IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.17, No.4, October 2023, pp. 395~404
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: https://doi.org/10.22146/ijccs.84242

Classification Of Maternal Health Risk Using Three Models Naive Bayes Method

Nurul Fathanah Mustamin[1], Firman Aziz*[2], Firmansyah[3], Pertiwi Ishak[4]
[1]Information Technology, Lambung Mangkurat University, Banjarmasin, Indonesia
[2, 3,4]Faculty of Mathematics and Natural Science, Pancasakti University, Makassar, Indonesia
e-mail: [1]nurul.mustamin@ulm.ac.id, ***[2]firman.aziz@unpacti.ac.id**,
[3]firmansyah17mb@gmail.com, [4]pertiwi.ishak@unpacti.ac.id

Abstrak

Kurangnya informasi terkait perawatan kesehatan ibu selama kehamilan dan pasca kehamilan terutama didaerah pedesaan mengakibatkan banyaknya kasus komplikasi kehamilan. Analisis resiko ibu hamil sangat dibutuhkan sebagai acuan dalam penanganan ibu hamil sehingga resiko terhadap ibu hamil dapat diminimalisir. Untuk menganalisis resiko ibu hamil dapat menggunakan teknik data mining dengan melakukan klasifikasi resiko ibu hamil. Penelitian ini mengusulkan untuk mengklasifikasi Maternal Health Risk menggunakan metode Naive Bayes dengan tiga model yaitu Gaussian, Multinomial, dan Bournolli. Data yang digunakan adalah data kesehatan ibu hamil berdasarkan intensitas resiko yang dikelompokkan menjadi tiga kelas yaitu low, mid, dan high risk. sedangkan untuk atribut yaitu Age, Systolic Blood Pressure as SystolicBP, Diastolic BP as DiastolicBP, Blood Sugar as BS, Body Temperature as BodyTemp, dan HeartRate. Hasil menunjukkan bahwa diantara tiga model Naïve Bayes yang memiliki kinerja terbaik adalah Multinomial dan Bournolli dengan akurasi sebesar 84.8% sedangkan Gaussian menghasilkan akurasi sebesar 82.6%.

***Kata kunci**— Klasifikasi, Resiko Ibu Hamil, Gaussian Naïve Bayes, Multinomial Naïve Bayes, Bournolli Naïve Bayes.*

Abstract

Lack of information related to maternal health care during pregnancy and post-pregnancy, especially in rural areas, results in many cases of pregnancy complications. Risk analysis for pregnant women is really needed as a reference in handling pregnant women so that the risk to pregnant women can be minimized. To analyze the risk of pregnant women can use data mining techniques by classifying the risk of pregnant women. This study proposes to classify Maternal Health Risk using the Naive Bayes method with three models, namely Gaussian, Multinomial, and Bournolli. The data used is the health data of pregnant women based on risk intensity which is grouped into three classes, namely low, mid and high risk. while the attributes are Age, Systolic Blood Pressure as SystolicBP, Diastolic BP as DiastolicBP, Blood Sugar as BS, Body Temperature as BodyTemp, and HeartRate. The results show that among the three Naïve Bayes models that have the best performance are the Multinomial and Bournolli with an accuracy of 84.8% while the Gaussian produces an accuracy of 82.6%.

***Keywords—** Classification, Maternal Health Risk, Gaussian Naïve Bayes, Multinomial Naïve Bayes, Bournolli Naïve Bayes.*

Received May 5th,2023; Revised July 30th, 2023; Accepted October 29th, 2023

1. INTRODUCTION

Pregnancy is a significant and often transformative phase in the lives of adult women, necessitating a keen focus on ensuring the well-being of both the expectant mother and the developing baby. Central to this concern is the imperative of managing and mitigating pregnancy-related risks to safeguard the health of both the mother and the unborn child. The early detection of pregnancy risks emerges as a pivotal approach to facilitating timely and targeted interventions by healthcare professionals. Regular antenatal check-ups with midwives or physicians offer a structured mechanism for addressing and minimizing these potential pregnancy risks. It is essential to underscore that a dearth of accessible information, particularly in rural areas, pertaining to maternal healthcare during and post-pregnancy has resulted in a multitude of cases characterized by pregnancy complications. Consequently, there arises an evident need for a systematic risk analysis framework tailored to pregnant women, aimed at not only identifying but also effectively mitigating potential risks. To this end, the application of data mining techniques for classifying pregnancy risks proves to be a valuable strategy. Data processing via data mining methods encompasses the intricate process of extrapolating likely scenarios based on historical data patterns and events. The concept of data mining essentially signifies the exploration of databases to unearth latent knowledge, thereby yielding novel and actionable insights in the context of maternal health and pregnancy management.

Research [1] proposes to classify using the Naïve Bayes method to predict births that will be experienced by pregnant women with the characteristics of the mother's age, height, Hb count, blood pressure, past pregnancy history, and congenital diseases. The results show that the greatest opportunity is in region 1. There are differences in the decision tree regarding the initial node of each region as the first causal factor for birth prediction. There are differences in program priorities for reducing MMR and IMR from each region referring to the initial node. Research [2] applied Pregnancy Risk Classification Using the Naïve Bayes Method with characteristics of age, number of children, height, and spacing of pregnancies, pregnancy failure, vacuum forceps, urethral intrusion, infusion, transfusion, cesarean section, anemia, malaria, pulmonary tuberculosis, poor heart disease, diabetes, sexually transmitted infections, facial swelling, twins, water twins, stillbirth, over months, breech position, transverse position, bleeding, and seizures. The results show that the calculation of the total probability is obtained by multiplying the probability calculation for each class and the attributes per class, then calculating the percentage for each class and compared to the largest one, which becomes the class resulting from the pregnancy risk classification. Research [3] proposes a Prediction Model for Maternal Health with the characteristics of Age, Pressure as SystolicBP, Systolic Blood, Blood Sugar as BS, Diastolic BP as DiastolicBP, HeartRate, and Body Temperature as BodyTemp. The results show that the highest accuracy performance uses a decision tree algorithm with 15 cross-fold validations. Research [4] proposes to classify the risks of pregnancy using Classification Tree-based Ensemble Learning. The results show that Poedji Rochyati's pregnancy risk classification using the Classification Tree-based Ensemble Learning method has succeeded in improving the accuracy of previous studies based on cost-sensitive learning. For accuracy, the Ensemble Learning method produces the best value of 76% compared to the cost-sensitive learning method which produces the best value of 73%. For recall, the Ensemble Learning method produces the best value of 89.5% compared to the cost-sensitive learning method which produces the best value of 77.9%. Research [5] proposes a Comparison of Classification between Naive Bayes and K-Nearest Neighbor on the Risk of Diabetes in Pregnant Women with the characteristics of Pregnancies Glucose, Blood Pressure, Skin Thickness, Insulin, Diabetes Pedigree Function, Age, and Outcome. The results show that for data sharing using K-Fold Cross Validation K=10 in the Naïve Bayes algorithm, the result is 75.78% and for processing using KNN with a value of K=25, the result is 74.48%. From these results, Naïve Bayes is better than K-Nearest Neighbor (KNN).

Building upon the context presented earlier, it is noteworthy that the Naïve Bayes classification method exhibits favorable attributes such as relatively high accuracy levels and efficient execution times when compared to alternative classification methodologies. However, it is essential to highlight that prior research endeavors have not thoroughly examined the diverse models encompassed within the Naïve Bayes approach. Hence, this study seeks to address this research gap by proposing a novel framework for classifying Maternal Health Risk, employing the Naive Bayes method with three distinct models: Gaussian, Multinomial, and Bernoulli. The dataset under scrutiny comprises health-related data pertaining to pregnant women, which are categorized based on their risk intensity into three distinct classes: low, moderate, and high risk. The attributes considered for this classification process encompass variables such as Age, Systolic Blood Pressure (SystolicBP), Diastolic Blood Pressure (DiastolicBP), Blood Sugar (BS), Body Temperature (BodyTemp), and Heart Rate. By employing these diverse Naïve Bayes models, this study endeavors to not only provide a comprehensive analysis of maternal health risk but also compare the efficacy of these models in addressing the complexities and nuances associated with the domain of pregnancy-related healthcare.

This research presents a significant contribution by examining the performance of three distinct Naive Bayes models: Gaussian, Multinomial, and Bernoulli. The main objective of this study is to comprehensively evaluate and compare the effectiveness of these models in classifying the risk levels of pregnant women. By conducting a rigorous analysis of these Naive Bayes models, the research aims to gain valuable insights into their respective strengths and limitations, providing valuable guidance for researchers and practitioners when selecting the most suitable model for specific scenarios. Employing meticulous experimental procedures and a valid methodology, the study assesses the three Naive Bayes models using well-curated and representative datasets. The anticipated outcomes of this research are expected to offer clear recommendations on the most efficient model to assess risks for pregnant women, considering various data characteristics.

2. METHODS

In this study, there are several stages of research that are interconnected as shown in Figure 1 below.

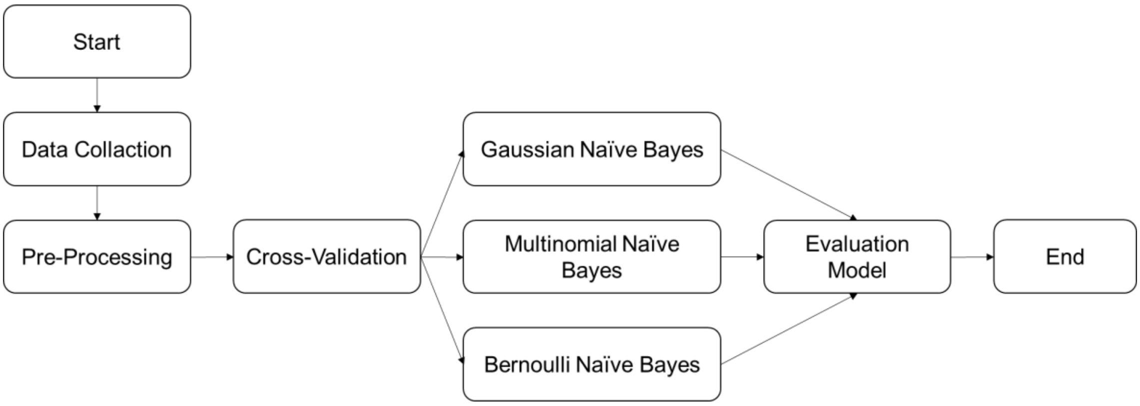

Figure 1 Research Stages

2.1 Data Collection

The process of data collection is a fundamental activity, focused on gathering relevant information and datasets, as previously elucidated [6]. In the context of this research paper, the data acquisition procedure involves the utilization of the UCI repository dataset site as the primary source. The dataset selected and employed for this particular study pertains to risk assessment data for pregnant women, providing the necessary foundation for subsequent analysis and evaluation.

2.2 Preprocessing

Data preprocessing, an integral phase in the data analysis pipeline, serves the purpose of preparing raw data for further processing and utilization within a classification algorithm, as referenced in sources [7], [8]. This preparatory stage encompasses a series of key processes, which include data cleaning, a crucial step in eliminating inconsistencies and errors, data transformation, to structure and format the data appropriately, and data normalization, which standardizes the data to facilitate equitable comparisons and accurate classification outcomes.

2.3 K-Fold

Cross-validation, a widely recognized model validation technique in the field of statistical analysis, is employed to evaluate the extent to which the outcomes of a given statistical analysis can be extrapolated and applied to independent datasets, as stipulated in reference [9]. This method serves as a robust means of assessing the generalizability and reliability of the statistical model, thus enhancing its applicability to real-world scenarios and diverse datasets that are not part of the original training data.

2.4 Naïve Bayes

The process of classification employing the Naive Bayes algorithm represents a crucial step in the data analysis pipeline, as outlined in reference [10]. This method entails the utilization of three distinct Naive Bayes models, each designed to handle specific aspects of the classification task. These models, when applied to the dataset, enable the system to make informed decisions based on probabilistic reasoning, resulting in effective categorization and insightful data analysis.

2.4.1 Gaussian Naïve Bayes

In the domain of classification, the execution of the Gaussian Naive Bayes algorithm is founded on the underlying assumption that the features utilized within the algorithm conform to a Gaussian distribution, as elaborated in reference [11]. This statistical approach, which relies on the normal distribution of features, serves as a pivotal component in the classification process, facilitating the accurate categorization and analysis of data.

$$P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right) \tag{1}$$

The parameters σ_y and μ_y are estimated using the maximum likelihood.

2.4.2 Multinomial Naïve Bayes

Is data that is distributed multinomially and is represented based on the number of vectors [12]. The distribution is parametrized by the vector $\theta_y = (\theta_{y1} \ldots \theta_{yn})$ for each class y, where *n* is the number of features and θ_{yi} is the probability $P(x_i \mid y)$ of feature *i* appearing in the class sample.

The parameter θ_y is estimated using the maximum likelihood version, which calculates the relative frequencies as follows:

$$\hat{\theta}_{yi} = \frac{N_{yi}+\alpha}{N_y+\alpha n} \tag{2}$$

2.4.3 Bernoulli Naïve Bayes

The implementation of the Naïve Bayes algorithm in the training phase, specifically when dealing with data distributed based on the Bernoulli distribution, necessitates the consideration of multiple features that can be regarded as Bernoulli Boolean variables, as indicated in references [13], [14]. Consequently, it becomes essential to acquire a suitable sample representation that aligns with the Bernoulli, Boolean characteristic.

In this context, the decision rule employed, known as Bayes' Naive Bernoulli decision rule, provides the guidelines for making data-driven decisions and classifications within this framework. This rule is a critical element in the classification process, allowing for the effective handling of binary and categorical data, ultimately contributing to the accuracy of the classification outcomes.

$$P(x_i \mid y) = P(x_i = 1 \mid y)x_i + (1 - P(x_i = 1 \mid y))(1 - x_i) \tag{3}$$

2.5 Performance Evaluation

In order to compute the error value associated with the classification method, the tool of choice is the utilization of a confusion matrix, which enables a comprehensive evaluation of the classification method's performance, as exemplified in the tabulated data presented in Table 1, as outlined in reference [15]. This matrix offers a detailed breakdown of true positives, true negatives, false positives, and false negatives, which are instrumental in gauging the method's accuracy, precision, recall, and overall effectiveness in the context of the specific classification task at hand.

Table 1 Confusion Matrix

Actual	Prediction		
	Low Risk	Mid Risk	High Risk
Low Risk	T_0	F_{01}	F_{02}
Mid Risk	F_{10}	T_1	F_{12}
High Risk	F_{20}	F_{21}	T_2

Description:
T0 (True 0) = The actual value is zero, and the result of the prediction model is zero.
T1 (True 1) = The actual value is one, and the result of the prediction model is one.
T2 (True 2) = The actual value is two, and the result of the prediction model is two.
F01 (False 01) = The actual value is zero, and the result of the prediction model is one.
F02 (False 02) = The actual value is zero, and the result of the prediction model is two.
F10 (False 10) = The actual value is one, and the result of the prediction model is zero.
F12 (False 12) = The actual value is one, and the result of the prediction model is two.
F20 (False 20) = The actual value is two, and the prediction model result is zero.
F21 (False 21) = The actual value is two, and the result of the prediction model is one.

Accuracy, as a fundamental metric, serves as a gauge of the proximity of a measurement result to its corresponding actual value. The quantification of accuracy, in this context, is accomplished through the utilization of a specific equation denoted as Equation 4, which encapsulates the precise mathematical relationship for assessing how well the measurement aligns with the true or expected value.

$$accuracy = \frac{\sum_j T_i}{N} \tag{4}$$

3. RESULTS AND DISCUSSION

The classification of pregnant women into three distinct risk categories, namely low-risk, mid-risk, and high-risk, will serve as the foundation for the dataset. This dataset encompasses a comprehensive total of 1014 records, each representing a holistic set of attributes, as visually depicted in Figure 2 below

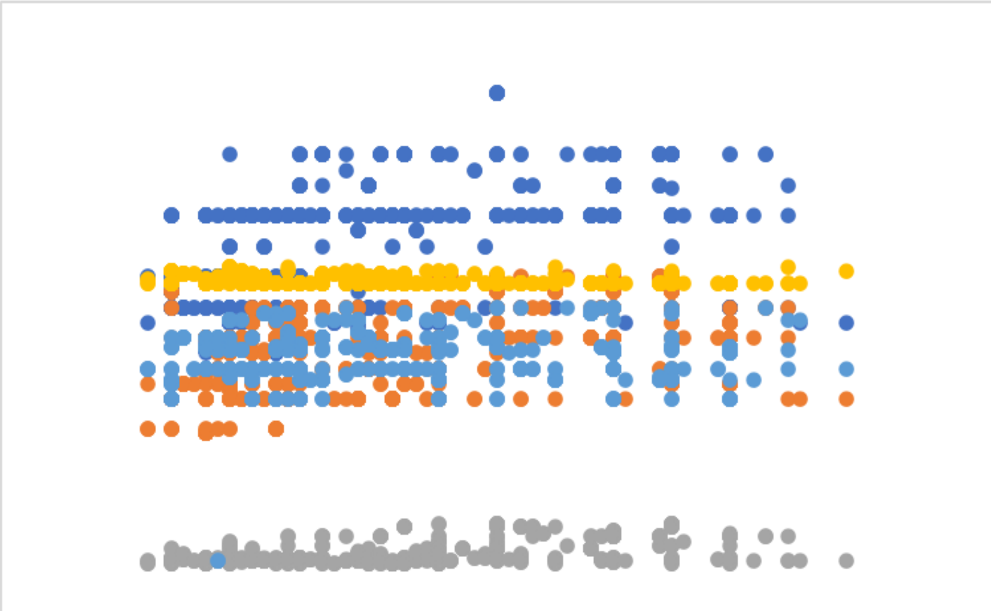

Figure 2 Overall Representation of Maternal Risk

Following this, the risk data associated with pregnant women embarks on an elaborate data processing journey, involving the meticulous application of a cross-validation technique with K=5 folds. This method partitions the dataset into distinct training and testing subsets, paving the way for rigorous evaluation and model validation.

In the scope of this study, the classification process is executed with precision, capitalizing on the integration of three distinctive Naïve Bayes models, namely Gaussian Naïve Bayes, Multinomial Naïve Bayes, and Bernoulli Naïve Bayes. The overarching objective is to conduct a comprehensive evaluation of the classification performance and the accuracy levels achieved by each of these models when applied to the intricate pregnancy risk dataset.

The entire process of classification and data analysis is meticulously orchestrated within the MATLAB software platform. This sophisticated software environment empowers the research team to extract an extensive array of results, facilitating a deep dive into the data for further analysis and the generation of insightful findings.

Table 2 confussion matrix result of model Gaussian Naïve Bayes

Actual	Prediction		
	Low Risk	Mid Risk	High Risk
Low Risk	331	59	16
Mid Risk	54	264	18
High Risk	11	18	243

Table 3 confussion matrix result of model Multinomial Naïve Bayes

Actual	Prediction		
	Low Risk	Mid Risk	High Risk
Low Risk	334	64	8
Mid Risk	45	277	14
High Risk	10	13	249

Table 4 confussion matrix result of model Bournolli Naïve Bayes

Actual	Prediction		
	Low Risk	Mid Risk	High Risk
Low Risk	336	63	8
Mid Risk	39	284	13
High Risk	15	17	240

Upon a meticulous examination of the outcomes derived from the confusion matrices associated with the three Naïve Bayes models, thoughtfully presented in Figures 3 to 5, the subsequent phase involves the implementation of performance calculations using the formula designated as Equation 4. These calculations are executed with a high degree of precision to assess the overall performance metrics of the three Naïve Bayes models.

The summative results of these performance calculations, which encompass key metrics such as accuracy, precision, recall, and other relevant indicators, have been thoughtfully compiled and are visually depicted in a comprehensive tabulation found in Table 3, offering a succinct yet thorough overview of the effectiveness and efficacy of each of the three Naïve Bayes models in the context of the study.

Table 5 Classification Results

Classification	**Accuracy**
Gaussian Naïve Bayes	82.6 %
Multinomial Naïve Bayes	84.8%
Bournolli Naïve Bayes	84.8%

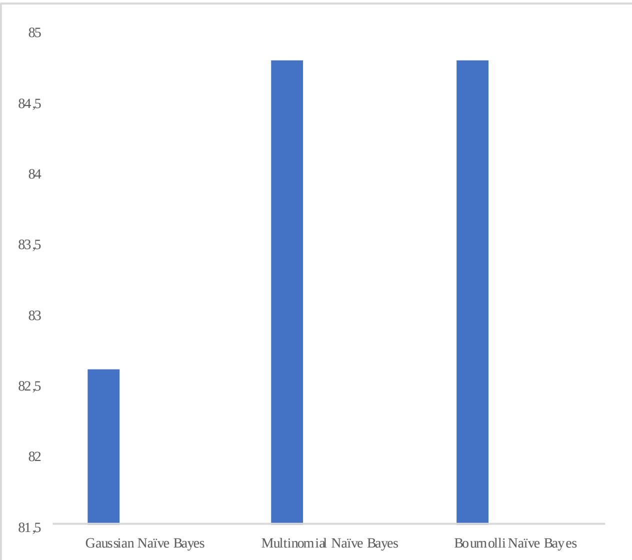

Figure 6 Classification method accuracy results

As elucidated in Table 3, a detailed analysis of the results underscores the notable distinction in accuracy among the three Naive Bayes models. Notably, the Multinomial and Bernoulli models emerge as the top performers, attaining the highest accuracy among the trio. In stark contrast, the Gaussian model lags behind, registering a comparatively lower accuracy rate. Specifically, the Multinomial and Bernoulli models exhibit a commendable accuracy rate of 84.8%, while the Gaussian model falls slightly short with an accuracy score of 82.6%.

The results gleaned from this comprehensive evaluation affirm the successful performance of all three Naive Bayes models in effectively classifying the risk levels associated with pregnant women. These results derive their strength from the extensive dataset, which underscores the reliability of the models in real-world scenarios.

However, it is imperative to acknowledge the obstacles encountered during this study. It has come to light that when dealing with an exceptionally wide data range, the classification performance tends to diminish, revealing an intrinsic challenge that warrants further investigation. This insight is pivotal for understanding the limitations of the models and the necessity for refining and optimizing the classification process in scenarios with expansive data ranges.

Table 6 Comparison of Maternal Health Classification Outcomes

Author	**Classification**	**Accuracy**
H. Amalia	Decision Tree	61,54%
B. Delvika, et.al	K-Nearest Neighbor	74.48%
Poedji Rochyati	Ensemble Learning based on cost sensitive learning	73%
M.A. Hidayat	Ensemble Learning based on Classification Tree	76%
Proposed	Gaussian Naïve Bayes	82.6 %
	Multinomial Naïve Bayes	84.8%
	Bournolli Naïve Bayes	84.8%

In this comparative analysis, several authors have applied different classification techniques to assess their accuracy in predicting specific outcomes. H. Amalia utilized a Decision Tree classification method, yielding an accuracy of 61.54%. Meanwhile, B. Delvika and colleagues employed the K-Nearest Neighbor approach, achieving an accuracy of 74.48%. Poedji Rochyati's study involved Ensemble Learning based on cost-sensitive learning, which resulted in an accuracy of 73%. In M.A. Hidayat's work, Ensemble Learning based on Classification Tree was implemented, achieving a 76% accuracy rate. Furthermore, the proposed approach used three Naïve Bayes models—Gaussian Naïve Bayes with 82.6% accuracy, as well as Multinomial Naïve Bayes and Bernoulli Naïve Bayes, both obtaining a notable accuracy rate of 84.8%. These results collectively reflect the effectiveness and diversity of classification techniques in addressing their respective research objectives.

4. CONCLUSIONS

The evaluation of the success of proposing the Naive Bayes model in this study was predominantly grounded on the assessment of accuracy values. The performance results of the Naive Bayes model are comprehensively presented in Table 3, offering valuable insights into their effectiveness. It is discernible from the table that the volume of data exhibits a discernible influence on the performance of the Naive Bayes model, shedding light on a significant factor in the classification process. Notably, among the trio of Naïve Bayes models, the most robust performers are the Multinomial and Bernoulli models, boasting an impressive accuracy rate of

84.8%, while the Gaussian model, while still commendable, records a slightly lower accuracy of 82.6%. However, this study has brought to the forefront a critical challenge. When dealing with datasets characterized by an extensive data range, there is a noticeable decline in the performance levels of the models. This issue warrants careful examination and consideration for future research endeavors, with the aim of devising strategies to overcome this inherent challenge in classification tasks.

REFERENCES

[1] A. N.- BIMIPA and undefined 2013, "Klasifikasi Naïve Bayes untuk Prediksi Kelahiran pada Data Ibu Hamil," *journal.ugm.ac.id*, vol. 23, no. 3, 2013, Accessed: Jan. 27, 2023. [Online]. Available: https://journal.ugm.ac.id/bimipa/article/view/13859.

[2] Q. Hasanah, A. Andrianto, M. H.-B. Sainstek, and undefined 2018, "Sistem Informasi Posyandu Ibu Hamil dengan Penerapan Klasifikasi Resiko Kehamilan Menggunakan Metode Naïve Bayes," *jurnal.unej.ac.id*, Accessed: Jan. 27, 2023. [Online]. Available: https://jurnal.unej.ac.id/index.php/BST/article/view/7554.

[3] M. Ahmed, M. K.-2020 2nd I. C. on, and undefined 2020, "IoT based risk level prediction model for maternal health care in the context of Bangladesh," *ieeexplore.ieee.org*, doi: 10.1109/STI50764.2020.9350320.

[4] M. H.-I. I. Journal and undefined 2021, "Klasifikasi Resiko Kehamilan Menggunakan Ensemble Learning berbasis Classification Tree," *jurnal.unej.ac.id*, vol. 6, no. 3, p. 177, 2021, Accessed: Jan. 27, 2023. [Online]. Available: https://jurnal.unej.ac.id/index.php/INFORMAL/article/view/28396.

[5] J. Homepage, B. Delvika, S. Nurhidayarnis, P. D. Rinada, N. Abror, and A. Hidayat, "Perbandingan Klasifikasi Antara Naive Bayes dan K-Nearest Neighbor Terhadap Resiko Diabetes pada Ibu Hamil: Comparison of Classification Between Naive," *journal.irpi.or.id*, vol. 2, pp. 68–75, 2022, Accessed: Jan. 27, 2023. [Online]. Available: https://journal.irpi.or.id/index.php/malcom/article/view/432.

[6] M. G.-R. with children and young people and undefined 2009, "Data collection and analysis," *books.google.com*, Accessed: Jan. 29, 2023. [Online]. Available: https://books.google.com/books?hl=en&lr=&id=DVBrL1UiatcC&oi=fnd&pg=PA65&dq=data+collection&ots=qv8L5s9o7b&sig=gAVcNOmYsXGEtIw2ZCH5bgJYRS8.

[7] L. Breiman, "Random Forests," *Mach. Learn. 2001 451*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[8] S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez, and F. Herrera, "Big data preprocessing: methods and prospects," *Big Data Anal.*, vol. 1, no. 1, Dec. 2016, doi: 10.1186/S41044-016-0014-0.

[9] S. Wager, "Cross-Validation, Risk Estimation, and Model Selection," Sep. 2019, Accessed: Jan. 29, 2023. [Online]. Available: http://arxiv.org/abs/1909.11696.

[10] S. Chen, G. Webb, L. Liu, X. M.-K.-B. Systems, and undefined 2020, "A novel selective naïve Bayes algorithm," *Elsevier*, Accessed: Jan. 29, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705119306185.

[11] Z. Bi, Y. Han, C. Huang, M. W.-2019 I. Conference, and undefined 2019, "Gaussian naive Bayesian data classification model based on clustering algorithm," *atlantis-press.com*, 2019, Accessed: Jan. 29, 2023. [Online]. Available: https://www.atlantis-press.com/proceedings/masta-19/125913250.

[12] G. Singh, B. Kumar, … L. G.-… C. on A., and undefined 2019, "Comparison between multinomial and Bernoulli naïve Bayes for text classification," *ieeexplore.ieee.org*, Accessed: Jan. 29, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8776800/.

[13] M. Singh, M. Bhatt, H. Bedi, U. M.-M. T. Proceedings, and undefined 2020, "Performance of bernoulli's naive bayes classifier in the detection of fake news,"

Elsevier, Accessed: Jan. 29, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214785320385333.

[14] M. A.-P. C. Science and undefined 2021, “Review the performance of the Bernoulli Naïve Bayes Classifier in Intrusion Detection Systems using Recursive Feature Elimination with Cross-validated selection of,” *Elsevier*, Accessed: Jan. 29, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050921013156.

[15] F. Provost, R. K.-M. learning, and undefined 1998, “Guest editors’ introduction: On applied research in machine learning,” *academia.edu*, Accessed: Mar. 08, 2022. [Online]. Available: https://www.academia.edu/download/40088606/Guest_Editors_Introduction_On_Applied_R20151116-31348-ccsgqm.pdf.