COMPARISON OF K-NN AND NAÏVE BAYES CLASSIFIER FOR ASPHYXIA FACTOR

Finki Dona Marleny[1)], Mambang[2)]
[1, 2)]STIKES SARI MULIA BANJARMASIN
Jl. Pramuka No.2 Banjarmasin
e-mail: finkidona@gmail.com[1)], mmbg1283gmail.com[2]

ABSTRACT

Asphyxia is influenced by several factors, including the factors affecting the Immediate Was maternal factors That relates Conditions mother Pregnancy and childbirth such as hypoxia mother, Asphyxia factor data can be modeled using the classification approach. this paper will be compared k-nearest neighbor algorithm and Naive Bayes classifier to classify asphyxia factor. Naive Bayes uses the concept of Bayes' Theorem which assuming the independency between predictors. Basically, Bayes theorem is used to compute the subsequent probabilities. Analysis of the two algorithms has been done on several parameters such as Kappa statistics, classification error, precision, recall, F-measure and AUC. We achieved the best classification accuracy with KNN algorithm, 92,27%, for k=4. are lower than the rates achieved with Naïve Bayes 83,19%.

Keywords*: Naive Bayes, k nearest neighbor,Asphyxia, factor, Classifier*

I. INTRODUCTION

The World Health Organization reports that 4 to 9 million cases of newborn asphyxia occur each year. Of these, death accounts for 20% while a million who survived develops permanent neurological sequels such as mental retardation, cerebral palsy, speaking/hearing/visual and learning disabilities [1]. Asphyxia is a condition caused by insufficient oxygen intake, commonly found in infants[2]. The causes of asphyxia is deficiency in oxygen that occurs on the first day after birth [3]. Asphyxia is influenced by several factors, including the factors affecting the Immediate Was maternal factors That relates Conditions mother Pregnancy and childbirth such as hypoxia mother, maternal age less than 20 years or more than 35 years, parity, illnesses suffered by the mother such as hypertension, hypotension, Impaired uterine contractions and others[4]. Asphyxia factor data can be modeled using the classification approach, In classification,there is a target categorical variable,such as income bracket, which, for example, could be partitioned into three classes or categories: high income ,middle income, and low income [5].

The k nearest neighbor (k-NN) classifier is one of the most popular a method to perform the classification of objects based on the data of learning The closest distance to the object [6]. Nearest-neighbor classifiers are based on learning by analogy, that is, by comparing a given test tuple with training tuples that are similar to it. The training tuples are described by n attributes. Each tuple represents a point in an n-dimensional space. In this way, all of the training tuples are stored in an n-dimensional pattern space. When given an unknown tuple,a k-nearest neighbor classifier search esthe pattern space for the k training tuples that are closes to the unknown tuple. These k-training tuples are the k -nearest neighbors of the unknown tuple [7].

Naive Bayes combination introduced by Domingos and Pazzani also needs training to estimate the prior and posterior probabilities [8]. Naïve Bayesian classifiers assume that the effect of an attribute value on a given class is independen to the values of the other attributes. This assumption is called class conditional independence. It is made to simplify the computations involved and, in this sense, is considered "naïve." Bayesian belief networks are graphical models, which unlike naïve Bayesian classifiers, allow there presentation of dependencies among subsets of attributes. Bayesian belief networks can also be used for classification [7].

In this paper will be compared concerning the application of k-nearest neighbor and Naive Bayes classifier to classify asphyxia factor based on the data used in this research is secondary data that the data obtained from medical records at MAS Hospital Banjarrmasin.

II. RELATED WORK

The previous study of Infant Cries with Asphyxia by Azlee Zabidi et al [9] using Classification to investigates the performance of the Multilayer Perceptron (MLP) classifier in discriminating between healthy and infants with asphyxia from their cries, the accuracy of MLP classifier while reducing the computation load. The highest MLP

classification accuracy of 94% is obtained with 40 filter banks, 12 highly ranked MFC coefficients and 15 hidden nodes.

Sahak, Mansor, et al [10] using classification of infant cry with asphyxia using integration of Orthogonal Least Square and Support Vector Machine with Radial Basis Function kernel (OLS-SVM) and integration of Orthogonal Least Square with Multilayer Perceptron. Classification accuracy was computed to evaluate the performance of both methods. The OLS-SVM has produced high classification accuracy (94.34%) compared to OLS-MLP when C and l' were set to 1 and 0.013 respectively, and the selection of coefficients is 30% of 33 filter banks.

Charles C. Onu [11] explored the approach of machine learning in developing a low-cost diagnostic solution. With designed a support vector machine-based pattern recognition system that models patterns in the cries of known asphyxiating infants (and normal infants) and then uses the developed model for classification of 'new' infants as having asphyxia or not. Our prototype has been tested in a laboratory setting to give prediction accuracy of up to 88.85%.

Marleny, F,D, et al [12] in their study using the data asphyxia factor for classifying factors that affect the incidence of asphyxia. In this study, there are 13 factors used to support the MLP neural network method in classifying factors affecting asphyxia, The results of training with MLPNN using output layer activation function identities can only result in a classification accuracy of 84.3% and 81.5%. To test the accuracy obtained 87.7% and 84.4%. As for the MLP training results using sigmoid activation function output layer produces an accuracy of 94.5% and 91.6%. For the test results obtained better accuracy is 90.5% and 89.0%.

III. METHODOLOGY

A. DATA

The data for this study were obtained by medical records at MAS(Moh. Ansari Saleh) Hospital Banjarrmasin. Data is divided into several parts of the data on mothers, babies and childbirth complications are:

TABLE 1. Structure Of Dataset Data On Mother

No	Attribute	Description
1	Age	age of mother during childbirth, < 20, 20-35, > 35 years old
2	Parity	secure parity is the number of children born to mothers living parity 2-3, unsafe parity 1 and more than 3
3	disease story	Anemia, asthma, hypertension, diabetes mellitus (DM), no
4	Type of brith	Sectio Caesaria(SC), Pervaginam
5	Placenta	Previa placenta, retained placenta,

TABLE 2. Structure Of Dataset Data babies

No	Attribute	Description
1	Age	age of mother during childbirth, < 20, 20-35, > 35 years old
2	Parity	secure parity is the number of children born to mothers living parity 2-3, unsafe parity 1 and more than 3
3	disease story	Anemia, asthma, hypertension, diabetes mellitus (DM), no
4	Type of brith	Sectio Caesaria(SC), Pervaginam
5	Placenta	Previa placenta, retained placenta,

TABLE 3. Structure Of Dataset Data On childbirth complications

No	Attribute	Description
1	Multiple Pregnancy	Yes, No
2	Breech Birh	Yes, No
3	Prolonged second stage	Yes, No
4	Distocya	Yes, No

5	*Cepalo PelvicDisoroportion*	Yes, No
6	*premature rupture of membranes*	Yes, No

The final dataset consisted of 375 and divided into training data (80%), and test data (20%).

A. Proposed Model

a. K-NN (k-nearest neighbor)

K-Nearest Neighbor classifier searches the pattern space for the k training tuples that are closes to the unknown tuple. These k training tuples are the k nearest neighbors of the unknown tuple. "Closeness" is defined in terms of a distance metric, such as Euclidean distance.

The Euclidean distance between two points or tuples, say, X1 = (x11,x12,...,x1n) and X2 =(x21,x22,...,x2n), is

$$dist(X1, X2) = \sqrt{\sum_{i=1}^{n} (x1i - x2i)^2} \quad (1)$$

b. Naïve Bayes

Naïve Bayes is a statistical method based on Bayes theorem and potential to classify the data because of its simplicity. Naive Bayes classifier assumes that the presence of a particular feature in a class is not related to the presence of any other feature.

$$P(c/w)=[P(w/c)P(c)]/P(w) \quad (2)$$

where P(c/w) is probability of class c given word is w. P(c) is probability of class c and P(w) is probability of word w .
Naive Bayes classifier will be

$$c^*=arg\ max_c\ P(c/w) \quad (3)$$

c. Experiment

The data will be used 245 rows of data. A total of 70% were used for training and 30% testing accuracy. Label Results for asphyxia is the Medium and Heavy.
Naive Bayes uses the concept of Bayes' Theorem which assuming the independency between predictors. Basically, Bayes theorem is used to compute the subsequent probabilities. The analysis and results of applying the algorithm reveals in an AUC (Optimistic) of 0.647, AUC (Pessimistic) as 0.608, depicted in fig.1,2 and table 4.

The final dataset consisted of 375 and divided into training data (80%), and test data (20%).

TABLE 4. Naïve Bayes Classifier Result of validation performance

Validation Performance	***Result***
Classification error	16,81%
Kappa	0.177
Precision	58,33%
Recall	15.50%
F_measure	25,45%
Accuracy	83,19%

The best KNN classification rates were attained for asphyxia factors number of k=4. We used Euclidean distance metric to determine the best value of k to maximize the classification performance. The analysis and results of applying the KNN algorithm for k=4 reveals in an AUC (Optimistic) of 0.978, AUC (Pessimistic) as 0.871, depicted in fig.3,4 and table 5.

TABLE 4. Naïve Bayes Classifier Result of validation performance

Validation Perfomance	***Number of k***							
	K=1	***K=2***	***K=3***	***K=4***	***K=5***	***K=6***	***K=7***	***K=8***
Classification error	8,95%	8,57%	8,97%	7,73%	10,60%	8,95%	11,38%	12,60%
Kappa	0.691	0,640	0,665	0,672	0.592	0.636	0.583	0.492
AUC (optimistic)	0.992	0.985	0.977	0.978	0.967	0.952	0.938	0.913
AUC (Pessimistic)	0.750	0.831	0.845	0.871	0.873	0.872	0.838	0.834
Precision	73,43%	80,38%	73,05%	89,17%	73,67%	83,17%	70,00%	76,17%
Recall	81,00%	66,50%	74,00%	64,00%	64,00%	64,00%	66,50%	48,50%
F_measure	74,24%	73,42%	71,59%	71,03%	64,96%	68,23%	64,70%	55,83%
Accuracy	91,02%	91,47%	91,03%	92,27%	89,40%	91,05%	88,62%	87,40%

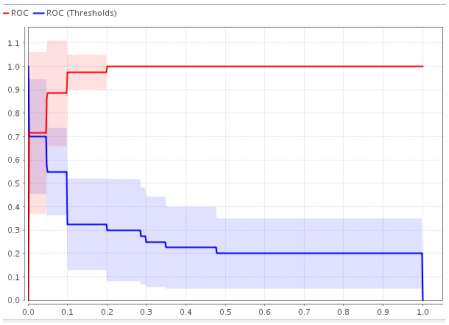

Fig. 3 AUC (Optimistic)

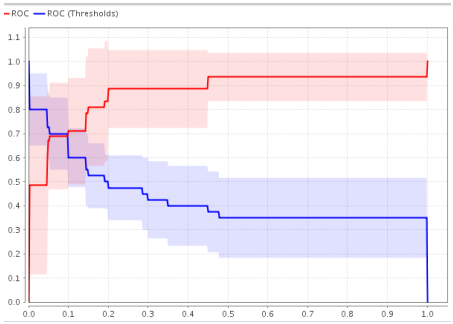

Fig. 4 AUC (Pessimistic)

IV. RESULT

We achieved the best classification accuracy with KNN algorithm, 92,27%, for k=4. Analysis of the two algorithms has been done on several parameters such as Kappa statistics, classification error, precision, recall, F-measure and AUC. A comparison between k-NN classifier and Naïve Bayes, computed for the asphyxia factor, is presented in table 6.

TABLE 6. Comparison result

Validation Performance	***Naïve Bayes***	***k-NN***
Classification error	16,81%	7,73%
Precision	58,33%	89,17%
Recall	15.50%	64,00%
F_measure	25,45%	71,03%
accuracy	83,19%	92,27%
Kappa	0.177	0.672
AUC(Optimistic)	0.647	0.978
AUC(Pessimistic)	0.608	0.871

V. CONCLUSION

The results of the study show that our proposed model is very helpful in classification model. We can see that the best percentages of accuracy obtained with k-NN 92,27%, for k=4, are lower than the rates achieved with Naïve Bayes 83,19%. We obtained better classification rates with k-NN for almost all of the subjects. In conclusion, the classifier used in asphyxia factor can be improved to obtain better accuracy and must be subject oriented.

REFERENCES

[1] C. f. P. Office of Health and Nutrition, Health and Nutrition, Bureau for Global Programs, Field Support and Research, U.S., "Detecting and Treating Newborn Asphyxia," Maternal Neonaltal & Health.

[2] Low, J. A., Muir, D. W., Pater, E. A., Karchmar, and E. Jane, "The association of intrapartum asphyxia in the mature fetus with newborn behavior," American Journal of Obstetrics and Gynecology, pp. 11311135, 1990

[3] J. A. Low, "Intrapartum fetal asphyxia: definition, diagnosis, and classification," American Journal of Obstetrics and Gynecology, vol. 176, pp. 957-959, 1997

[4] Wiknjosastro, Hanifa. "Ilmu kebidanan." Jakarta: Yayasan Bina Pustaka Sarwono Prawirohardjo (2005): 45-51.

[5] Larose, Daniel T. Discovering knowledge in data: an introduction to data mining. John Wiley & Sons, 2014.

[6] Wang, Bing, Yong Zeng, and Yupu Yang. "Generalized nearest neighbor rule for pattern classification." Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress on. IEEE, 2008.

[7] Han, Jiawei, Jian Pei, and Micheline Kamber. Data mining: concepts and techniques. Elsevier, 2011.

[8] DOMINGOS,P.andPAZZANI,M.(1997):OntheoptimalityofthesimpleBayesianclassifier under zero- loss, Machine Learning, 29, 103-130.

[9] Zabidi, Azlee, et al. "Classification of infant cries with asphyxia using multilayer perceptron neural network." Computer Engineering and Applications (ICCEA), 2010 Second International Conference on. Vol. 1. IEEE, 2010.

[10] Sahak, R., et al. "Performance of combined support vector machine and principal component analysis in recognizing infant cry with asphyxia." Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE. IEEE, 2010.

[11] Onu, Charles C. "Harnessing infant cry for swift, cost-effective diagnosis of Perinatal Asphyxia in low-resource settings." Humanitarian Technology Conference-(IHTC), 2014 IEEE Canada International. IEEE, 2014.

[12] Marleny, Finki Dona, et al. "Klasifikasi Faktor Yang Mempengaruhi Asfiksia Menggunakan Multilayer Perceptron Neural Network." Proceedings Konferensi Nasional Sistem dan Informatika (KNS&I) (2015).