JURNAL RISET INFORMATIKA **P-ISSN: 2656-1743 |E-ISSN: 2656-1735**
Vol. 5, No. 3. June 2023 DOI: https://doi.org/10.34288/jri.v5i3.238
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Classification of Blighted Ovum Factors in Pregnant Women Using PSO-Based Naïve Bayes

Febryo Ponco Sulistyo[1], Endang Sri Palupi [2*)]

Teknik Informatika
[1]Universitas Mercu Buana
Jakarta, Indonesia
[1]febryo.ponco@mercubuana.ac.id

Sistem Informasi
[2]Universitas Bina Sarana Informatika
Jakarta, Indonesia
[2*)] endang.epl@bsi.ac.id

(*) Corresponding Author

Abstract

Classification of Blighted Ovum Factors or undeveloped fetuses is carried out considering that many cases occur in pregnant women. Blighted Ovum is 60% of the causes of miscarriage. In Indonesia, it is found in 37% of every 100 pregnancies. Classification uses Naïve Bayes based on Particle Swarm Optimization (PSO), which only requires small training data to determine the parameter estimates needed in the classification process, and the use of Particle Swarm Optimization can increase accuracy and solve optimization problems with the process of selecting variable data and attribute data to create a questionnaire as a data collection method. The results of the classification of blighted Ovum in pregnant women using the Naïve Bayes algorithm with the Rapid Miner framework obtained an accuracy value of 71.56% with an Area Under Curve (AUC) of 0.674 included in the excellent classification category. After using the PSO optimization, the accuracy value rose to 79.82% with an Area Under the Curve of 0.764, including a good classification category. Naïve Bayes is a suitable method for classification, and PSO can improve the accuracy and AUC values .

Keywords: Classification, naïve Bayes, particle swarm optimization

Abstrak

Klasifikasi Faktor Blighted Ovum atau janin tidak berkembang dilakukan mengingat kasus Blighted Ovum banyak terjadi pada ibu hamil. Blighted Ovum merupakan 60% dari penyebab keguguran, di Indonesia ditemukan 37% dari setiap 100 kehamilan. Klasifikasi menggunakan Naïve Bayes berbasis Particle Swarm Optimization (PSO) yang hanya membutuhkan data training yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian dan penggunaan Particle Swarm Optimization dapat meningkatkan nilai akurasi serta memecahkan masalah optimasi. Dengan proses pemilihan data variable dan data attribute untuk membuat kuisioner sebagai metode pengambilan data. Hasil klasifikasi blighted ovum pada wanita hamil menggunakan algoritma Naïve Bayes dengan framework Rapid Miner mendapatkan nilai akurasi sebesar 71,56% dengan Area Under Curve (AUC) 0,674 termasuk dalam kategori klasifikasi yang baik. Setelah menggunakan optimasi PSO nilai akurasi naik menjadi 79,82% dengan Area Under Curve 0,764 termasuk kategori klasifikasi yang baik. Naïve bayes merupakan metode yang cocok untuk klasifikasi, dan PSO bisa membuat nilai akurasi dan AUC lebih baik lagi.

Kata kunci: Klasifikasi, naïve bayes, particle swarm optimization

INTRODUCTION

Cases of miscarriage caused by Blighted Ovum reach 60%. Until now, the cause of blighted Ovum cannot be detected because the symptoms are not specific. Factors - determinants of Blighted Ovum: Age factor, parity factor, immunological factor, genetic disorder factor (Anggrayni et al., 2022). Because of that, the authors want to classify the Blighted Ovum factor in pregnant women, and

classification is the process of finding a model (or function) that describes and distinguishes class or concept data that aims to be used to predict the class of objects whose class labels are unknown. (Han & Kamber, 2006). Empty pregnancy or blighted Ovum is a pregnancy that occurs without an embryo in the uterus. This pregnancy problem, also known as anembryonic gestation in the health world, is one of the causes of miscarriage in the first trimester of pregnancy. Usually, blighted Ovum occurs due to chromosomal abnormalities due to imperfect cell division or poor quality of sperm and Ovum. When an empty pregnancy occurs, sperm cells and Ovum still meet or fertilize. However, the results of the fertilization did not develop into an embryo. The condition of an empty pregnancy can be marked by pain in the stomach until bleeding occurs. Unfortunately, this pregnancy problem can only be known after an ultrasound examination. This is because various pregnancy symptoms, such as nausea, vomiting, positive results on a quick pregnancy check or pack test, and breasts that turn harder, are also experienced by pregnant women with an empty pregnancy. (Dr Rizal Fadli, 2022)

The author's goal in conducting this research is to classify the factors of blighted Ovum so that pregnant women are more aware of these factors and can carry out more detailed examinations and take anticipatory steps by consulting a specialist doctor.

In 2019 the journal entitled Analysis of the Bayes Theorem Method in Diagnosing Miscarriage in Pregnant Women Based on the Type of Food written by Fricles Ariwisanto Sianturi, This study used the Microsoft Visual Studio 2008 expert system application design, while the authors in this study used the Rapid Miner Framework. This study's authors classify the Blighted Ovum factor using the PSO-based Naïve Bayes Algorithm. At the same time, the journal Fricles Ariwisanto Sianturi applies the Bayes theorem method to diagnose types of food and can provide fast diagnosis results along with a risk level value for each food effect consumed. (Sianturi, 2019)

In the journal Early Diagnosis of Preeclampsia in Pregnant Women Using the K-Nearest Neighbor (KNN) Method in 2021 resulted in an accuracy of 88%, while the authors classified the blighted ovum factor in pregnant women using the PSO-based Naïve Bayes method resulting in an accuracy of 79.82%. (Normawati et al., 2021)

The following study, Detection of Fetal Health Using a Decision Tree and Feature Forward Selection, will be conducted in 2022. This study aims to detect fetal health using a decision tree classification algorithm without feature selection resulting in an accuracy of 89.84%. At the same time, using the forward selection feature in this decision tree algorithm produces an accuracy of 91.06%. This shows that using the forward selection feature can increase accuracy by 1.22%. (Sulihati et al., 2022)

In 2021 research on the Classification of Diagnosis of Gestational Diabetes in Pregnant Women uses the Neighbor Weighted K-Nearest Neighbor (NWKNN) Algorithm, the NWKNN Algorithm is superior to the KNN in classifying positive/diabetic classes. (Yolanda et al., 2021) In this study, the authors used the Naïve Bayes Algorithm based on PSO, while the journal written by Vinesia Yolanda and friends used the NWKNN Algorithm, namely the K-NN Algorithm, which was optimized using Neighbor Weighted. Both studies used optimization to get better classification results.

Fetal Health Recommendations Using the C5.0 Algorithm Using the Classifying Cardiotocography Dataset created in 2021 by Muhamad Rian and colleagues, this study This research proposes that the C5.0 Algorithm utilizes the Cardiotocography dataset regarding fetal conditions. The Cardiotocography dataset consists of 2,126 records, each with 22 attribute columns and 3 classification classes, namely; normal, suspect, and pathological. In conclusion, the greater the training data, the better the accuracy. (Rian Santoso & Musa, 2021) At the same time, the author uses PSO optimization to get better accuracy.

RESEARCH METHODS

The author conducted research at the Evasari Women's and Children's Hospital in Central Jakarta from February to April 2023. Taking data on abortion cases from 2019 to 2022, from this data Blighted Ovum causes 70% of abortion cases. Researchers created a questionnaire and distributed it to obstetrics and gynecology patients with blighted ovum attribute factors. To get more varied data, the author also distributed questionnaires to content discussion forums and the pregnancy program WhatsApp group that the author participated in. The questionnaires used blighted ovum factor attributes. The collected data set was 362. Questionnaires or questionnaires are data collection techniques that are carried out by giving respondents a set of questions or written statements to answer. (Sugiyono, 2017)

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v5i3.238
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Table 1. Blighted Ovum Factor Dataset

AGE	FARITY	GENETIC DISORDERS	IMMUNOLOGICAL DISORDERS	SPERM QUALITY	EGG CELL QUALITY	HAVE EXP BO	BO
40	NO	NO	NO	GOOD	GOOD	YES	BO
38	YES	NO	NO	GOOD	GOOD	YES	BO
35	NO	NO	NO	GOOD	GOOD	YES	BO
37	YES	NO	NO	GOOD	GOOD	NO	BO
35	YES	NO	NO	BAD	GOOD	NO	BO
42	NO	NO	NO	GOOD	BAD	NO	BO
35	NO	NO	YES	GOOD	CURRENTLY	NO	NO
25	NO	NO	YES	GOOD	CURRENTLY	NO	NO
38	YES	NO	NO	CURRENTLY	GOOD	NO	NO
34	NO	NO	NO	GOOD	GOOD	NO	BO
27	NO	NO	NO	GOOD	GOOD	NO	BO
30	YES	NO	NO	CURRENTLY	GOOD	NO	NO
33	YES	NO	YES	GOOD	GOOD	YES	BO
34	NO	NO	YES	GOOD	CURRENTLY	NO	NO

There are 4 determinant factors for Blighted Ovum, namely: age factor, parity factor, immunological disorders, and genetic disorders. (Anggrayni et al., 2022). Causes of blighted Ovum include: genetic disorders, egg cell quality, sperm quality, experienced blighted Ovum, and certain diseases.(Annisa Karnesyia, 2020)

The author uses an analytical descriptive method so that researchers can use it to identify deeper data so that we can know the characteristics of the data. Analytical descriptive is a method that describes or provides an overview of an object under study through data or samples collected as they are without analyzing to make general conclusions. (Sugiyono, 2013)

The author also uses data mining classification to define the similarity of characteristics in a group or class. Classification of data mining is one of the most common methods to use. This method aims to estimate the class of an object whose label is unknown. Classification techniques that can be used are Support Vector Machine, AdaBoost, Naïve Bayes, Constant, KNN, Tree, Random Forest, Stochastic Gradient Descent, and CN2 Rule. Naïve Bayes Classifier is a classification method rooted in Bayes' theorem. The main feature of this Naïve Bayes Classifier is the very strong (naïve) assumption of the independence of each condition/event. (Wibawa et al., 2018)

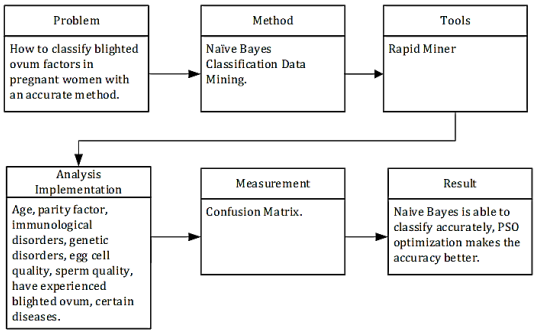

Figure 1. Research Stages

The general research stages are described in Figure 1 as follows.

1. Problem
   In this study, the problem of classifying blighted ovum factors in pregnant women uses an accurate method.
2. Method
   The method used to classify blighted ovum factors in pregnant women is Naïve Bayes.
3. Tools
   The tool in this study uses the Rapid Miner framework.
4. Implementation
   Classification using data training, analyzed using data mining classification Naïve Bayes.
5. Measurement
   Measurements were made using the Confusion Matrix.
6. Results
   Analyzing the data processing results and measuring the accuracy of the Naïve Bayes classification data mining method.

RESULTS AND DISCUSSION

The data training process uses the Naïve Bayes classification to determine the results of class distribution Yes (experiencing blighted Ovum) or No (not experiencing blighted Ovum). A training dataset is a data set used to train or build a model. The model is trained using a training dataset, then the performance during the exercise is tested using a validation dataset. This aims to see the model's ability during training and whether it can recognize patterns in general. (Salsabila MR, 2023)

The following Figure 2 classifies the training data using the Naïve Bayes Algorithm. Data mining algorithm performance in most cases depends on the dataset's quality because the training data's low quality can lead to weak classification. Thus, data techniques are needed for preprocessing to prepare the data.

This can improve data quality, which helps improve the accuracy and efficiency of the data mining process. Several data preprocessing techniques include data cleaning: deleting data that contains errors. Data integration: combining data from various sources. Data transformation: data normalization and reduction: reducing data size by concatenating and removing redundant features.(Khadafy & Wahoho, 2015)

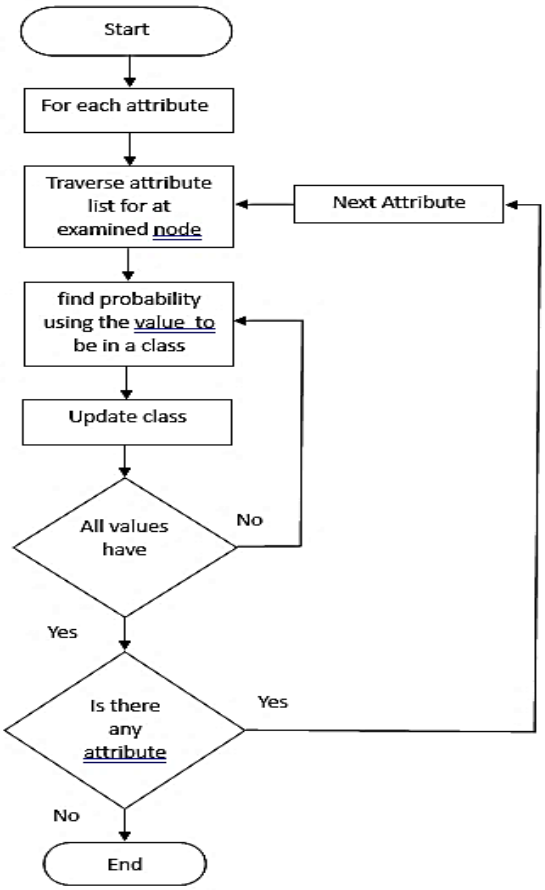

Figure 2. Naïve Bayes Classification Training Data Process

Table 2. Simple Distribution Model

Class	Mark
YES	0.310
NO	0.690

The results of the training data classification using the Naïve Bayes method are divided into 2 classification classes, namely YES (Blighted Ovum) class 0.310 and NO (Normal) 0.690 in Figure 3 and Table 2.

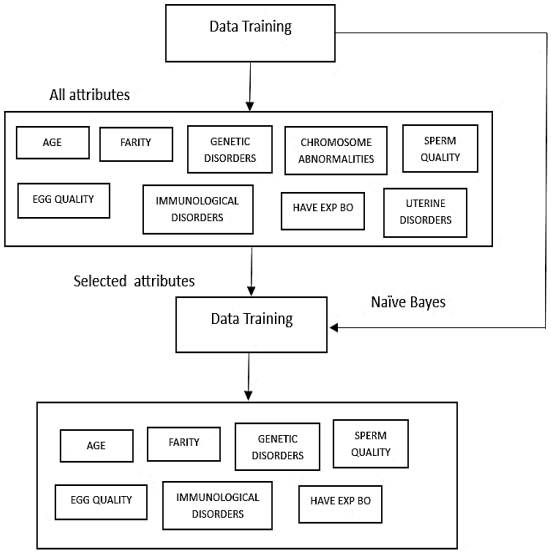

Figure 3. The Process of Selecting Attributes

In Figure 4, the attribute selection process aims to get the minor decision tree in size. Selection (selection) of data from a set of operational data needs to be done before extracting information in the KDD (Knowledge Discovery In Database) begins. The selected data for the mining process is stored in a file, separate from the operational database. (Adi Mhsd, 2021)

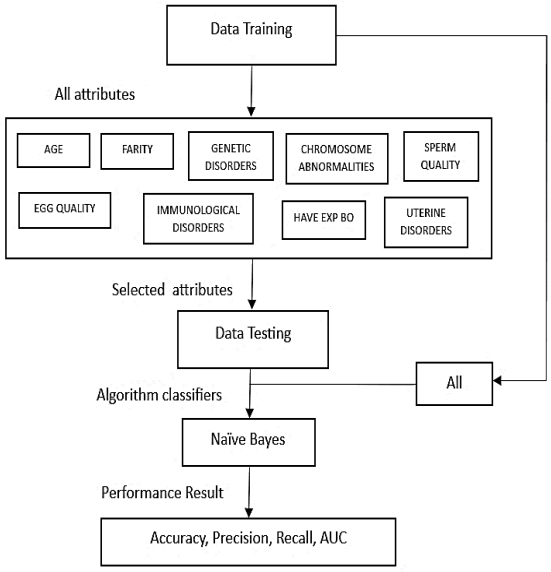

Figure 4. Classification process using Naïve Bayes

We start the classification process using the Naïve Bayes algorithm using Rapid Miner in Figure 5, then get the accuracy results in Table 3 of

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v5i3.238
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

71.56% with class precision BO 55.56% and class precision NO 76.83%.

Table 3. Naïve Bayes Algorithm Accuracy Results

Accuracy 71.56%	True BO	True NO	Class Precision
Prediction BO	15	12	55.56%
Prediction NO	19	63	76.83%
Class Recall	61.76%	88.00%	

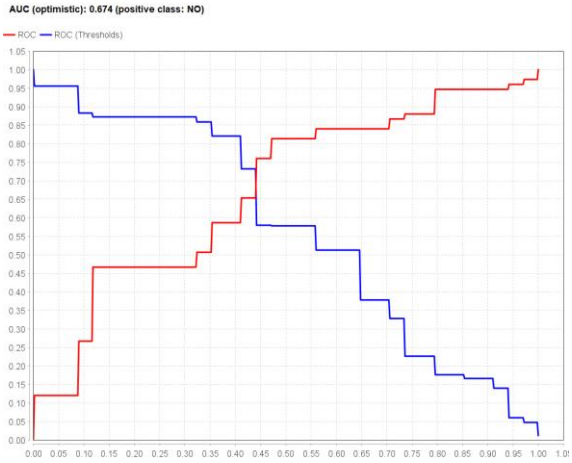

Figure 5. Area Under Curve using Naïve Bayes Algorithm

PerformanceVector

```
PerformanceVector:
accuracy: 71.56%
ConfusionMatrix:
True:    BO      NO
BO:      15      12
NO:      19      63
precision: 76.83% (positive class: NO)
ConfusionMatrix:
True:    BO      NO
BO:      15      12
NO:      19      63
recall: 84.00% (positive class: NO)
ConfusionMatrix:
True:    BO      NO
BO:      15      12
NO:      19      63
AUC (optimistic): 0.674 (positive class: NO)
AUC: 0.664 (positive class: NO)
AUC (pessimistic): 0.653 (positive class: NO)
```

Figure 6 . Performance Vector using Naïve Bayes Algo

The AUC value for classification using the Naïve Bayes Algorithm of 0.674 is included in the poor classification. AUC (Area Under Curve) is the area under the ROC curve. The model obtained is more accurate if the value is close to one. (Sayed & Burhanuddin, 2018)

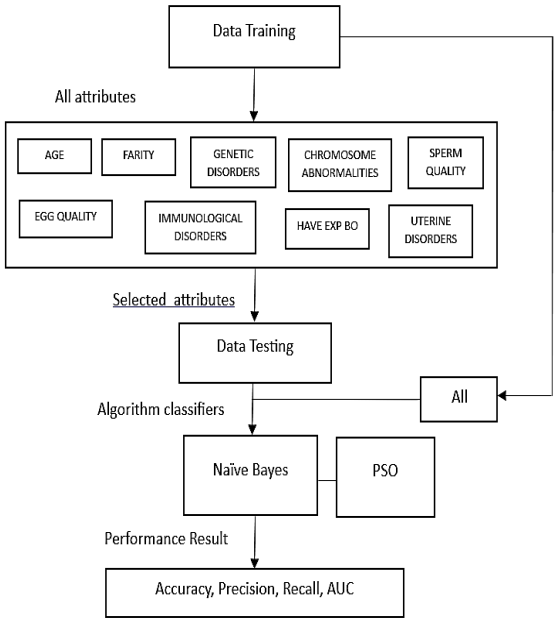

Figure 7. Classification Process Using the PSO-Bsed Naïve Bayes

Furthermore, the authors also classify the blighted ovum factor using the Naïve Bayes algorithm based on Particle Swarm Optimization to get a better accuracy value. PSO is one of the basic techniques of a swarm intelligence system to solve optimization problems in space search as a solution. PSO was first proposed by James Kennedy and Eberhart (1995) and designed to simulate birds searching for food. Swarm intelligence systems deploy innovative intelligence in solving optimization problems by taking inspiration from biological examples, such as the phenomenon of groups (swarms) in animals, where each group has individual behavior in carrying out joint actions to achieve the same goal. (Mansur et al., 2014) The Rapid Miner Framework process in Figure 7.

Table 4. Naïve Bayes Algorithm Accuracy Result

Accuracy 79.82%	True BO	True NO	Class Precision
Prediction BO	21	9	70.00%
Prediction NO	13	66	83.54%
Class Recall	61.76%	88.00%	

Table 4 shows the accuracy results using the PSO-based Naïve Bayes algorithm are 79.82%, with an actual YES recall class of 61.76% and a valid NO recall class of 88.00%.

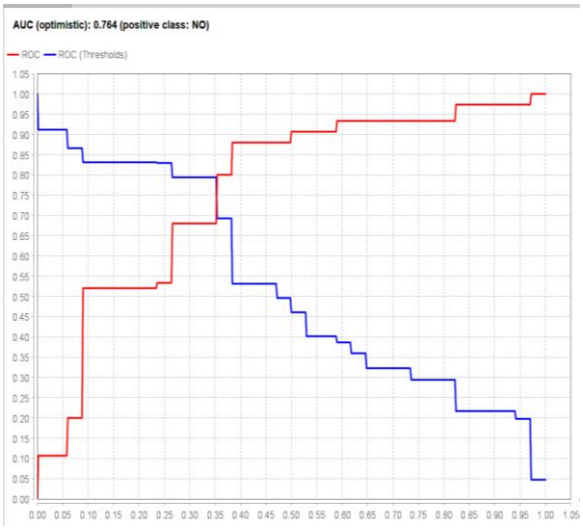

Figure 8. Area Under Curve results using the PSO-based Naive Bayes algorithm

PerformanceVector

```
PerformanceVector:
accuracy: 79.82%
ConfusionMatrix:
True:   BO      NO
BO:     21      9
NO:     13      66
precision: 83.54% (positive class: NO)
ConfusionMatrix:
True:   BO      NO
BO:     21      9
NO:     13      66
recall: 88.00% (positive class: NO)
ConfusionMatrix:
True:   BO      NO
BO:     21      9
NO:     13      66
AUC (optimistic): 0.764 (positive class: NO)
AUC: 0.756 (positive class: NO)
AUC (pessimistic): 0.749 (positive class: NO)
```

Figure 9. PerformanceVector result using the PSO-based Naive Bayes algorithm

Then in Figure 8, it can be seen that the result of the Area Under Curve is 0.764, which is included in the fair classification category. (Romi Satrio, 2020)

CONCLUSIONS AND SUGGESTIONS

Conclusion

The results of classifying the blighted ovum factor in pregnant women use the Naïve Bayes Algorithm with an accuracy value of 71.56% and an Area Under Curve of 0.674, included in the poor classification category. While the results of the classification using the Naïve Bayes Algorithm based on Particle Swarm Optimization with an accuracy value of 79.82% and an Area Under a Curve of 0.764 are included in the fair classification category. The Naïve Bayes Algorithm method is very suitable for data mining classification. If you want to get better accuracy and AUC values , use PSO as an optimization to determine the process parameters that produce the optimum response value.

Suggestions

In the future, it can be calculated using other algorithms and optimization methods to get better results, with better feature selection, to produce more accurate data.

REFERENCES

Adi Mhsd. (2021). *Proses KDD (Knowledge Discovery In Database)*. https://adi.unggas.id/lecture/proses-kdd-knowledge-discovery-in-database/

Anggrayni, N. H., Mas'udah, E. K., & Triningsih, R. W. (2022). Faktor Determinan Kejadian Blighted Ovum. *Jurnal Kebidanan*, *11*(2), 380. https://doi.org/https://doi.org/10.47560/keb.v11i2.380

Annisa Karnesyia. (2020). *Blighted Ovum*. https://www.haibunda.com/bundapedia/20220815101912-211-281708/blighted-ovum

Dr Rizal Fadli. (2022). *Blighted Ovum*. Halodoc.Com. https://www.halodoc.com/kesehatan/blighted-ovum

Han, J., & Kamber, M. (2006). *Data Mining Concepts and Techniques Second Edition*. Morgan Kaufmann.

Khadafy, A. R., & Wahoho, R. S. (2015). Penerapan Naive Bayes untuk Mengurangi Data Noise pada Klasifikasi Multi Kelas dengan Decision Tree. *Journal of Intelligent Systems*, *1*(2), 136–142.

Mansur, Prahasto, T., & Farikhin. (2014). Particle Swarm Optimization Untuk Sistem Informasi Penjadwalan Resource Di Perguruan Tinggi. *Jurnal Sistem Informasi Bisnis*, *01*, 11–19.

Normawati, D., Akbari, R., & Nurhusna, A. (2021). *Diagnosis Dini Penyakit Preeklamsia Pada Ibu Hamil Dengan Metode K-Nearest Neighbor (Knn)*. *13*(2), 69–78. https://www.unisbank.ac.id/ojs/index.php/fti2/article/view/8839

Rian Santoso, M., & Musa, P. (2021). Rekomendasi Kesehatan Janin Dengan Penerapan Algoritma C5.0 Menggunakan Classifying

Cardiotocography Dataset. *Jurnal Simantec*, *9*(2), 65–76.

Romi Satrio, W. (2020). *Data Mining*. https://romisatriawahono.net/dm/

Salsabila MR. (2023). *3 Tipe Proses Dataset dalam Machine Learning*. https://dqlab.id/3-tipe-proses-dataset-dalam-machine-learning

Sayed, F., & Burhanuddin. (2018). Penggunaan Metode Support Vector Machine Untuk Mengklasifikasi Dan Memprediksi Angkutan Udara Jenis Penerbangan Domestik dan Penerbangan Internasional Di Banda Aceh. *Jurnal Sistem Informasi*, *2*(2), 1–10.

Sianturi, F. A. (2019). Analisa Metode Teorema Bayes dalam Mendiagnosa Keguguran pada Ibu Hamil Berdasarkan Jenis Makanan. *Teknik Informasi Dan Komputer (Tekinkom)*, *2*(1), 87–92. http://jurnal.murnisadar.ac.id/index.php/Tekinkom/article/view/78

Sugiyono. (2013). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung. Alphabet.

Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D* (Cetakan 26). CV. Alfabeta.

Sulihati, I., Syukur, A., & Marjuni, A. (2022). *Deteksi Kesehatan Janin Menggunakan Decision Tree dan Feature Forward Selection*. *4*(3), 1658–1664. https://doi.org/10.47065/bits.v4i3.2672

Wibawa, A. P., Guntur, M., Purnama, A., Akbar, M. F., & Dwiyanto, F. A. (2018). Metode - Metode Klasifikasi. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, *3*(1), 134–138.

Yolanda, V., Cholissodin, I., & Adikara, P. P. (2021). Klasifikasi Diagnosis Penyakit Diabetes Gestasional pada Ibu Hamil menggunakan Algoritme Neighbor Weighted K-Nearest Neighbor (NWKNN). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *5*(4), 1310–1321.

This sheet is intentionally left blank