# Prediction of Birth Rate of Babies at Regional Hospitals in Salatiga City for Future Planning Using the Naïve Bayes Algorithm

**Adelliya Dewi Saputri[1*], Hendry[1]**

[1]Faculty of Technology & Information, Satya Wacana Christian University, Jl. Diponegoro 52-60, Salatiga, Central Java, Indonesia

* adellsz02@gmail.com

**Abstract** . Birth rates have a significant impact on population growth and large populations can be a burden on development. In the Salatiga City Regional Hospital, the numbers tend to change every year, with the current population density making it a special concern for the City of Salatiga. Therefore, it is hoped that the application of Data Mining Techniques with the Naive Bayes algorithm can help predict the number of births in the future using the RapidMiner Application. In this research, the population used was Population Data from Salatiga City with a total of 989,674 residents. Then the sample used was 4699 babies from the Salatiga City Regional Hospital. All data was taken from 2019 – 2023 by conducting observations, literature studies and documentation. By analyzing the pattern of each variable and testing the training data against the testing data, a calculation was produced which shows the Testing Data Prediction, namely the "High" label with the number 4.77192E-06, with this the predicted result of the Baby Birth Rate in the Salatiga City Regional Hospital which is influenced by Population Density in 2024 it will be even higher.

## 1. Introduction

Rapid population development refers to the rapid growth of the human population over a certain period of time. One of the things that can influence population development is the birth rate [21] . The birth rate is something that can influence the increase in population growth [16]. The number of births in Indonesia tends to increase every year. Birth rate, also known as birth rate or fertility rate, refers to the number of births that occur in a population within a certain period of time [ 1 ] [ 17 ].

Salatiga City Regional General Hospital (RSUD) is a government-owned hospital located on Jl. Osamaliki No.19, Mangunsari, Sidomukti District, Salatiga City, Central Java 50721. Optimal population distribution based on a balance of population and power supported by community quality and environmental quality is the government's hope for each region, especially in Salatiga City. The problem is that not all regions experience population balance and the current increase in population needs to be taken into account [14]. This is because the number of babies born at the Salatiga City Regional Hospital is increasing every year [18] .

Data mining is the process of analyzing data from different points of view and forming it into useful

information . In the context of the Salatiga City Hospital, data mining is used to analyze the birth rate of babies [2] . So it can be a reference for overcoming or anticipating population density in the future [15] [20] [22 ] . . The Naive Bayes algorithm is used by the Salatiga City Regional Hospital to manage population data, especially birth rates. By using data for 2019-2023, including spontaneous births, caesareans, and population, this algorithm carries out a training process to group the data into negative clusters with a value of 0 and positive clusters with a value of 1.

The previous research that has been carried out in predicting the birth rate of babies at regional hospitals using the Naïve Bayes algorithm which is used as a reference in this research, includes:

a.  In research conducted by Nur Isnaini Parihah, Sari Hartini, Juarni Siregar (2020) with the title "Prediction of Baby Birth Rates in Tridaya Sakti Village Using the Naïve Bayes Algorithm" regarding the prediction of baby birth rates in Tridaya Sakti Village showed significant results, especially P( X | Description = "High") is 5.63839E-10. The Tridaya Sakti Village Office can use this prediction as a guide for evaluating population data, as well as as a reference for reducing birth rates, such as increasing the socialization of Family Planning which is expected to be implemented to achieve the goal [3].

b.  In research conducted by Fajar Romadhon, Adi Suwondo, Hidayatus Sibyan (2022) with the title "Implementation of the Naïve Bayes Algorithm in the Annual Population Growth Prediction Application in Wonosobo Regency". The prediction results show low population growth because the value (P|low) is greater than (P|high). The population is predicted using the average growth value in the training data which has the same prediction results [4].

c.  In research conducted by Sinta Ayu Sabilla, Banni Satria Andoko, Moch Zawaruddin Abdullah (2021) with the title "Naïve Bayes Method for Predicting Baby Birth (Case Study: Midwife S Clinic)". The birth age prediction results obtained at Midwife Sulikah can provide an accuracy rate of 94%, precision of 93.33%, and recall of 96.55% from 200 training data and 50 test data using the Naïve Bayes algorithm. This success is proven by consistent test results between manual calculations and applications [5].

Based on the research above, the author discusses a topic that is slightly different from several previous studies. The writer will conduct research about predicting the birth rate of babies at the Salatiga City Regional Hospital using the *Naïve algorithm Bayes Classifier* which is based on research Previously, the *Naive Bayes Classifier algorithm* was suitable for predicting the number of births in the coming year using baby birth data from 2019 – 2023 . After carrying out the training data processing process by grouping the data into negative clusters which have a value of 0 or low and positive values of 1 or high, data is produced whose level of prediction numbers increases from year to year. It is hoped that this research will be useful for the Salatiga City Regional Hospital and also to reduce population density in the coming year.

## 2.  Method

This research was conducted at the Salatiga City Regional General Hospital. To predict an event that is likely to occur in the future based on data from the last 5 years and information from the past, a Naïve *Bayes algorithm can be used* [ 6]. This algorithm is a machine learning method that uses probability calculations. The basic concept used by Bayes is Bayes' Theorem [7], namely carrying out classification by calculating probability values. Classification is carried out to determine the category of existing data [9]. An advantage of the Naïve Bayes Classifier is that it will only require a small amount of training data to estimate the parameters required for classification [ 8].

Bayes' theorem is the basic rule of the Naïve Bayes Classifier. The following Bayes' theorem will be presented in the equation.

$$P\ (H \mid X) = \frac{P(X|H)P(H)}{P(X)}$$

Information:
X      = Data sample that has an unknown class (label).
H      = Hypothesis that X is class data (label)
P(H)  = Probability of hypothesis H
P(X)  = Probability of the observed sample data
P(X|H) = Probability of sample data X if it is assumed that the hypothesis is true

Naïve Bayes Formula Numerical Data Type

| Gaussian distribution | Mean | Standard Deviation |
|---|---|---|
| $$\mathcal{g}(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi.\sigma}}exp^{\frac{-(x-\mu)^2}{2\sigma^2}}$$ | $$\mu = \sum_{i}^{n} x_i$$ | $$\sigma = \frac{\sqrt{\sum_{i}^{n}(x_i - \mu)^2}}{n-1}$$ |
| information : <br> $\mathcal{g}$ : Normal Distribution <br><br> x : Variable Value <br><br> μ : Calculated Average (Mean) <br><br> $\sigma$: Standard Deviation | information : <br> μ : Calculated Average <br><br> i : Sample Sequence <br><br> xi : Value of the i-th Sample <br><br> n : Number of Samples | information : <br> σ : Standard Deviation <br><br> xi: Value of X to –I <br><br> μ : Calculated Average <br><br> n : Number of Samples |

In this study, the population used was from Salatiga City Population data of 989,674 residents [10]. Then the sample used in this research was 4699 babies at the Salatiga City Regional Hospital. Meanwhile, the methodology that will be used in this research is shown in Figure 1 where these steps will be carried out during the research process. As for the detailed methods in this research, there are stages that can be seen as follows:
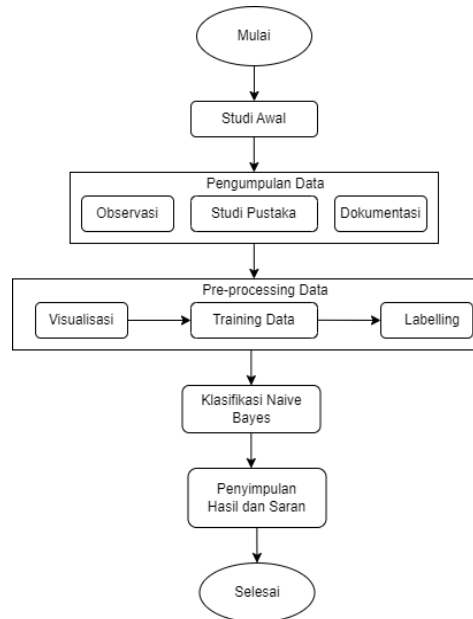
**Figure 1.** Flowchart of Research Method Stages

This research began with a Preliminary Study at the Salatiga City Regional Hospital to identify birth problems [12]. Data was collected through observation, literature study and documentation. Data processing focuses on identifying problems in RSUD, with visualization using scatter plots for the relationship between population density and baby births. The Kmeans method is used to group data into positive clusters with a value of 1 and negative values with a value of 0 [13]. then the Naïve Bayes algorithm can carry out classification using probability and statistical methods, namely predicting future opportunities based on previous experience. The final stage is Conclusion of Results and Suggestions, where the classification results are used to conclude the birth rate of babies in the Salatiga City Regional Hospital [11]. So that later accurate predictions, suggestions and solutions can be submitted to the Salatiga City Regional Hospital to overcome the potential for increasing population density in the coming year.

## 3. Results and Discussion

### 3.1 Data analysis

Naïve Bayes can carry out classification using probability and statistical methods, namely predicting future opportunities based on previous experience. In analyzing Salatiga City Population data using the Naïve Algorithm Bayes begins by determining the training data. After that, the author determines the test data and then uses the Naive Bayes Algorithm to calculate predictions or possibilities.

**Table 1.** Training Data

| Year | Population density | Number of Baby Births | Clusters |
|------|-------------------|----------------------|----------|
| 2019 | 195563 | 1448 | Low |
| 2020 | 196082 | 1081 | Low |
| 2021 | 196440 | 662 | Tall |
| 2022 | 200220 | 683 | Tall |
| 2023 | 201369 | 825 | Tall |

After determining the data that will be used as training data, then determine the data that will be used as testing data as shown in the following table.

**Table 2.** Testing Data

| Year | Population density | Number of Baby Births | Clusters |
|------|--------------------|-----------------------|----------|
| 2024 | 202659 | 447 | ? |

## 3.2 Calculations using the Naïve Bayes method

1. Calculating the P(Ci) Value of Class (label) Description

   **P(Ci)**
   P( Description = Height) 3/5 = 0.6
   P( Description = Low) 2/5 = 0.4

2. Calculating the Mean Value

   **Mean**

$$\mu = \sum_{i}^{n} x_i$$

   Variable: Population, Description: Height

$$\mu = \frac{196440 + 200220 + 201369}{3} = 199343$$

   Variable: Population, Description: Low

$$\mu = \frac{195563 + 196082}{2} = 195823$$

   Variable: Number of Baby Births, Information: Height

$$\mu = \frac{662 + 683 + 825}{3} = 723,3$$

   Variable: Number of Baby Births, Description: Low

$$\mu = \frac{1448 + 1081}{2} = 1114,5$$

3. Calculating Standard Deviation Values
   **Standard Deviation**

$$\sigma = \sqrt{\dfrac{\sum_i^n (x_i - \mu)^2}{n - 1}}$$

Variable: Population, Description: Height

$$\sigma = \sqrt{\dfrac{(196440 - 119343)^2 + (200220 - 119343)^2 + (201369 - 119343)^2}{3 - 1}} = 66347$$

Variable: Population, Description: Low

$$\sigma = \sqrt{\dfrac{(195563 - 195823)^2 + (196082 - 195823)^2}{2 - 1}} = 67341$$

Variable: Number of Baby Births , Information: Height

$$\sigma = \sqrt{\dfrac{(662 - 723,3)^2 + (683 - 723,3)^2 + (825 - 723,3)^2}{3 - 1}} = 6014,14$$

Variable: Number of Baby Births , Description: Low

$$\sigma = \sqrt{\dfrac{(1448 - 1114,5)^2 + (1081 - 1114,5)^2}{2 - 1}} = 335,18$$

1. Calculate the P Value ( XK | Ci) for each class i

**P( X K | Ci )**

$$g(x, \mu, \sigma) = \dfrac{1}{\sqrt{2\pi.\sigma}} exp^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Variable: Population, Description: Height

$$g(x, \mu, \sigma) = \dfrac{1}{\sqrt{2.3,14.66347}} exp^{\frac{-(202659-199343)^2}{2.\,66347^2}} = 0,001547274$$

Variable: Population, Description: Low

$$g(x, \mu, \sigma) = \dfrac{1}{\sqrt{2.3,14.67341}} exp^{\frac{-(202659-195823)^2}{2.\,67341^2}} = 0,001529828$$

Variable: Number of Baby Births , Information: Height

$$g(x,\mu,\sigma) = \frac{1}{\sqrt{2 \times 3{,}14 \times 6014{,}14}} exp^{\frac{-(447-723{,}3)^2}{2 \times 6014{,}14^2}} = 0{,}005140142$$

Variable : Number of Baby Births , Description: Low

$$g(x,\mu,\sigma) = \frac{1}{\sqrt{2 \times 3{,}14 \times 335{,}18}} exp^{\frac{-(447-1114{,}5)^2}{2 \times 335{,}18^2}} = 0{,}00300046$$

2. Calculating the P(X|Ci) Value for each class (label)

**P(X|Ci)**

P (X | Description = "High" )

=0,001547274 × 0,005140142 = 7.95321E-06

P (X | Description = "Low" )

=0,001529828 × 0,00300046 = 4.59019E-06

3. Calculating the Value of P(X| Ci)* P(Ci)

P (X | Description = "Height" ) * P (Description = "Height" )

= 7.95321E-06 ×0.6 = 4 .77192E-06

P (X | Description = "Low" ) * P (Description = "Low" )
= 4.59019E-06 × 0.4 = 1.83608E-06

4. Determine the Class Results of the cases that have been calculated the final calculation by multiplying the probability values of this case, the author sees that the value of P ( "Low" 1.83608E-06 then the description class (label) is **"High"** .

**Table 3.** Results of Naïve Bayes calculations

| Year | Total Population Density | Number of Baby Births | Information |
|------|--------------------------|-----------------------|-------------|
| 2024 | 202659 | 447 | Tall |

5. Prediction Results using the *RapidMiner Application RapidMiner* Application calculations show the Testing Data Prediction, namely the information label "High", meaning that the predicted results of the Baby Birth Rate at the Salatiga City Regional General Hospital which is influenced by Population Density in 2024 will be higher.

**Table 4.** Testing Data Prediction Results from the *RapidMiner Application*

| No. | Predictions (Information) | Confidence (Tall) | Confidence (Low) | Year | Number Birth Baby | Amount Density Resident |
|-----|---------------------------|-------------------|------------------|------|-------------------|-------------------------|
| 1 | Tall | 100.00% | 0.00% | 2024 | 447 | 202659 |

In this study, the prediction level for the birth rate of babies in the Salatiga City Hospital in 2024 was obtained by calculating the probability from the dataset using the Naive Bayes algorithm to produce a description label P( X | Description = "High") = 4.77192E-06. This means that the predicted results of the baby birth rate in the Salatiga City Regional Hospital which is influenced by population density in 2024 will be higher. This is in line with the journal [ 3] that the results of the prediction of the birth rate of babies in Tridaya Sakti Village in 2020 experienced a high birth rate of 5.63839E-10 using the Naive Bayes Algorithm Method so that by making this prediction, steps can be taken to take action to reduce the birth rate of babies in Tridaya Sakti Village. This is also supported by the journal statement [ 5] which discusses the prediction of birth age at Midwife Sulikah provides an accuracy rate of more than 85%. Namely, with an accuracy level of 94 %, precision of 93.33% and recall of 96.55% from 200 training data and 50 test data. This test uses test data with the aim of determining the level of precision, recall and accuracy in the Naïve Bayes method.

**Conclusion**

Using Data Mining can find important information from data warehouses whose benefits have not been known until now. Through the Naïve Bayes algorithm, classification can be carried out using probability and statistical methods, namely predicting future opportunities based on previous experience. The results of the predicted birth rate at the Salatiga City Regional General Hospital in 2024 can be used by the Salatiga City Population Service to assist population officers in evaluating data reports and also as a reference for determining the number of births in the future. The Naive Bayes algorithm can calculate the probability of a dataset. The predicted results for the 2024 birth rate are labeled P( X | Description = "High") = **4.77192E-06** . From the results of this research, the Salatiga City Population Service can take steps or actions to reduce the birth rate by further increasing socialization and family planning development activities.

**Acknowledgments**

**References**

[1] Novia, Anggun Elni, "Prediction of Birth Rates in 2022 in North Sumatra Using the Naive Bayes Method," vol. 1, pp. 8-14, 2022.

[2] Idris, Muhammad, "Implementation of Data Mining Using the Naïve Bayes Algorithm to Predict Crime Scenes in Ponorogo Regency," *Paper Knowledge . Toward a Media History of Documents,* vol. 7, no. 1, pp. 1-33, 2020.

[3] Parihah, Nur Isnaini, "Prediction of Infant Birth Rates in Tridaya Sakti Village Using the Naive Bayes Algorithm," *Journal of Students' Research in Computer Science,* vol. 1, no. 2, pp. 77-88, 2020.

[4] Informatics, Engineering, "Implementation of the Naïve Bayes Algorithm in the Wonosobo Application," vol. 3, no. 2, pp. 301-310, 2022.

[5] Sabilla, Sinta Ayu, "Naive Bayes Method for Predicting Baby Birth (Case Study: Midwife S Clinic)," *Polinema Applicative Informatics Seminar,* pp. 2019-2022, 2021.

[6] Syamsiyah, Nur, "Design of a Loan Prediction Information System at the Panca Bhakti Bekasi Cooperative Using the C4.5 Algorithm," *Journal of Science & Technology,* vol. 9, no. 1, pp. 28-43, 2019.

[7] Rizal, Saiful, "Application of the Naïve Bayes Algorithm to Predict New Student Admissions at Al-Amien Vocational School, Wonorejo," *Explore IT: Scientific Journal and Applications of Informatics Engineering,* vol. 10, no. 1, pp. 14-21, 2018.

[8] Qisthiano, M Riski, "Development of a Model for Predicting Student Graduation Rates on Time Using the Naïve Bayes Method," *Budidarma Media Informatics Journal,* vol. 5, no. 3, p. 987, 2021.

[9] Tri Romadloni, Nova, "Comparison of Naive Bayes, Knn and Decision Tree Methods for Krl Commuter Line Transportation Sentiment Analysis," *IKRA-ITH Informatics Journal,* vol. 3, no. 2, pp. 1-9, 2019.

[10] Amin, Nur Fadilah, "General Concepts of Population and Samples in Research," *Pilar Journal,* vol. 14, no. 1, pp. 15-31, 2023.

[11] Indraswari, Naisha Rahma, "Application of Birth Age Prediction Using the Naive Bayes Method," Simetri: Journal of Mechanical Engineering, Electrical and Computer Science, vol. 9, no. 1, pp. 129-138, 2018.

[12] Damar Rani, Handini Arga, "Prediction System for Infant Birth Conditions using Naïve Bayes Classification," *Joined Journal (Journal of Informatics Education),* vol. 3, no. 2, p. 48, 2020.

[13] Sihombing, Pardomuan Robinson, "Application of Machine Learning Methods in Classifying the Risk of Low Birth Weight Events in Indonesia," MATRIK: Journal of Management, Informatics Engineering and Computer Engineering, vol. 20, no. 2, pp. 417-426, 2021.

[14] Emaniar, Ade Emaniar, "Increasing Maternal Knowledge in Preventing Stunting Through Community-Based Nutrition Education," *Advance Sustainable Science Engineering and Technology,* vol. 5, no. 2, p. 0230205, 2023.

[15] Pusadan, Mohammad Yazdi, "Implementation of Data Mining to Predict the Status of the Labor Process in Pregnant Women Using the Naive Bayes Algorithm," *Technomedia Journal,* vol. 8, no. June 1, pp. 137-153, 2023.

[16] Z. Zulham, "Prediction of Baby Birth Rates in Data Mining Using the Multi Regression Method," *Warta Dharmawangsa,* vol. 17, no. 1, pp. 30-40, 2023.

[17] WC Wahyudin, "Prediction of Stunting in Toddlers in Semarang City Hospitals Using Naive Bayes," vol. 2019, pp. 32-36, 2023.

[18] DFI Lestari, "Factors That Influence the Number of Births in Rapak Dalam Subdistrict," *Innovation,* vol. 14, no. 1, p. 8, 2018.

[19] W. Ananda, "Prediction of Palm Oil Harvest Results Using the Naive Bayes Algorithm," *TIN: Applied Informatics Nusantara Vol,* vol. 1, no. 10, pp. 513-519, 2021.

[20] MT Anwar, "Rain Prediction Using Rule-Based Machine Learning Approach," *Advance Sustainable Science, Engineering and Technology,* vol. 2, no. 1, pp. 1-6, 2020.

[21] VAV Setyawati, "Mapping Exclusive Breastfeeding Coverage And Toddler Stunting Prevalence In Indonesia Based On Web Geographic Information System," *Advance Sustainable Science, Engineering and Technology,* vol. 2, no. 2, pp. 1-5, 2020.

[22] MF Rifai, "Application of the Naïve Bayes Algorithm in the Graduation Level Prediction System for Microsoft Office Specialist (MOS) Certification Participants," *Petir,* vol. 12, no. 2, pp. 131-144, 2019.