JMSCOWA 6 (1) (2025) Pages:. 12-16

Journal homepage: https://pcijournal.org/index.php/jmscowa

Journal of Mathematics and Scientific Computing with Applications

Published: Pena Cendekia Insani

# CLASSIFICATION OF BASIC INFANT IMMUNIZATION MOTIVATION USING BINARY LOGISTIC REGRESSION AND SUPPORT VECTOR MACHINE (SVM)

**Sari Fathul Jannah[1], Faisal Muhammad[2]**

[1]Department of Mathematics, Universitas Negeri Makassar, Makassar, Indonesia

[2]Department of Applied Mathematics, IPB University, Bogor, Indonesia

**Article Info**

***Article history:***

Received 08 13, 2024
Revised 11 30, 2024
Accepted 01 5, 2025

***Keywords:***

Immunization Motivation, Binary Logistic Regression, Support Vector Machine (SVM), Accuracy

**ABSTRACT**

Immunization is an effort to actively increase a person's immunity against a particular disease so that if one day they are exposed to it, they will not get sick or only experience mild illness. Then motivation is a driving force for carrying out certain actions, where the level of motivation can determine the level of results obtained. We conducted this research to identify the factors that influence motivation and determine the optimal classification accuracy. The data used is primary data by distributing questionnaires to mothers who have babies under one year old in Simatahari Village, Kotapinang Labuhan Batu Selatan District, with the response variable being motivation for basic baby immunization, where (1) is interested and (0) is not interested. The predictor variables used are education, employment, attitudes, knowledge, family support, and health services. The method used to classify basic infant immunization motivation is the Support Vector Machine (SVM) method and will be compared with the Binary Logistic Regression method. Based on the results of research using the Binary Logistic Regression method, it indicates that the variables that influence motivation for basic infant immunization are knowledge and health services with a classification accuracy of 83.3%. While the classification accuracy using SVM was the highest, the classification accuracy using the RBF kernel reached 88.33%.



***Corresponding Author:***

Sari Fathul Jannah,
Department Of Mathematic,
Universitas Negeri Makassar
Email: sarifathuljannah@gmail.com

## 1. INTRODUCTION

Immunization is an effort to actively increase a person's immunity against a particular disease so that if one day they are exposed to it, they will not get sick or only experience mild illness. Immunization can prevent various diseases that claim the lives of more than 1.4 million children worldwide every year. Immunization is one of the most effective and widely used efforts to prevent child deaths.

There are various factors that influence immunization, such as knowledge, attitudes, actions, social culture, side effects , and family support (Astrida, 2019). The mother's characteristics, including her education and attitude, will determine the completeness of immunization. mother's perception. Maternal characteristics are factors that impact the completeness of immunization (Hartati et al., 2019). Parents often choose not to complete the immunization process due to concerns about potential side effects (Nur Aziah & Mifbakhudin, 2015). Fever and swelling from injection sites are the most common complaints. Ordinary

people are more worried about the side effects of immunization than the disease itself and its complications, which can cause disability and death (PA Siregar, 2020).

Where South Labuhanbatu district is one of the districts that has just been expanded from Labuhanbatu district, this means that South Labuhanbatu district still does not have complete health facilities like other districts that have been independent for a long time. It can be seen from the fact that there are still many incomplete facilities and infrastructure, including in the health sector, so it does not rule out the possibility that health facilities and health personnel are still incomplete as in other districts (Labuhan Batu Selatan, 2020). Of the five sub-districts, the first position is the Kotapinang sub-district, where there are 7 cases of dengue fever and 78 cases of TB (tuberculosis).

The analysis used in this research uses binary logistic regression and support vector machine (SVM). The binary logistic regression model is used to analyze the relationship between one response variable and several predictor variables, while "SVM is a technique for finding hyperplanes that can separate two sets of data from two different classes. SVM has advantages, including determining distances using support vectors so that the computing process is rapid (Kasmawati & Bekti, 2019).

## 2. RESEARCH METHOD

The data used in this research is primary data, obtained from mother respondents who have babies under 1 year old in Simatari village, Kotapinang subdistrict, by filling in a questionnaire that has been distributed to the research sample. A questionnaire serves as a tool for data collection, incorporating questions that align with the objectives of the relevant theory.

This research uses a predictor variable (independent variable) X and a response variable (dependent variable) Y, as follows:

1. The predictor variable used is:
   $X_1$ = education
   0 = low; 1 = tall
   $X_2$ = work
   0 = doesn't work; 1 = work
   $X_3$ = attitude
   0 = not good; 1 = good
   $X_4$ = knowledge
   0 = not good; 1 = good
   $X_5$ = dukungan keluarga
   0 = does not support; 1 = support
   $X_6$ = health services
   0 = not good; 1 = good
2. The response variable used is basic infant immunization motivation with two categories (classes) Y, namely: interest in immunization (1) and no interest in immunization (0).

### Analysis Steps

The analysis steps in this research are as follows:

1. Determine the response variables and predictor variables.
2. Create a research plan that includes place, time and location, data sources, population, number of research samples, and sample collection methods.
3. Divide the data into random training and testing data with the help of RapidMiner Studio 9.10 software.
4. Modeling using binary logistic regression to find out the factors that influence the baby's basic immunization motivation.
   a. Create an initial model
   b. Carrying out a likelihood ratio test to find out whether the predictor variables contained in the model have a real effect or not as a whole
   c. Carry out a Wald test to find out which predictor variables have a strong relationship with the response variable
   d. Carry out model suitability tests
   e. Determine the final model
   f. Calculating the accuracy of binary logistic regression classification for testing data
5. Carry out classifications that influence motivation for basic infant immunization using the Support Vector Machine (SVM) method. The modern classification method Support Vector Machine (SVM) was first introduced by Vapnik in 1992 and presented at the Annual Workshop on Computational Learning Theory. SVM is a machine learning method that works on the principle of Structural Risk Minimization (SRM) with the aim of finding an inverse hyperplane that separates two classes in the input (Novianti &

Purnami, 2012). In research (Octaviani et al., 2014), to get the best separating function, look for a separating function that is located in the middle between two class boundary planes, and getting the best separating function is the same as maximizing the margin or distance between two sets of objects from different classes. The following is the iSVM method algorithm:

a. Carry out data transformation according to the iSVM software format, namely RapidMiner Studio 9.10.
b. Determining the modeling kernel function.
   The kernel function provides various conveniences, because in the SVM learning process to determine the support vector, we only need to know the kernel function used, and do not need to know the form of the non-linear function ϕ. According to Hibe (2009), the kernel functions that are usually used in the SVM literature are as follows:
   Linear: $K(x, y) = x.y$ (1)
   Polynomial: $K(x, y) = (x.y + c)^d$ where $d = 1, \ldots$ (2)
   Raldial Basis Function (RBF): $K(x, y) = exp(-\gamma\|x - y\|^2)$, for $\gamma = \frac{1}{2\sigma^2}$ (3)
c. Determine the values of kernel parameters and cost parameters for optimization. Where, $x$ and $y$ are pairs of two data from all parts of the training data. Parameters $\sigma, c, d > 0$, are constants (Damanik et all., 2015).
d. Determining the best parameter values for optimizing training data for testing data classification
e. Calculating classification accuracy for testing data

6. Compare the classification accuracy obtained from binary logistic regression with SVM.
7. Conclusion.

## 3. RESULT AND ANALYSIS

### 3.1 Data Collection

The data analyzed in this research is data from mothers who have babies under one year old who participated or did not participate in immunization in Simatahari Village, Kotapinang District, South Labuhanbatu Regency. The data is in the form of questionnaires distributed to 150 respondents, including Padangrie HTI Hamlet, Sejahtera Hamlet, Sentosa Hamlet, Perjuangan Hamlet, Makmur Hamlet, Bakti Hamlet, and Aek Hije Hamlet.

### 3.2 Binary Logistic Regression Method

Testing the Significance of Parameters Simultaneously and Partially The purpose of this test is to determine the significant impact of the used parameters on the model. Namely, the Likelihood Ratio test, where the initial step is taken to find out whether the independent variables contained in the model have a real influence or not as a whole. After testing partial significance with the likelihood ratio, the next step is the Wald test. This was done to find out which predictor variables significantly influence motivation for basic immunization in babies in Simatari village. Finally, we use the model suitability test to determine the appropriateness of the resulting model. In calculating the classification accuracy value, 60 testing data are used, where we divide the training data into 60% (90) and testing data into 40% (60). The following are the results of the accuracy of binary logistic regression testing data classification:

Table 1. Binary Logistic Regression Classification Results

| Observation | Prediction | | Total |
|---|---|---|---|
| | Not interested | Interested | |
| Not interested | 50 | 0 | 50 |
| interested | 10 | 0 | 10 |
| Total | 60 | 0 | 60 |

$$Accuracy = \frac{0 + 50}{0 + 50 + 10 + 0} \times 100\% = 83{,}3\%$$

$$Error = \frac{10 + 0}{0 + 50 + 10 + 0} \times 100\% = 16{,}7\%$$

Based on the calculations above, it shows the overall classification accuracy results from predictions using 90 training data and 60 testing data, resulting in a classification accuracy of 83.3%.

### 3.3 Support Vector Machine (SVM)

The classification of data on the motivation for basic infant immunization uses the SVM method, which in this study employs the Radial Basis Function (RBF) kernel function. The model created from the training data is used to classify the testing data, which consists of 60 samples, with 60% (90) used for training and 40% (60) for testing. The results of this classification will show how accurate the classification model obtained is.

Table 2. Support Vector Malchine Classification Results

| Observation | Prediction | | Total |
|---|---|---|---|
| | Not interested | Interested | |
| Not interested | 49 | 6 | 55 |
| interested | 1 | 4 | 5 |
| Total | 50 | 10 | 60 |

$$Accuracy = \frac{4+49}{4+49+1+6} \times 100\% = 88{,}33\%$$

$$Error = \frac{1+6}{4+49+1+6} \times 100\% = 11{,}67\%$$

Based on the calculation above, it shows the overall classification accuracy results from predictions using 90 training data and 60 testing data, resulting in a classification accuracy of 88.33%.

### 3.4 Comparison of Classification Methods of Binary Logistic Regression and Support Vector Machine (SVM)

The following accuracy was obtained by evaluating the classification results in this analysis and comparing the classification accuracy values through several experiments to determine the best classification using binary logistic regression and the support vector machine method:

Table 3. Comparison of Classification of Binary Logistic Regression Methods in SVM

| Data Testing | Binary Logistic Regression | *Suppor Vector Malchine* (SVM) |
|---|---|---|
| 60 | 83.3 | 88,33 |
| 30 | 80.0 | 86.67 |
| 15 | 80.0 | 80.0 |

From the table above, information is obtained that the best classification is using the SVM method with testing data of 60 data, with the RBF kernel function, the classification accuracy is 60%, where the parameter C = 62.7 and the Gamma value (γ) is 0.125.

## 4. CONCLUSION

Based on the results of the analysis of the Basic Infant Immunization Motivation Classification research, which was completed using Binary Logistic Regression and Support Vector Machine (SVM), the results showed that 25 respondents fell into the category of interest in immunization, and 125 respondents fell into the category of no interest in immunization from 150 mother respondents who had babies under one year old in Simatahari Village, South Kotapinang Labuhanbatu District. The conclusion was obtained using the binary logistic regression method, where the variables that influence motivation for basic infant immunization are knowledge and health services, with the best model being:

$$\pi(x) = \frac{e^{-24.774+0.994X_4+0.957X_6}}{1+e^{-24.774+0.994X_4+0.957X_6}}$$

Using training and testing data, namely 90:60, classification accuracy is 83.3%. Meanwhile, the classification accuracy using the SVM method, namely using the Radial Basis Function (RBF) kernel function, the highest classification accuracy with three trials reached 88.33%. Therefore, the Support Vector Machine (SVM) method outperforms binary logistic regression in classification completion.

## 5. REFERENCES


[1] Astrida Budiarti. (2019). Hubungan Faktor Pendidikan, Pekerjaan, Sikap Dan Dukungan Keluarga Terhadap Imunisasi Dasar Di RW 03 Kelurahan Kedung Cowek Kenjeran Surabaya. Jurnal Kesehatan Mesencephalon, 5(2), 53–58.

[2] Damanik, S. M. S., Ispriyanti, D., & Sugito. (2015). Klasifikasi Lama Studi Mahasiswa FSM Universitas Diponegoro Menggunakan Regresi Logistik Biner Dan Support Vector Machine (SVM). Jurnal Gaussian, 4(1), 123–132.

[3] Hartati, I., Irawan, D., & Maulida, Al. (2019). Faktor-Faktor Yang Mempengaruhi Status Imunisasi Dasar Lengkap Pada bayi Usia 0-12 Buan Di Desa Suka Mulia Kecamatan Rantau Kabupaten Aceh Tamiang. Jurnal Pendidikan Dan Praktik Kesehatan, 2(1), 41–53.

[4] Hibe, J. M. (2009). Logistic Regression Models. www.staltisticallalssocialtes.com

[5] Kasmawati, & Bekti, R. D. (2019). Klasifikasi Status Human Immunodeficiency Virus (HIV) Menggunakan Metode Support Vector Machine ( SVM ) Dan Regresi Logistik Biner ( Studi Kalsus : Rumah Sakit Tiom Kabupaten Lanny Jaya Provinsi Papua ) Kasmawati , Rokhana Dwi Bekti Terdapat Beberapa Mas. Jurnal Statistik Industri Dan Komputasi, 4(2), 65–73.

[6] Labuhan Batu Selatan. (2020). Kecamatan Kotapinang Dalam Angka 2020. In Journll Of Chemical Information And Modeling.

[7] Novianti, F. Al., & Purnami, S. W. (2012). Analisis Diagnosis Pasien Kanker Payudara Menggunakan Regresi Logistik Dan Support Vector Machine (SVM) Berdasarkan Hasil Mamografi. Jurnal Sains Dan Seni ITS, 1(1), 147–152.

[8] Nur Aziza, Mifbakhuddin, L. M. (2015). Faktor-Faktor Yang Berhubungan Dengan Kelengkapan Imunisasi Dasar Pada Bayi Usia 9 - 11 Bulan Di Desa Sumberejo Kecamatan Mranggen Demak. Jurnal UNIMUS, 1–7.

[9] Octaviani, P. Al., Wilandari, Y., & Ispiriyanti, D. (2014). Penerapan Metode SVM Pada Data Akreditasi Sekolah Dasar Di Kabupaten Magelang. Jurnal Gaussian, 3(4), 811–820. [10] Supranto, Johanes. 2004. Analisis Multivariat Arti Dan Interpretasi. Rineka Cipta: Jakarta.

[10] Pa Siregar. (2020). Analisis Faktor Berhubungan Dengan Tindakan Imunisalsi Campak Pada Batita Di Wilayah Kerja Puskesmas Kota Pinang Kabupaten labuhanbatu Selatan.