#### PAPER • OPEN ACCESS

# Comparison of Data Mining Algorithms Using Artificial Neural Networks (ANN) and Naive Bayes for Preterm Birth Prediction

To cite this article: Diah Puspitasari et al 2020 J. Phys.: Conf. Ser. 1641 012068

View the article online for updates and enhancements.



## IOP ebooks<sup>™</sup>

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection-download the first chapter of every title for free.

### **Comparison of Data Mining Algorithms Using** Artificial Neural Networks (ANN) and Naive Bayes for Preterm Birth Prediction

#### Diah Puspitasari<sup>1</sup>, Kresna Ramanda<sup>2</sup>, Adi Supriyatna<sup>3</sup>, Mochamad Wahyudi<sup>4</sup>, Erma Delima Sikumbang<sup>5\*</sup>, Sulaeman Hadi Sukmana<sup>6</sup>

<sup>1</sup> Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Indonesia

<sup>2</sup> Program Studi Teknik Informatika, Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri, Indonesia

<sup>3</sup>,<sup>5</sup> Program Studi Sistem Informasi Akuntansi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Indonesia

<sup>4</sup> Program Studi Teknologi Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Indonesia

 $^{6}$  Program Studi Sistem Informasi, Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri, Indonesia

E-mail: erma@bsi.ac.id

Abstract. Premature birth is still a big problem in Indonesia, in general, 15 million babies are born prematurely every year, more than 1 million babies die from complications due to premature birth. The main purpose of this study is to compare the Artificial Neural Network and Naive Bayes datamining algorithm models to predict preterm birth so as to obtain clinical evidence in preterm birth long before confinement so that sudden preterm birth can be converted to normal nativity. The model proposed in research on the prediction of preterm birth is by applying an Artificial Neural Network (ANN) algorithm and Naive Bayes algorithm. Where the two algorithms will be compared the level of accuracy and the value of the AUC against the prediction of preterm birth The results obtained that the prediction of preterm birth using the Artificial Neural Network (ANN) algorithm produces an accuracy value of 90.67% and an ROC value of 0.954. While the Naive Bayes algorithm produces an accuracy value of 84.53% and an ROC value of 0.929. For this reason, it can be concluded that the Artificial Neural Network (ANN) algorithm has a superior accuracy of 6.14% and 0.025 for its ROC value in predicting preterm birth.

#### 1. Introduction

Preterm birth is still a big problem in Indonesia, in general, 15 million babies are born prematurely every year, more than 1 million babies die from complications due to premature birth. According to the Health Technology Assessment Indonesia, preterm confinement is a birth that occurs in pregnant women whose gestational age is below 37 weeks [1][2][3]. There are many factors that can cause premature birth; they may include: previous preterm birth, race and body mass index [4]. Many cases of infant deaths occur due to premature birth besides that, premature birth also causes the baby has the possibility of physical disability. As many as

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Published under licence by IOP Publishing Ltd

7,000 newborns die worldwide each day while in Indonesia 185 per day, one of the main causes of death is premature. Increased confinement complications and increased risk of neonatal death, in some cases also causes maternal behavioral and psychological problems this is due to the delay in obtaining clinical evidence for preterm birth [5]. The short term, complications can occur in breathing, brain, heart, blood and the immune system. Complications can also occur in the long term in cognitive disorders, hearing loss, cerebral palsy, psychological to chronic health problems [6]. To reduce the number of preterm births and reduce neonatal deaths several studies have been conducted. The following are widely used algorithms: Decision Tree, Naive Bayes, Neural Networks, Rough Sets, K-Nearest Neighbor, Rule Based, Memory Based Reasoning, Support Vector Machines [7]. Artificial Neural Networks show effective results in creating general models [8]. Other studies comparing data mining algorithms are naïve Bayes and decision trees. The test results show that naïve Bayes has better accuracy on the decision tree algorithm [9]. Contribution of this study is to compare the Artificial Neural Network and Naïve Bayes datamining algorithm models to predict preterm birth so as to obtain clinical evidence in preterm birth long before confinement so that sudden preterm birth can be converted to normal nativity. At the prediction stage, the dataset is used to predict the learning model label. Examination of model performance is obtained from the comparison of datasets on predictive labels. At the classification stage, often using performance metrics, accuracy, loyalty, recall, etc. In regression problems, root mean squared error (RMSE) is used to measure errors. [10].

Research that has been conducted on premature births, such as, Artificial Neural Networks (ANN) is used as a screening tool for preterm birth in heterogeneous maternal populations; Risk estimation uses obstetric variables that are available to doctors before 23 weeks' gestation. The aim is to assess whether Artificial Neural Networks (ANN) have the potential to be used in estimating obstetric results in a population of low-risk mothers [8]. Soft computation techniques such as Softmax regression using Neural Networks and Gradient Descent Optimizers, Are used to practice the prediction model. The success rate of the prediction obtained is 89.99% with an average phased cost of 0.52. Therefore, this model has proven to be a reliable predictor of identifying high preterm risk in women, so that to plan the necessary antenatal and clinical interventions during pregnancy can provide sufficient time. [5]. This study intends to conduct research on how to predict patients who will give birth prematurely with the C4.5 model algorithm. The results obtained are the C4.5 algorithm produces an accuracy value of 93.60% and an AUC value of 0.946 for the diagnostic level of Very Good Classification [11].

#### 2. Research Methodology

#### 2.1. Artificial Neural Network

The structure that emulates and mimics biological neural networks is a computational model known as artificial neural networks. Artificial neural networks are systems that obtain inputs and outputs that are determined by complex relationships used in non-linear statistical data. Artificial neural networks can learn through the input and output produced, and returned to the network. [5]. In terms of classification, artificial neural networks (ANN) are effective in preterm birth research. The EHG signal is used in classifying terms in preterm birth. Artificial neural networks produce increased generalizations and recognition of the back-propagation algorithm and the resulting accuracy is 70.82% compared to other classification techniques.[12].

#### 2.2. Naive Bayes

Naïve Bayes is a classification method that is widely used based on Bayes theory. Based on the estimated conditional density of the class and the probability of the previous class, the probability of the posterior class from the test data point can be derived and the test data will be assigned to the class with the maximum posterior class probability. The main problem in the Naïve Bayes method is the estimation of class conditional density. Traditionally conditional

density classes are estimated based on data points. For uncertain classification problems, it is a must to study the conditional density of classes of uncertain data objects represented by probability distributions [13].

#### 2.3. Novelty description

The novelty in this paper is the attribute of the dataset used. The attributes are obtained from existing studies and confirmed by obstetricians and gynecologists.

#### 3. Result and Discussions

#### 3.1. Dataset

The dataset used was from two hospitals in Jakarta, The first hospital had 500 records gynecological patient data, and the second hospital had 250 records gynecological patient data , so that the total data used was 750 records consisted of 11 variables or attributes. There are variables which are classified as predictor variables or predictors which are used as determinants of preterm birth, and destination variables are variables which are used as birth results. The predictor variables are age, systole, diastole, history of high blood pressure, history of miscarriage, history of premature, stress (trauma), cigarette consumption, multiple pregnancy, vaginal discharge. The following is Table 1. Which describes the used attributes.

Attributes	Values
Age	0 = Not Prone; 1 = Prone
Systole	0 = Low; 1 = Normal; 2 = High
Diastole	0 = Low; 1 = Normal; 2 = High
History of high blood pressure	0 = Yes; 1 = No
History of miscarriage	0 = Yes; $1 = $ No
History of premature	0 = Yes; 1 = No
Stress (Trauma)	0 = Yes; $1 = $ No
Cigarette consumption	0 = Yes; 1 = No
Multiple pregnancy	0 = Yes; $1 = $ No
Vaginal discharge	0 = Yes; 1 = No
Results	Premature; Not Premature

 Table 1. Attributes of Gynecology Patients.

#### 3.2. Data mining Modelling Using Rapidminer

The model proposed in research on the prediction of preterm birth is by applying an Artificial Neural Network (ANN) algorithm and Naïve Bayes algorithm. Where the two algorithms will be compared the level of accuracy and the value of the AUC against the prediction of preterm birth. Application of Artificial Neural Network (ANN) algorithm and Naïve Bayes algorithm in prediction of preterm birth is using Rapidminer Studio applications or tools.

#### 3.3. Artificial Neural Network (ANN)

To measure the accuracy of the Artificial Neural Network (ANN) classification algorithm, the method used is confusion matrix. Following are the results of the confusion matrix value based on the dataset applied using Rapidminer Studio.

Accuracy=90.67%	True Premature	True Not Premature	Class Precision
Premature Pred.	335	32	91.28%
Not Premature Pred.	38	345	90.08%
Class Recall	89.81%	91.51%	

 Table 2. Confusion Matrix Table of Artificial Neural Network (ANN) Algorithm

Based on Table 2. shows that, the number of True Positive (TP) is 335 records classified as True Premature and False Negative (FN) classified as Premature but Not Premature is 32 notes. The next 345 records for True Negative (TN) are classified as Not Premature, and classified as Not Premature, but Premature is 38 False Positive Notes (FP). And the level of accuracy generated by using an Artificial Neural Network (ANN) algorithm is 90.67%.

#### 3.4. Naive bayes

To measure the accuracy of the Naïve Bayes classification algorithm, the method used is confusion matrix. Following are the results of the confusion matrix value based on the dataset applied using Rapidminer Studio.

Table 3. Confusion Matrix Table of Naive Bayes Algorithm

Accuracy=84.53%	True Premature	True Not Premature	Class Precision
Premature Pred.	281	24	92.13%
Not Premature Pred. Class Recall	$92 \\ 75.34\%$	353 93.63%	79.33%

Based on Table 3. shows that, the number of True Positive (TP) is 281 records classified as True Premature and False Negative (FN) is 24 records classified as Premature but Not Premature. Next 353 records for True Negative (TN) are classified as Not Premature, and 92 False Positive records (FP) are classified as Not Premature, but Premature. And the level of accuracy generated by using the Naïve Bayes algorithm is 84.53%.

#### 3.5. Performance Accuracy

Based on 750 records datasets that have been processed, using Artificial Neural Network (ANN) algorithm and Naïve Bayes algorithm in the prediction of preterm birth. The Table 4. Comparison of accuracy, specificity, sensitivity, NPV and PPV values in the two algorithms is produced.

Based on table 4 that the Artificial Neural Network (ANN) algorithm has an accuracy value of 90.67%, a sensitivity value of 84.14%, a specificity value of 87.74%, a PPV value of 88.20%, and an NPV value of 83.55%. While for the Naïve Bayes algorithm has an accuracy value of 84.53%, a sensitivity value of 92.13%, a specificity value of 79.33%, a PPV value of 75.34%, and an NPV value of 93.63%.

Value	ANN	Naive Bayes
Accuracy	90.67%	84.53%
Sensitivity	91.28%	92.13%
Specificity	90.08%	79.33%
PPV	89.81%	75.34%
NPV	91.51%	93.63%

 Table 4. Comparison of accuracy, specificity, sensitivity, NPV and PPV values

1641 (2020) 012068

#### 3.6. ROC Curve

The following are graphic images of ROC on Artificial Neural Network (ANN) algorithm and Naïve Bayes algorithm.



Figure 1. ROC Curve with Artificial Neural Network (ANN) Algorithm

Figure 1 illustrates the results of the Very Good classification diagnosis on the Artificial Neural Network (ANN) algorithm. With the Area Under Curve (AUC) value of 0.954 on the ROC graph.



Figure 2. ROC Curve with Naive Bayes Algorithm

Meanwhile, Figure 2 illustrates the results of the Excellent classification for the Naive Bayes algorithm. with an Area Under Curve (AUC) value of 0.929 on the ROC graph.

#### 4. Conclusion

Based on the research conducted on the prediction of premature birth, by comparing the accuracy value and the ROC value on two different algorithms. Namely, the Artificial Neural Network (ANN) algorithm and the Naive Bayes algorithm. The results obtained that the prediction of preterm birth using the Artificial Neural Network (ANN) algorithm produces an accuracy value of 90.67% and an ROC value of 0.954. While the Naive Bayes algorithm produces an accuracy value of 84.53% and an ROC value of 0.929. For this reason, it can be concluded that the Artificial Neural Network (ANN) algorithm has a superior accuracy of 6.14% and 0.025 for its ROC value in predicting preterm birth. further research can utilize optimization algorithms to increase accuracy.

#### References

- [1] Carolina I and Kresna R, 2018 Klasifikasi kelahiran prematur menggunakan algoritma c4.5 p. 668–672.
- [2] Esty A Frize M Gilchrist J and Bariciak E, 2018 Applying Data Preprocessing Methods to Predict Premature Birth in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2018 p. 6096–6099.
- [3] Kocak A T and Yilmaz A, 2018 Segmentation and classification of contractions in uterine 16 channels EMG signals for preterm birth prediction in 2018 26th Signal Processing and Communications Applications Conference (SIU) p. 1–4.
- [4] Goldenberg R L Culhane J F Iams J D and Romero R, Jan. 2008 Epidemiology and causes of preterm birth Lancet 371, 9606 p. 75–84.
- [5] Raghav H K V S Devi S P Rengaraj N and Thanranikumar E, 2018 Prediction of Preterm Pregnancies using Soft Computing Techniques Neural Networks and Gradient Descent Optimizer 2018 Int. Conf. Comput. Commun. Informatics March 2016 p. 1–4.
- [6] Janjarasjitt S, 2017 Evaluation of performance on preterm birth classification using single wavelet-based features of EHG signals in 2017 10th Biomedical Engineering International Conference (BMEiCON) 2017–Janua p. 1–4.
- [7] Gorunescu F, 2011 Data Mining Concepts, Models and Te chniques Springer.
- [8] Catley C Frize M Walker R C and Petriu D C, Jul. 2006 Predicting High-Risk Preterm Birth Using Artificial Neural Networks IEEE Trans. Inf. Technol. Biomed. 10, 3 p. 540–549.
- [9] Septiani W D, 2017 Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis J. Pilar Nusa Mandiri 13, 1 p. 76–84.
- [10] Pari R Sandhya M and Sankar S, 2017 Risk factors based classification for accurate prediction of the Preterm Birth in 2017 International Conference on Inventive Computing and Informatics (ICICI) Icici p. 394–399.
- [11] Puspita A and Wahyudi M, 2015 Algoritma C4.5 Berbasis Decision Tree untuk Prediksi Kelahiran Bayi Prematur Konf. Nasinal Ilmu Pengetah. dan Teknol. 1, 1 p. 97–102.
- [12] Idowu I O Fergus P Hussain A Dobbins C and Askar H Al, 2014 Advance Artificial Neural Network Classification Techniques Using EHG for Detecting Preterm Births in 2014 Eighth International Conference on Complex, Intelligent and Software Intensive Systems p. 95–100.
- [13] Ren J Lee S D Chen X Kao B Cheng R and Cheung D, 2009 Naive Bayes Classification of Uncertain Data in 2009 Ninth IEEE International Conference on Data Mining 60703110 p. 944–949.