         |
| 12   | 1        | 0        | -5           | -1           | -6        |
| 13   | 4        | 2        | -10          | -2           | 6         |
| 14   | 0        | 0        | 0            | -2           | -4        |
| 15   | -4       | -2       | 10           | 4            | 4         |
| 16   | 0        | 0        | 5            | 0            | 0         |
| 17   | -1       | 0        | 0            | 0            | 0         |
| 18   | 5        | 3        | -15          | -4           | 1         |
| 19   | 0        | 1        | 0            | 0            | 1         |
| 20   | 0        | 0        | 0            | 1            | -1        |
| 21   | -3       | -3       | 5            | 1            | -2        |
| 22   | 3        | 1        | -5           | -1           | 3         |
| 23   | 0        | 2        | -5           | -2           | -2        |
| 24   | 0        | -1       | 0            | 3            | 2         |
| 25   | 0        | 0        | 5            | -2           | 0         |
| 26   | 0        | -1       | 5            | 2            | 1         |

MSE Error 0.28 0.15 0.38 0.04 0.076

In the real time forecasting chart, parameters of minimum and maximum temperature are shown in figure 5 and figure 6. In actual forecasting chart of minimum and maximum, humidity parameters can be seen in figure 7 and figure 8. Graphic for actual forecasting of wind speed is shown in figure 9.

Figure 5. Real time forecasting chart for minimum temperature

Figure 6. Real time forecasting chart for maximum temperature
4. Conclusion
Based on weather-data forecasting system tested using the BMA, forecasting error values generated by MSE are 0.28 for minimum temperature, maximum temperature 0.15 to 0.38, and 0.04 for minimum humidity in maximum moisture and 0.076 for wind speed. The lower the value of MSE shows, the more optimal the accuracy in forecasting is. Forecasting process with BMA has met the objective function of the problem which is to help the process of forecasting in period of 3 to 7 days.

In the future study, we would like to do forecasting using our classification engine such as Distributed Adaptive Neural Network [10-12] and improved by using fast learning algorithm [13].

5. References
[1] Cahyono W E 2010 Pengaruh pemanasan global terhadap lingkungan bumi Jurnal Sains Dirgantara 8(2) 28-31
[2] Kresnawan A 2008 Penerapan model jaringan saraf tiruan untuk memprediksi gangguan cuaca maritim Tugas Akhir Jurusan Teknik Fisika – FTI - ITS. Surabaya
[3] Dewi C, Kartikasari D P and Mursityo Y T 2014 Prediksi cuaca pada data time series menggunakan adaptive neuro fuzzy inference system (ANFIS) Jurnal Teknologi Informasi dan Ilmu Komputer 1(1): 18-24
[4] Sofia D, Kunto and Soenarnatalina 2014 Pemodelan bayesian model averaging (BMA) pada kasus pneumonia pada belita Jurnal biometrika dan kependudukan 3(1): 34-42
[5] Montgomery J and Nyhan B 2010 Bayesian model averaging: theoretical developments and practical applications *Political Analysis* 18(2): 245-270

[6] Sembiring J H 2010 Peramalan curah hujan bulanan di kota medan dengan metode box-jenkins. Skripsi. Universitas Sumatera Utara. R. Nicole, “Title of paper with only first word capitalized,” *J. Name Stand. Abbrev.*, in press

[7] Rahmat R F, Rizki A, Alharthi A F and Budiarto R 2016 4th *Saudi Int. Conf. on Information Technology* (Big Data Analysis) 1–5

[8] Siregar B, Butar-Butar I A, Rahmat R F, Andayani U and Fahmi F 2017 Comparison of exponential smoothing methods in forecasting palm oil real production *J. Phys.: Conf. ser.* 801(1) 12004

[9] Rahmat R F, Syahputra M F and Lydia M S 2016 *International Conf. on Informatics and Computing* 383–38

[10] Rahmat R, Pasha M, Syukur M and Budiarto R 2015 A gene-regulated nested neural network *Int. Arab J. of Information Technology* 12(6) 532–539

[11] Aaron O S, Sitompul and Rahmat R F 2014 *Int. Conf. on Cyber and IT Service Management* 63–68

[12] Pasha M F, Rahmat R F, Budiarto R and Syukur M 2010 A distributed autonomous neuro-gen learning engine and its application to the lattice analysis of cubic structure identification problem *Int. J. of Innovative Computing, Information and Control* 6(3) 1005–1022

[13] Alwajeeh T, Alharthi A F, Rahmat R F and Budiarto R 2017 *IOP Conf. Ser.: Mater. Sci. Eng.* 190 12015