Naive_Bayes_classifier_for_infant_weight_prediction

by Anita Desiani

PAPER · OPEN ACCESS

Naive Bayes classifier for infant weight prediction of hypertension mother

To cite this article: Anita Desiani *et al* 2019 *J. Phys.: Conf. Ser.* **1282** 012005

View the article online for updates and enhancements.

Naive Bayes classifier for infant weight prediction of hypertension mother

Anita Desiani [1,a], Rifkie Primartha [2], Muhammad Arhami[3], Osvari Orsalan [2]

[1] Universitas Sriwijaya Fakultas Matematika dan Ilmu pengetahuan Alam
[2] Universitas Sriwijaya, Fakultas Ilmu Komputer
[3] Politeknik Negeri Lhokseumawe

email: [a] anita_desiani@unsri.ac.id

Abstract :Classification is one method of data analysis in data mining that is used to form a model in order to describe the appropriate data class or model that predicts data trends. The Usage of classification has been applied in various areas, including in health areas. One of the classification methods used is Naive Bayes. This study aims to predict the weight of infants in maternal hypertensive and nonhypertensive conditions with Naive Bayes method. Data were taken as many as 219 data from pregnant women based on the medical record in Obstetrics and Gynecology of Muhammadiyah Palembang Hospital from January 1017 until September 2017. Data is divided into two groups, 188 for training data and 31 data for testing data. The performance data analysis was using WEKA and the results showed that the Naive Bayes's accuracy is 80.372%. the accuracy score means Naive Bayes works well to predict the weight of infants in maternal hypertensive and nonhypertensive mothers. The result is expected to be a reference for others research by comparing it with other classification methods and incorporating other factors in pregnancy and multiple births or other factors.

1. Introduction

World health statistics 2014 report stated that one in three adults worldwidely experienced an increase in blood pressure that resulting hypertension [1]. WHO also states that the incidence of hypertension worldwidely is estimated to cause 7.5 million deaths, or about 12.8% of all deaths (Global Health Observatory). WHO 2014 health report stated that hypertension in Indonesia was still sufficiently high about 32.5% occurs for adult males and about 29.3% occur in adult women [1]. Mothers who have a history of hypertension, have the possibility to give birth to the infant with abnormal weight. Furthermore, the mother who gave birth to babies under normal weight, can affect the condition of high blood pressure to the mother in the future [2]. The infant's birth weight is divided into 3 groups namely Normal Birth Weight (NBW), Low Birth Weight (LBW) and macrosomia is a large infant weight [3]. Many things can affect the birth weight of babies, such as hypertension suffered by the mother during pregnancy [4-5].

Low infant weight is less than 2500 gram of birth weight [3]. Low-birth-weight infants have a higher risk of having perinatal death at greater risk than normal infants at 8:1. At the same gestational age, 50% having a likelihood of learning disability, mental retardation and decreased IQ, may also cause long-term disability including auditory and visual disturbances, susceptible to cardiovascular disease and diabetes in the future[6-[7]. Globally, every year 4 million newborns die in the first 4 weeks of life, 36%, the infant's death occurs in the surrounding Southeast Asian countries.

Approximately 1.4 million neonatal deaths occur in Southeast Asian countries [8]. In other words 50% of infant deaths are caused by neonatal deaths in Southeast Asia. Kearney in 2005 calculated the number of hypertension in 2000 of 972 million, consisting of 333 in developed countries and 639 in developing countries, and predicted if there was no action to cope with the number of hypertension available then it is estimated that by 2025 the number up to 413 million in developed countries and about 1.15 billion in developing countries [9-10]. Precautions to overcome the incidence of high blood pressure in the world is very necessary [11]. In Indonesia the incidence of low birth weight babies in 2015 occurs about 9% of the total number of births in Indonesia [10].

South Sumatra as one of the provinces in Indonesia with low birth weight is 1: 9 for normal weight birth [12]. Palembang as the capital of South Sumatra in 2013 has a low birth weight rate of 10% lower than in 2010, which was 11.4%. However, if compared in 2008, which was only 0.6%, it can be said that the birth rate of infants in Palembang City increased sharply [13]. Hypertension in Southern Sumatra is the third most common illness [12]. If the rate of hypertension increases then the birth rate of the infant with low weight is likely to increase as well. To cope with that, low birth weight babies of mothers with hypertension is to monitor the condition of the mother's health, starting from the beginning of pregnancy. LBW can serve as an indicator of infant's survival, growth, and psychosocial development [14-15]. Thus, it is very important to maintain the condition of maternal hypertension during pregnancy so that the incidence of low birth of infants can be avoided. We recommend that the infant's weight to be born can be predicted well before 37 weeks, so that the various treatments to make the infant's weight normal even for mothers with hypertension.

One of the techniques that can be used as a prediction is data mining technique. Data mining refers to extracting useful information from vast amounts of data [16]. It is also known as data warehouses, or other information repositories [17]. The use of data mining has been widely applied in various fields such as for the classification of marine mammals [18], Software Defect Prediction Problem [19-20], social network analysis and social media usage [21-22], biology image classification [23], also in the health field. Data mining has been applied in many areas of health management [24-25]. The use of data mining techniques in the field of health include the use of the diagnosis of Diabetes Disease [26-27], nutrition prediction [28], classification of immunization in rural and urban communities [29], prediction of breast cancer and prostate cancer [30-31]. In addition, data mining is also used in classifying normal or cesarean birth with Algorithm C.45 [32]. The prediction of infant weight using data mining techniques has been done by Senthileskumar and Paulraj by 2015 to predict the risk of infants who have birth weight [33] and predicting birth weight by batch image has been done by Noguchi and Katto in 2014 [34].

Many classification methods developed in data mining include SVM, Naive bayes, decession tree, and so on. Naive bayes are the simplest form of Bayesian network classifiers [35]. Naive Bayes is a simple and fast classification method, the most popular and most used with satisfactory results [19], [36- 38]. It is suitable for many learning scenarios, such as image classification [39], prediction of carcinogenic content in food [40], text classification [35-36], [41], on human robot interaction to recognize gesture [38], predict mutenegitas on ames method [42], Sarcasm detection in microblogs [43], use of electricity in household (alpha). Naive Bayes is very easy to understand and is applied in various problems, however, this method is limited to categorical or discrete data [37]. Naive Bayes is one of the classification methods based on the bayes theorem with the assumption of conditional freedom [44-46].

Rosnah Sultan, *et.al* in 2014, conducted a study in Malaysia to compare which factors affected the birth weight of the infant between maternal age, ethnicity, gravida, parity, gestational age, maternal ordering weight, height and body mass index (BMI), severe infant history low birth, birth interval, ordering of hemoglobin levels, hypertension, diabetes mellitus and mode of delivery using logistic regression but they have not used these factors to predict the infant delivered weight [47]. Parisa Ahmadi, *et.al* in 2017, used the Random Forest to predict the infant weight by observing gestational age, body mass index for the third three months of pregnancy, maternal age and body mass index for the first three months of pregnancy with the accuracy of 95% . Senthilkumar in 2015 [48] compared several data mining techniques (one of which was Naive Bayes) to predict the infant weight by involving the attributes of the last maternal weight before pregnancy, and the age of the mother with

the accuracy of 77.8% [33]. This study uses naive bayes to predict the weight of a newborn infant by using several different attributes from previous studies, namely age, maternal education background, gravida status, hypertension condition and mother's birth experience.

2. Methodology

Data Set: Sources of data used in this study is secondary data derived from maternal patient status book contained in Medical Record Muhammadiyah Hospital Palembang taken from January until October 2017.

Pre-Processing Data: The data was obtained 250 data, but since there are some lost and ambiguous data, the data used in this study were only 219 data, with attributes involved such as maternal age, maternal gravida, maternal education level, birth-to-birth, hypertension conditions, and birth weight.

Naive Bayes Classifier : Naive Bayes Classifier is one of the clasification that works based on the method of Bayes theory. The eqution of bayes theory is as follow:

$$P(c|A) = \frac{P(A|c)P(c)}{P(A)} \tag{1}$$

where c is a class variable, A is a class of data not yet known, $P(A|c)$ is Probability based on condition on hypothesis, $P(c)$ is probability of hypothesis (prior probability), while $P(A)$ ignored because of constant value for all classes. A further description of the Bayes formula is made by describing $P(c|A_1, A_2 ..., A_n)$ using the rules of multiplication.

This is where the assumption of independence is very high (naive), that each instruction is independent of each other. With these assumptions, then apply a similarity as follows:

$$P(A|c) = P(a_1, a_2, ..., a_n = \prod_{i=1}^{n} P(A_i|c) \tag{2}$$

the naive bayes classifier can be calculated as follow[42] :

$$a_n(A) = \frac{P(c=+)}{P(c=-)} \prod_{k=1}^{n} \frac{P(a_k|c=+)}{P(a_k|c=-)} \tag{3}$$

If there is a label that never appears in the event then the way to handle the probability value 0 (zero) is to add 1 piece of data by using Laplace Correction method or also known as Laplacian Estimator method. In this paper the performance of naive bayes is be tested in 3 ways: supplied testing data set, cross validation, where has been taken as 10, and precenttages of 66% as training data, 34% as testing data

Analysis: An analysis of the results of any performance that performed to predict the weight of a newborn, it has been evaluated by looking at prediction accuracy, Precision and Recall for the accuracy of the resulting model.

$$Accuracy\ Prediction = \frac{TP+TN}{TP+TN+FP+FN} x100\% \tag{4}$$

$$precission = \frac{TP}{TP+FN} x100\% \tag{5}$$

$$recall = \frac{TN}{TN+TP} x100\% \tag{6}$$

where TP (True Positive) is data that can be properly calcified on a positive label, TN (True Negative) is the amount of data correctly specified in the negati class. While FP is a lot of data that is recognized as a positive label when the actual value in the negative class, FN (Negative Flase) is the amount of data introduced as a negative class, it is classified as a positive class.

3. Result and Discussion

This study took samples at Muhammadiyah Hospital Palembang. Characteristics studied include maternal age, education, gravida and infant weight, respectively. The data in this study is seen from the installation of medical colleagues and recapitulation books in the obstetrics and gynecology of MuhammadiYesh Hospital since January 1, 2017 to 30 September 2017. Some data that have been collected in this research can be seen in table 1.

Table 1. Data of Birth Statistic

Age	Education	Gravida	Hypertensive	Birth	Infant Weight
19	Senior High School	1	NO	1	Normal
17	Elementary school	1	NO	1	Normal
31	Senior High School	2	NO	3	Large
21	Junior High School	1	NO	2	Normal
30	Junior High School	1	Yes	3	Low
28	Junior High School	1	Yes	2	Low
28	Junior High School	2	NO	2	Normal
19	Elementary school	1	NO	1	Normal
17	Junior High School	1	NO	1	Normal
31	Senior High School	1	NO	3	Normal
27	Elementary school	1	NO	2	Normal
25	Junior High School	2	NO	2	Normal
29	Diploma	1	NO	2	Large
23	Junior High School	1	NO	2	Normal
32	Senior High School	2	NO	3	Normal
21	Elementary school	1	NO	2	Normal
29	Senior High School	2	Yes	2	Low
21	Senior High School	1	Yes	2	Normal
38	Junior High School	2	Yes	4	Low

From table 1, the age of pregnant women was changed into 4 groups, namely the age of ≤20 years (group 1) as many as 4.6%, 21- 29 years(group 2) is 47%, 30-34 years (group 3) is 23.7% and age ≥35 years old (group 4) is 24.7%. The mother's education level is the last education of mothers who are both hypertensive and not hypertensive. From the data obtained that the education of the mothers has about 13.7% for elementary school, 31% for junior high school, 32% for senior high school, 16% for diploma and 20% for undergraduate background education. Gravida mother is divided into 2 namely primigravida (1) and multigravida (2). Primigravida is the first pregnancy, while multigravida is a woman who has been pregnant multiple times. The results showed that 96 people were included in the primigravida group and 123 were admitted to the multigravida group, of which 63 primigravida did not have hypertension, 23 with hypertensive, 85 non-hypertensive multigravides and 38 multigravides women with hypertension. The other atribute is condition of hypertension where were also grouped in 2, namely from the group of mothers who had hypertension and mother who did not have hypertension. The attribute as based for classification is status of infant weight. The status of infant weight has been divided into 3 classification that are low (less than 2500gram), Normal (2500gram-3500gram),

The data obtained actually consists of many attributes, since there are some missing data and the existence of ambiguous data contents then in table 1 is only selected some attributes that will be used in this research, namely age factor, hypertension condition, mother gravida, infant's weight condition, birth, and mother education. Total data obtained is 219 data. In this study the infant weight data is divided into two groups. The first group of 188 data is used as data training and second group 31 as data testing. All types of data are of nominal type since naive bayes work very well on data of a nominal or categorical type. The data in this study then classified as to predict infant weight with the help of WEKA software. The proportion of giving birth mother based on eductaion level can be seen in table 2.

Table 2. Proportion of Giving birth mother based on educational levels

Atributes	Class		
	Normal	Low	Large
Elementary School	20	7	7
Junior High School	31	12	2
Sebior High School	64	10	14
DIPLOMA	4	0	3
Bachelor	12	0	2
Total	131	29	28

In Naïve Bayes we always use Laplacian Estimator to avoid the zero probability caused by the absence of a possibility from a cluster of events. Laplace Correction (Laplacian Estimator) or additive smoothing is a way to handle a probability value of 0 (zero) for example in table 3.

Table 3. Laplacian Estimator of Table 2

Atributes	Class		
	Normal	Low	Large
Elementary School	21	8	8
Junior High School	32	13	3
Sebior High School	65	11	15
DIPLOMA	5	1	4
Bachelor	13	1	3
Total	136	34	33

The propotion of giving birth mother based education level in table 2. The Diploma and Bachelor do not have data for low class, so to avoid the emergence of probability zero we use The Estimator Laplacian, then plus one instant as additional data as 1 data with low classification for education diploma and bachelor, this also should apply to other education levels to keep data consistent. Table 3 is the result table of the Estimator laplacian for Proportion of giving birth mother based on education levels.

Naive Bayes method can be done as follow:

1. Calculating the Value of New Case Opportunities From Any Hypothesis with Existing Class (Label) "P (Ak | Ci)"
2. Calculating the Accumulated Value of Opportunities From Each Class "P (A | Ci)"
3. Calculating the Value of P (Ai | Ci) x P (Ci)
4. Determine the Class of the new Case.

In this study to measure the accuracy of the results of the naive bayes model 3 test options are used as supplied test, Cross Validation, Split Percentage.

3.1. Supplied test set

In this research 188 data is used as training, for testing 31 new data is supplied. Predicted results obtained from WEKA are as follows:

```
 a   b  c   <-- classified as
15   0  1 |  a = normal
 3   5  0 |  b = Low
 2   1  4 |  c = Large
```

From the confusion matrix the recall value for normal class is 93.8%, the recall value for the low class is 62.5% and the recall value for the large class is 57.1%. The precision values for each class were 75% Normal, 83.3% Low, and 80% Big, respectively. the accuracy of the Naive Bayes model for the prediction of a infant's weight condition is 77.419%, of which 24 can be correctly predicted and 7 data

can not be correctly classified. The results of the confusion matrix show that each data is correctly recognized as normal infant weight, 5 data can be correctly classified as a low-weight infant condition, and 4 data can be correctly classified under conditions of infants with large body weight, resulting in a total accuracy can be calculated as follows:

$$The\ Accuracy = \frac{15+5+4}{31}x100\% = 77.419\% \qquad (7)$$

3.2. Cross-validation

In cross-validation, the value of fold 10, where the 188 training data is divided into 10 subset. So, there are 10 tests. Where, each datum is test data 1 times, and become training data as much as k-1 times. Then, the error of the test average k is calculated. Confusion matrix results obtained from Cross validation are:

```
 a    b    c   <-- classified as
122   3    6 | a = normal
  9  20    0 | b = Low
 15   0   13 | c = Large
```

Recall values for each class were 93.1% Normal, 69% Low, 46.4% Large. Precision values of each class were 83.6% Normal, 87% Low, and 68.4% Large. Predictive results show that 122 data are recognized correctly as normal classes and 3 more are recognized as low, and the other 6 are identified as large classification data. There are 20 data that can be recognized precisely as a low class, and 9 again are recognized as normal classes. There are 13 data recognizable as large classes, but 14 more are recognized as normal. So the accuracy obtained is:

$$the\ accuracy\ = \frac{122+20+13}{188}x100\% = 82.4468\% \qquad (8)$$

3.3. Percentage split

On the split percentage the data is divided into 2 groups by percent respectively. In this study 66% of data used as training data, while the remaining 34% of data will be used as data testing. The confusion matrix obtained is:

```
 a   b   c   <-- classified as
39   0   1 | a = normal
 2   6   0 | b = Low
 9   0   7 | c = Large
```

The recall value of each class is 97.5% Normal, 75% Low, and 43.8% Large. Precision values of each class are 78% Normal, 100% Low and 87.5% Large. The accuracy of the model is:

$$The\ Accuracy = \frac{39+6+7}{188}x100\% = 81.25\% \qquad (9)$$

The comparison from the three different methods can be seen in table 4.

Table 4. Comparison of the Naive Bayes accuracy model

Test Option	Recall (%)			Precision(%)			Accuracy
	Normal	Low	Large	Normal	Low	Large	
Supplied Test	93.8	62.5	57.1	75	83,3	80	77.419
Cross-Validation	93.1	69	46.4	83.6	87	68.4	82.4468
Percentage Split	97.5	75	43.5	78	100	87.5	81.25

From table 4 can be seen the success rate of system in rediscovering an information (Recall value) for normal class is 94.8%, while for low class equal to 68.83%, and for big class average is 48.97% , although large classes are only 48.97%, this is because the amount of data that has a large class is not as much data for normal and low classes. the level of accuracy between the information requested by the user with the answers given by the system or referred to as Precisiond from each test option either, the precision value is shown for each class above 50% where, the normal class is 78.867%, for the low class 90.1%, and for large classes 78.63%. The accuracy of each test option also shows the prediction model of Naive Bayes is very well used to predict the infant's weight especially for mothers who are hypertensive.

Conclusion

The results of the study and discussion can be concluded that the Bayesian Naive method can be used to predict the weight of infants against the condition of mothers who are hypertensive and non-hypertensive. Naive bayes provide good predictive results with an average accuracy of 80.372%. This study only involves general factors so it is advisable to proceed for more specific data such as conditions during pregnancy and birth such as premature, twins and other factors.

References

[1]   WHO Library Cataloguing-in-Publication Data 2014 *World health statistics 2014*

[2]   Ediriweera D S, Dilina N, Perera U, Flores F and Samita S 2017 Risk of low birth weight on adulthood hypertension - evidence from a tertiary care hospital in a South Asian country Sri Lanka : a retrospective cohort study *BMC Public Health* 17 15–20

[3]   World Health Organization 2011 *Maternal Mortality in 2011: Estimates Developed by WHO, UNICEF, UNFPA and the World Bank*

[4]   Cunningham S C Y, F.G, Levono J.L, Rouse D.J, Bloom, S.L, Hauth, J.C, *Obstetri Wiliam*. Jakarta: EGC, 2006.

[5]   T. Podymow and P. August, "New Evidence in the Management of Chronic Hypertension in Pregnancy," *Semin. Nephrol.*, vol. 37, no. 4, pp. 398–403, 2017.

[6]   UNICEF, "Reduction of Low Birth Weight a South Asia Priority," 2002.

[7]   R. Mochtar, "Toksemia Gravidarum," in *Sinopsis Obsetri*, Jakarta: EGC, 1998, pp. 198–208.

[8]   M. Islam, "Newborn survival – the forgotten milestone for achieving MDG 4 Monir," *WHO South-East Asia J. Public Heal.*, vol. 1(3), pp. 227–231, 2012.

[9]   J. Kearney, P.M., Whelton, M., Reynolds, K., Muntner, P., Whelton, P.K., He, "Global burden of hypertension: analysis of worldwide data," *Lancet*, vol. 365, no. 9455, pp. 217–233, 2005.

[10]   K. Sohn, "The Association between Height and Hypertension in Indonesia," *Econ. Hum. Biol.*, vol. 27, pp. 74–83, 2017.

[11]   S. Bromfield and P. Muntner, "High Blood Pressure : The Leading Global Burden of Disease Risk Factor and the Need for Worldwide Prevention Programs," *Curr Hypertens Rep*, vol. 15. Springer Publishing International, pp. 134–136, 2013.

[12]   BPS-Statistics of Sumatera Selatan, "Sumatera Selatan Province in Figure 2016," Palembang, 2016.

[13]   D. K. K. Palembang, "PROFIL KESEHATAN KOTA PALEMBANG," Palembang, 2016.

[14]   M. Dahlui *et al.*, "Risk factors for low birth weight in Nigeria : evidence from the 2013 Nigeria Demographic and Health Survey Risk factors for low birth weight in Nigeria : evidence from the 2013 Nigeria Demographic and Health Survey," *Glob. Health Action*, vol. 9, no. February, 2017.

[15]   D. Mohammed, A. Rahman, A. Abdel, N. Musa, A. Gamal, and A. El-aziz, "Egyptian Pediatric Association Gazette Incidence , risk factors and complications of hyperglycemia in very low birth weight infants," *Egypt. Pediatr. Assoc. Gaz.*, vol. 65, no. 3, pp. 1–8, 2017.

[16]   Y. Li, "DATA MINING: CONCEPTS, BACKGROUND AND METHODS OF INTEGRATING UNCERTAINTY IN DATA MINING," pp. 2–7.

[17]   M. Han, J. & Kamber, *Data Mining Concepts and Techniques*, 2nd ed. San Francisco, CA, USA: Morgan Kauffman, 2006.

[18] F. R. González-hernández, L. P. Sánchez-fernández, S. Suárez-guerra, and L. A. Sánchez-pérez, "Marine mammal sound classification based on a parallel recognition model and octave analysis," *Appl. Acoust.*, vol. 119, pp. 17–28, 2017.

[19] Ö. F. Arar and K. Ayan, "A Feature Dependent Naive Bayes Approach and Its Application to the Software Defect Prediction Problem," *Appl. Soft Comput. J.*, vol. 59, pp. 197–209, 2017.

[20] I. Arora, V. Tetarwal, and A. Saha, "Open Issues in Software Defect Prediction," *Procedia - Procedia Comput. Sci.*, vol. 46, no. Icict 2014, pp. 906–912, 2015.

[21] M. Injadat, F. Salo, A. B. Nassif, M. Injadat, F. Salo, and A. B. Nassif, "Mining Techniques in Social Media : A Survey," *Neurocomputing*, vol. 214, no. C, pp. 654–670, 2016.

[22] M. Adedoyin-Olowe, M. M. Gaber, and S. Frederic, "A Survey of Data Mining Techniques for Social Media Analysis," *Soc. Inf. Networks*, vol. 3, no. 6, 2014.

[23] C. Affonso *et al.*, "Deep learning for biological image classification," *Expert Syst. Appl.*, vol. 85, pp. 114–122, 2017.

[24] Y. Shirisha, S. S. S. Rao, and D. Sujatha, "Improved weighted associative classification approach," *Glob. J. Comput. Sci. Technol.*, vol. 11, no. 22, 2011.

[25] S. Patel and H. Patel, "S URVEY OF DATA MINING TECHNIQUES USED IN HEALTHCARE DOMAIN," *Int. J. Inf. Sci. Tech.*, vol. 6, no. 1, pp. 53–60, 2016.

[26] A. M. Salem, "Reasoning Techniques for Diabetics Expert Systems," *Procedia - Procedia Comput. Sci.*, vol. 65, pp. 813–820, 2015.

[27] M. C. D. Thenmozhi, "Identifying Disease -Treatment Relations using Machine Learning Approach," *Procedia - Procedia Comput. Sci.*, vol. 87, pp. 306–315, 2016.

[28] S. Kusumadewi, "KLASIFIKASI STATUS GIZI MENGGUNAKAN," *CommIT*, vol. 3, no. 1, pp. 6–11, 2009.

[29] M. HEMALATH and S. MEGALA, "MINING TECHNIQUES IN HEALTH CARE: A SURVEY OF IMMUNIZATION," *J. Theor. Appl. Inf. Technol.*, vol. 25, no. 2, 2011.

[30] B. Durgalakshmi and V. Vijayakumar, "PROGONOSIS AND MODELLING OF BREAST CANCER AND ITS GROWTH NOVEL NAIVE BAYES," *Procedia - Procedia Comput. Sci.*, vol. 50, pp. 551–553, 2015.

[31] M. A. Tahir, A. Bouridane, F. Kurugollu, and A. Amira, "A Novel Prostate Cancer Classification Technique Using Intermediate Memory Tabu Search," *EURASIP J. Appl. Signal Processing*, vol. 14, pp. 2241–2249, 2005.

[32] L. M. Nindia, "Klasifikasi persalinan normal atau caesar menggunakan algoritma C4.5," Universitas Sanata Darma, 2017.

[33] D. Senthilkumar and S. Paulraj, "Prediction of Low Birth Weight Infants and Its Risk Factors Using Data Mining Techniques," in *International Conference on Industrial Engineering and Operations Management*, 2015, pp. 186–194.

[34] C. Noguchi and J. Katto, "Improvement of Height Estimation of Low Birth Weight Infants , Newborns and Infants Image Processing System Using Kinect," in *IEEE 3rd Global Conference on Consumer Electronics (GCCE)*, 2014, no. 6, pp. 221–222.

[35] L. Jiang, C. Li, S. Wang, and L. Zhang, "Deep feature weighting for naive Bayes and its application to text classification," *Eng. Appl. Artif. Intell.*, vol. 52, pp. 26–39, 2016.

[36] L. Zhang, L. Jiang, C. Li, and G. Kong, "Two Feature Weighting Approaches for Naive Bayes Text Classiþers," *Knowledge-Based Syst.*, vol. 100, pp. 137–144, 2016.

[37] C. Hsu and Y. Huang, "Extended Naive Bayes classifier for mixed data," *Expert Syst. with Appl. Elsevier*, vol. 35, pp. 1080–1083, 2008.

[38] H. J. Escalante, E. F. Morales, and L. E. Sucar, "A Naive Bayes baseline for early gesture recognition," *Pattern Recognit. Lett.*, vol. 73, no. C, pp. 91–99, 2016.

[39] M. Ontivero-ortega, A. Lage-castellanos, G. Valente, R. Goebel, and M. Valdes-sosa, "Fast Gaussian Naïve Bayes for searchlight classification analysis," *Neuroimage*, vol. In Press, 2017.

[40] H. Zhang, Z. Cao, M. Li, Y. Li, and C. Peng, "Novel naïve Bayes classification models for predicting the carcinogenicity of chemicals," *Food Chem. Toxicol.*, vol. 97, pp. 141–149, 2016.

[41] G. Feng, J. Guo, B. Jing, and T. Sun, "Feature Subset Selection Using Naive Bayes for Text Classiþcation," *Pattern Recognit. Lett.*, vol. 65, pp. 109–115, 2015.

[42] H. Zhang *et al.*, "Novel naïve Bayes classification models for predicting the chemical Ames mutagenicity," *Toxicol. Vitr.*, 2017.

[43] S. Mukherjee and P. K. Bala, "Sarcasm detection in microblogs using Naïve Bayes and fuzzy clustering," *Technol. Soc.*, vol. 48, pp. 19–27, 2016.

[44] E. George, G. P. Box, and C. Tiao, *Bayesian Inference in Statistical Analysis*, 2011th ed. John Wiley & Sons. Inc, 2011.

[45] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2013th ed. Springer-Verlag, 2013.

[46] L. Wang, "Combining decision tree and Naive Bayes for classification," *Knowledge-Based Syst.*, vol. 19, pp. 511–515, 2006.

[47] R. Sutan, M. Mohtar, A. N. Mahat, and A. M. Tamil, "Determinant of Low Birth Weight Infants : A Matched Case Control Study," no. March, pp. 91–99, 2014.

[48] P. Ahmadi, H. A. Majd, S. Khodakarim, L. Tapak, S. Safshekan, and F. Pazhuheian, "Original Article : Prediction of low birth weight using Random Forest : A comparison with Logistic Regression," vol. 8, 2017.

Naive_Bayes_classifier_for_infant_weight_prediction

14% SIMILARITY INDEX **10**% INTERNET SOURCES **14**% PUBLICATIONS **8**% STUDENT PAPERS

PRIMARY SOURCES

1 Submitted to Sriwijaya University
Student Paper
6%

2 Submitted to Harrisburg University of Science and Technology
Student Paper
1%

3 Kisron, Bima Sena Bayu Dewantara, Femando Ardilla. "Early Warning and IoT-based Reporting System for Mobile Trash Bin Robot Application", 2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC), 2018
Publication
1%

4 Kitae Sohn. "The Association between Height and Hypertension in Indonesia", Economics & Human Biology, 2017
Publication
1%

5 www.scirp.org
Internet Source
1%

6 Ikhsan Romli, Toga Pardamean, Sufajar Butsianto, Tri Ngudi Wiyatno, Effendi bin Mohamad. "Naive Bayes Algorithm
1%

Implementation Based on Particle Swarm Optimization in Analyzing the Defect Product", Journal of Physics: Conference Series, 2021
Publication

7   T S Jaya, M Yusman. "Predicting the Quality of Pineapple Using the Naive Bayes Classifier Method", IOP Conference Series: Earth and Environmental Science, 2022
Publication   1 %

8   ecsjournal.org
Internet Source   1 %

9   Dwi Andini Putri, Dinar Ajeng Kristiyanti, Elly Indrayuni, Acmad Nurhadi, Denda Rinaldi Hadinata. "Comparison of Naive Bayes Algorithm and Support Vector Machine using PSO Feature Selection for Sentiment Analysis on E-Wallet Review", Journal of Physics: Conference Series, 2020
Publication   1 %

10   A P Ardhana, D E Cahyani, Winarno. "Classification of Javanese Language Level on Articles Using Multinomial Naive Bayes and N-Gram Methods", Journal of Physics: Conference Series, 2019
Publication   1 %

11   journals.sbmu.ac.ir
Internet Source   1 %

12 Sugandi Yahdin, Anita Desiani, Ali Amran, Desty Rodiah, Solehan. "Pattern recognation for study period of student in Mathematics Department with C4.5 algorithm data mining technique at the Faculty of Mathematics and Natural Science Universitas Sriwijaya", Journal of Physics: Conference Series, 2019
Publication

1 %

13 iopscience.iop.org
Internet Source

1 %

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		