ISSN 2502-5872 (Print)
ISSN 2622-3627 (Elektronik)

Volume 10 Number 3, August 2025, 649-658

BAYESIAN MODEL AVERAGING (BMA) MODELING IN THE ANALYSIS OF FACTORS AFFECTING CHILDHOOD PNEUMONIA CASES IN MEDAN CITY

Iren Salsalina[1*], Hendra Cipta[2]
[1,2]Departemant of mathematics, Faculty of Science and Technology, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
*Correspondence: iren0703212033@uinsu.ac.id

ABSTRACT
Pneumonia is one of the leading causes of death among toddlers worldwide, including in Indonesia. Medan, as a densely populated city, faces significant challenges in managing pneumonia in toddlers, with a still low case detection rate. The high incidence of pneumonia is influenced by various risk factors such as environmental density, air pollution, malnutrition, smoking habits, and lack of public awareness. Therefore, this study aims to identify the most influential factors affecting the incidence of pneumonia in toddlers in Medan City. This study uses the Bayesian Model Averaging (BMA) approach to address model uncertainty and produce more accurate estimates by considering various possible combinations of risk factors. The BMA method is considered superior in the context of complex health data because it can probabilistically combine information from multiple models. The results of this study are expected to provide data-based recommendations for more effective strategies for the prevention and control of pneumonia in toddlers in densely populated urban areas such as Medan City.
Keywords: Pneumonia, Linear Regression, Bayesian Model Averaging (BMA)

How to Cite: Iren, S., & Hendra, C. (2025). Bayesian Model Averaging (BMA) Modeling In The Analysis Of Factor Affecting Childhood Pneumonia Cases In Medan City. *Mathline: Jurnal Matematika dan Pendidikan Matematika, 10*(3), 649-658. http://doi.org/10.31943/mathline.v10i3.998

PRELIMINARY

Pneumonia is the leading cause of death among children under five in the world and in Indonesia. Pneumonia causes more deaths in infants worldwide than the combined total of AIDS, Malaria, and Measles. Every year, it is estimated that more than 2 million toddlers die from pneumonia. In developing countries, 60% of pneumonia cases are caused by bacteria, while in developed countries, the cause of this disease is viruses. The Indonesian government is paying attention to these pneumonia cases (Kemenkes RI, 2023). Data in Indonesia shows that the number of pneumonia cases is higher in the age group of 1-4 years compared to those under one year old (Kemenkes RI, 2019).

As one of the largest cities in North Sumatra, Medan faces a major problem in handling and preventing pneumonia in toddlers. Pneumonia cases in the city of Medan, particularly Mycoplasma pneumonia type, are found in children and toddlers. The Medan

City Health Office has issued a circular to increase the use of masks and the reporting of cases to healthcare facilities. The main cause is a bacterial infection, with a higher risk in densely populated areas. This disease often presents with symptoms of fever, dry cough, and shortness of breath, which can become severe if not promptly treated (Media Indonesia, 2023).

Pneumonia cases will be more effectively reduced by identifying the factors that influence pneumonia, allowing for various actions and preventive measures to be taken against those influencing factors. Several risk factors that increase the risk of pneumonia infection include advanced age, toddlers, smoking habits, environmental exposure, malnutrition, previous pneumonia history, chronic bronchitis, asthma, functional disorders, poor oral hygiene, and the use of acid secretion inhibitors (Ministry of Health, 2022).

Research by Merlinda Permata Sari on pneumonia using the Autoregressive Integrated Moving Average (ARIMA) method, the results of this research using this method are based solely on the pattern of pneumonia patient data in the city of Semarang recorded by the Semarang City Health Office with ARIMA data results (Sari, 2019). Research by Budi & Eva using the Convolutional Neural Network (CNN) method provides important information on the performance of pneumonia classification (Nugroho & Puspaningrum, 2021). Research by Andi & Siska using the K-Nearest Neighbor (KNN) method analyzes performance metrics (accuracy, precision, recall, and f-measure) in the classification process of pneumonia and non-pneumonia image datasets (Halim & Anraeni, 2021).

The Bayesian model averaging (BMA) method was chosen because it has the advantage of handling model uncertainty and producing more accurate estimates by considering various possible models. This approach is relevant in cases of complex health data such as childhood pneumonia, because Bayesian Model Averaging (BMA) combines the results of several models using posterior probabilities. In the national context, BMA has been used in spatial analysis studies on economic data, demonstrating the method's ability to enhance the accuracy and reliability of analysis results (Sarimah et al., 2019).

The research gap compared to previous studies in this research is:

Table 1. Research Gap In The Study

Researcher and Year	Methods	Research Focus	Gap
Sari (2019)	ARIMA	Prediction of pneumonia cases in Semarang city.	Only modeling case trends without analyzing the risk factors influencing pneumonia.
Nugroho &	Convolutional	Pneumonia image	Focus on classification

Puspaningrum (2021)	Neural Network (CNN)	classification	performance; does not investigate the causative factors or risk factors for pneumonia occurrence.
Halim & Anraeni (2021)	K-Nearest Neighbor (KNN)	Analysis of the classification performance of the pneumonia image dataset.	Failure to identify risk factors; failure to consider model uncertainty in the analysis.
This research	Bayesian Model Averaging (BMA)	Analysis of dominant risk factors for pneumonia in children under five in Medan.	Filling the gap with a comprehensive risk factor analysis in Medan City using the accurate BMA method.

METHODS

The type of research used is quantitative research. The research was conducted using secondary data from the Medan City Health Office. The research was conducted in April – May 2025. The unit of analysis in this study is 41 community health centers in Medan City in 2024.

The response variable in this study is the percentage of pneumonia in children under five (Y), while the predictor variables are the percentage of exclusive breastfeeding (X1), percentage of receiving early initiation of breastfeeding (X2), percentage of children under five receiving vitamin A (X3), percentage of low birth weight infants (X4), percentage of malnourished children under five (X5), number of non-smokers in the house (X6), percentage of infant health services (X7), percentage of child health services (X8), and percentage of clean and healthy living behaviors (X9). The data used are secondary data obtained from the Medan City Health Office in 2024.

Research flowchart:

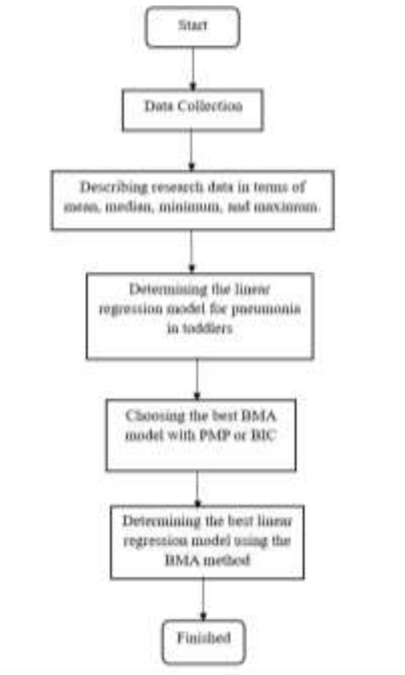

Figure 1. Research Flowchart

RESULT AND DISCUSSION

The percentage data of pneumonia in toddlers and the factors influencing it will be modeled using linear regression to determine the relationship between the response variable and the predictor variable.

The F-test is conducted to determine the joint effect of the predictor variables on the response variable or to assess the suitability of the model. The results show that the predictor variables collectively influence the response variable. This can be seen through the calculated F value (4.636) > table F value (1.30). The value of 0.574% and only 0.254% are influenced by other variables outside the predictor variables in the study. The large coefficient of determination (R2) value, while the regression model is not significant in the model fit test, may be caused by the failure to meet the linear regression assumptions, including the presence of multicollinearity. Table 2 presents the results of the multicollinearity test for each variable.

Table 2. Multicollinearity Test Results

Variable predictor	VIF
X1	1.437
X2	6.402
X3	7.640
X4	1.338
X5	1.324
X6	8.194
X7	150.281
X8	136.419
X9	7.445

Partial or individual tests are conducted to see the contribution of each predictor variable to the response variable using the t-test. Partial tests are conducted if at least one variable affects the response variable, but the research results showed that no variable affects the response variable, so partial tests should not have been conducted. Table 3 presents the results of the partial test.

Table 3. Partial Test Results

Variable Predictor	Koefisien	Standart Error (SE)	P *value*
constant	-0,729	0,562	0,204
X1	-0,030	0,106	0,781
X2	1,180	0,488	0,022*
X3	-1,387	0,472	0,006*
X4	0,104	0,060	0,094
X5	-0,126	0,084	0,144
X6	1,096	0,254	0,000*
X7	-7,984	2,061	0,001*
X8	8,518	1,914	0,000*
X9	-0,082	0,267	0,761

Note: *indicates a significant variable

Table 3 shows that there are 5 significant predictor variables because the p-value < α 0.05. The variables are the percentage of receiving IMD (X2), the percentage of toddlers receiving Vit A (X3), the percentage of non-smoking in the house (X6), the percentage of infant health services (X7), and the percentage of toddler health services (X8). The value of 0.574% and only 0.254% are influenced by other variables outside the predictor variables in the study.

The obtained linear regression model is: Percentage of pneumonia in toddlers = -0.729 - 1.180 percentage receiving IMD - 1.387 percentage of toddlers receiving Vit A + 1.096 percentage of non-smoking in the house - 7.984 percentage of infant healthcare services + 8.518 percentage of toddler healthcare services.

Model Bayesian Model Averaging (BMA)

This method aims to identify the factor that most influences pneumonia in toddlers among the 9 variables in this study using the BMA method. Table 4 presents the posterior probabilities, parameter coefficients, and standard errors of the predictor variables.

Table 4. posterior probabilities, parameter coefficients, and standard errors of the predictor variables

Variable Predictor	Probabilitas Posterior (%)	Koefesien	Standard Error
constant	100	-0,72843	0,56162
X1	0,782685	-0,02941	0,10570
X2	0,021756*	1,17967	0,48820
X3	0,006218**	-1,38689	0,47240
X4	0,094351	0,10424	0,06040
X5	0,144552	-0,12534	0,08373

X6	0,000150***	1,09682	0,25407
X7	0,000516***	-7,98844	2,06151
X8	0,000103***	8,52096	1,91391
X9	0,760510	-0,08222	0,26738

Note: *indicates a significant variable

The selection of the best BMA model is based on the highest Posterior Model Probability (PMP) value with the smallest BIC. Table 5 presents the best model generated by BMA.

Table 5. Best Model BMA

Variable	P!=0	Model 1	Model 2	Model 3	Model 4	Model 5	Average Estimate
constant	100,0	-0,7057	-0,7530	-0,5860	-0,6573	-0,2457	-0,58954
X1	10,6	-	-	-	-	-	-0,001882
X2	66,5	0,9700	1,1042	0,9500	1,1107	-	0,681977
X3	86,4	-1,4175	-1,4205	-1,3180	-1,3343	-0,8840	-1,077957
X4	54,8	0,1206	0,1074	-	-	0,1177	0,062670
X5	35,8	-	-0,1248	-	-0,1471	-	-0,046083
X6	100,0	1,0480	1,0341	1,1484	1,1191	1,0879	1,105504
X7	100,0	-8,5385	-7,8773	-7,8016	-7,1169	-8,0974	-7,988996
X8	100,0	9,0970	8,4755	8,2518	7,6277	8,8645	8,484428
X9	11,2	-	-	-	-	-	-0,008851
Probabilitas posterior		0,179	0,116	0,099	0,095	0,069	
BIC		-64,7589	-63,9002	-63,5825	-63,5039	-62,8681	

Table 5 shows that there are three variables with an inclusion probability of 100%, namely:

1. X6 (percentage of non-smoking in the house),
2. X7 (percentage of infant health services), and
3. X8 (percentage of toddler health services).

This indicates that these three variables consistently appear in all the best models selected by BMA, thus it can be concluded that all three are very important predictors in determining the level of pneumonia in children.

Additionally:

1. X3 (percentage of toddlers receiving vitamin A) has an inclusion probability of 86,4%,
2. X2 (percentage receiving IMD) of 66,5%, and
3. X4 (percentage of low birth weight infants) of 54,8%.

BMA Modeling Results

Searching in the BMA regression model:

$$\hat{Y}_{BMA} = \beta_0 + \beta_1 X_2 + \beta_2 X_3 + \beta_3 X_4 + \beta_4 X_6 + \beta_5 X_7 + \beta_6 X_8$$

$$\hat{Y} = -0,58954 + 0,681977 \cdot X_2 - 1,077957 \cdot X_3 + 0,062670 \cdot X_4 + 1,105504 \cdot X_6 - 7,988996 \cdot X_7 + 8,484428 \cdot X_8$$

Table 6. Significant Variable

Predictor Variable	Value(%)
X_2 (IMD)	66%
X_3 (Vit A)	86%
X_4 (BBLR)	54%
X_6 (No Smoking)	100%
X_7 (Baby Services)	100%
X_8 (Child Care Services)	100%

Based on the posterior percentage values already listed in the table, this study uses significant variables in the BMA regression modeling, specifically 6 significant variables found in Table 6.

$$\hat{Y} = -0,58954 + 0,681977 \cdot X_2 - 1,077957 \cdot X_3 + 0,062670 \cdot X_4 + 1,105504 \cdot X_6 - 7,988996 \cdot X_7 + 8,484428 \cdot X_8$$

In addition, the variables of IMD administration (X2), Vitamin A administration (X3), Low Birth Weight (X4), No smoking in the house (X6), Infant Health Services (X7), and Toddler Health Services (X8) were found to have the most significant contributions in the models with the highest PMP, thus can be considered as the main factors influencing the increase in cases of pneumonia in toddlers.

CONCLUSION

The conclusions that can be drawn from this study are as follows:

1. The Bayesian Model Averaging (BMA) method was successfully applied to build a regression model in analyzing the factors affecting the incidence of pneumonia in toddlers. BMA allows for the probabilistic selection of the best model by considering the uncertainty in model selection.
2. Based on the BMA model estimation results, it was found that the significant factors affecting pneumonia cases in toddlers in Medan City are the IMD provision variable (X2), Vitamin A provision (X3), Low Birth Weight (BBLR) (X4), No smoking indoors (X6), Baby Health Services (X7), and Toddler Health Services (X8).
3. When compared to conventional statistical methods such as multiple linear regression, the model produced by BMA shows better accuracy. Proven by the significant variables in linear regression, there are 5 significant predictor variables, namely the percentage of receiving IMD (X2), the percentage of toddlers receiving Vit A (X3), the percentage of not smoking indoors (X6), the percentage of infant health services (X7), and the

percentage of toddler health services (X8). Meanwhile, in BMA, there are 6 significant predictor variables, namely the IMD provision variable (X2), Vit A provision variable (X3), LBW variable (X4), not smoking indoors variable (X6), infant health services variable (X7), and toddler health services variable (X8). This indicates that BMA is more effective in capturing the complexity of the relationships between variables in cases of pediatric pneumonia.

Advice :

a. Further research should use a larger sample size.
b. Further research should use fewer research variables to focus solely on direct causes.
c. Further research should complete the data and avoid missing data in the research variables.
d. Future research should avoid using factors with nearly identical meanings as predictor variables.

REFERENCES

Artanto, Noviesag., & Agoestanto, Arief. 2015. *Perbandingan Peramalan Curah Hujan Dengan Metode Bayesian Model Averaging Dan Kalman Filter*. *UNNES Journal of Mathematics,* UJM 4 (1) (2015).

Cher, Nobella, Asyunita. 2019. *Penerapan Bayesian Model Averaging Pada Regresi Linear Dalam Menentukan Faktor-Faktor Yang Mempengaruhi Angka Kematian Ibu Di Provinsi Jawa Tengah.* Skripsi. Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret. https://digilib.uns.ac.id/

Halim, Dzariah, Ainun, Andi., & Anraeni, Siska. 2021. *Analisis Klasifikasi Dataset Citra Penyakit Pneumonia Menggunakan Metode K-Nearest Neighbor (KNN). Indonesian Journal of Data and Science (IJODAS),* Vol 2, No 1, March 2021, pp. 01-12. https://doi.org/10.33096/ijodas.v2i1.23

Islamiyah, Ilma, nurul. 2018. *Pemodelan Generalized Poisson Regression Pada Faktor-Faktor Yang Mempengaruhi Kasus Pneumonia Pada Balita Di Provinsi Sulawesi Selatan Tahun 2018.* Skripsi. Jurusan Matematika pada Fakultas Sains dan Teknologi Universitas Islam Negeri Alauddin Makassar. https://repositori.uin-alauddin.ac.id/xxxx

Muharsyah, Robi., & Ratri, Nur, Dian. 2021. *Meningkatkan Kualitas Prediksi Curah Hujan Musiman saat Fase ENSO Menggunakan Metode Bayesian Model Averaging (BMA), Studi Kasus: Pulau Jawa. Jurnal Tanah dan Iklim,* Vol. 45 No. 1, Juli 2021: 13-25. https://doi.org/10.21082/jti.v45n1.2021.13-25

Muharsyah, Robi., Hadi, Wahyu, Tri., & Indratno, Wahyu, Sapto. 2020. *Penerapan Metode Bayesian Model Averaging untuk Kalibrasi Prediksi Ensemble Curah Hujan Bulanan di Pulau Jawa. Journal Ipb ,* Agromet 34 (1): 20-33, 2020. https://doi.org/10.29244/j.agromet.34.1.20-33

Nugroho, Budi., & Puspaningrum, Yulia, Eva. 2021. *Kinerja Metode Cnn Untuk Klasifikasi Pneumonia Dengan Variasi Ukuran Citra Input*. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, Vol. 8, No. 3, Juni 2021, hlm. 533-538. https://doi.org/10.25126/jtiik.2021834515

Paramita., Annas, Suwardi., & Aidid, Muhammad, Kasim. 2020. *Aplikasi Metode Bayesian Model Averaging (BMA) dengan Pendekatan Markov Chain Monte Carlo (MCMC) untuk Peramalan Curah Hujan di Stasiun Meteorologi Kota Makassar*. *Journal of Statistics and Its Application on Teaching and Research*, Vol. 2 No. 3 (2020), 147-157. *https://doi.org/10.35580/variansiunm14642*

Pawestri, Vincentia., Setiawan, Adi., Linawati, Lilik. 2019. *Pemodelan Data Penjualan Mobil Menggunakan Model Autoregressive Moving Average Berdasarkan Metode Bayesian*. *Jurnal Sains dan Edukasi Sains*, Vol.2, No.1, Februari 2019: 26-35. https://doi.org/10.24246/juses.v2i1p26-35

Putera, Setiarno, Luthfi, Muhammad. 2017. *Bayesian Model Averaging Dan Geostatistical Output Perturbation Untuk Prakiraan Cuaca Jangka Pendek Terkalibrasi*. Tesis. Program Magister Departemen Statistika Fakultas Matematika Dan Ilmu Pengetahuan Alam Institut Teknologi Sepuluh Nopember Surabaya.

Rahmat, R, F., Nasution, F, R., Seniman., Syahputra, M, F., Sitompul O, S. 2018. *Implementation of bayesian model averaging on the weather data forecasting applications utilizing open weather map*. *Materials Science and Engineering*, 309 (2018) 012054. *https://doi.org/10.1088/1757-899X/309/1/012054*

Sari, Permata, Melinda. 2019. *Penggunaan Metode Arima (Autoregressive Integrated Moving Average) Untuk Prakiraan Penderita Pneumonia Balita Di Kota Semarang Tahun 2019-2021*. Jurusan Ilmu Kesehatan Masyarakat Fakultas Ilmu Keolahragaan Universitas Negeri Semarang. https://repository.unnes.ac.id/xxxx

Sarimah., Djuraidah, Anik., & Wigena, Hamim, Aji. 2019. *Pemodelan Autoregresif Spasial Menggunakan Bayesian Model Averaging Untuk Data Pdrb Jawa*. *Indonesian Journal of Statistics and Its Applications*, Vol 3 No 3 (2019), 287–294. file:sarimahpaper_31e98183-b9fc-42a8-bcb8-7069dec3dd60%20(2).pdf..

Trisnawati, Y. 2012. Hubungan perilaku merokok Oramg Tua dengan Kejadian ISPA pada Balita Di Wilayah Kerja Puskesmas Rembang Kabupaten Purbalingga. Jurnal Kesmas Indonesia, 6(1), 35–42.

Veridiana, Nyoman, Ni., Octaviani., & Nurjana, Agus, Made. 2021. *Faktor Internal dan Eksternal Kejadian Pneumonia pada Anak Bawah Dua Tahun di Indonesia*. *Buletin Penelitian Kesehatan*, Vol. 49, No. 3, September 2021: 145-154. https://doi.org/10.22435/bpk.v49i3.4802

World Health Organization. (2008). *The world health report 2008: Primary health care – now more than ever*. Geneva: WHO Press. https://iris.who.int/handle/10665/43949

Winiarti, S. 2008. Pemanfaatan Teorema Bayes dalam Penentuan Penyakit THT. Jurnal Informatika Vol. 2 hal. 2.

Zeugner, S. 2011. Bayesian Model Averaging with BMS. http://cran.r project.org/web/packages/BMS/vignettes/bms.pdf. https://bms.zeugner.eu/tutorials/bms.pdf