International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT206657

Developing an Expert System Application to Detect Childs' Lung Disease

Sulis Sandiwarno
Department of Computer Science, Universitas Mercu Buana University, Indonesia
sulis.sandiwarno@mercubuana.ac.id

Article Info
Volume 6, Issue 6
Page Number: 285-290
Publication Issue :
November-December-2020

Article History
Accepted : 24 Nov 2020
Published : 10 Dec 2020

ABSTRACT

The development of information technology has supported many activities, especially in terms of health. Artificial Intelligence (AI) is the application of information technology that is currently developing well. Several previous studies have evaluated models from expert systems to diagnose lung disease in children using Naïve Bayes (NB) and Support Vector Machine (SVM). However, in conducting these evaluations they do not try to make an integrated application to facilitate evaluation. In this study we propose to build a system that integrates NB and SVM classifiers. Furthermore, in this study we used a sample of data from a clinic in Indonesia. The results of this study, we conclude that the existence of this system will make it easier to evaluate the lung disease experienced by children.

Keywords: Artificial intelligence, NB, SVM, Lungs

I. INTRODUCTION

The application of information technology today is the focus of life that leads to the advancement of science education. The development of information technology is supported by the lives of everyone who wants to use this information technology to help solve existing problems [1–3]. Information technology development in the world of health has a very big role in everyday life in solving existing problems.

Expert system is a computer program that contains knowledge from one or more human experts regarding a specific field [4]. The general form of an expert system is a program based on a set of rules that analyzes information (usually provided by the user of a system) about a specific class of problems as well as a mathematical analysis of the problem. In the standard system, there are two commonly used models, namely forward chaining rules and backward chaining rules [4]. Forward chaining is forward tracking that starts from a set of facts by looking for rules that match the existing assumptions / hypotheses to conclusions. Meanwhile, backward chaining is backward tracking that starts reasoning from the conclusion (goal), by looking for a set of hypotheses to the facts that support a set of these hypotheses.

Previous research has conducted an evaluation of expert systems in the field of health, especially for the lungs in children by applying models from machine learning such as Naïve Bayes (NB) and Support Vector Machine (SVM) [5, 6]. Although previous research has implemented various types of classifiers such as NB and SVM, they have not considered creating a system that can make it easier to conduct research evaluations simultaneously. Therefore, in this study we propose to design a system that can perform simultaneous evaluation by utilizing NB and SVM classifiers in the detection of lung disease in children.

II. RELATED WORK

In this chapter we divide references based on our research topic, namely NB and SVM classifiers in detecting Childs' lung disease.

2.1 Naïve Bayes (NB) to Detect Childs' Lung

Naïve Bayes (NB) is of the one most elegant machine learning technique that is practically used. NB is an efficiency, effectiveness, and iterability algorithm to classify the data [7–9]. Previous studies have used NB in evaluating lung symptoms in children [4]

2.2 Support Vector Machine (SVM) to Detect Childs' Lung

SVM is one of the successful methods to develop classification in supervised learning. In addition, Huajuan [10] illustrated SVM useful to apply data classification in data mining and machine learning, because SVM has successfully solved and predicted the problem in high dimensionality of data classification such as text categorization with excellent performances. Previous studies have used SVM in evaluating lung symptoms in children [5, 6].

III. PROPOSED MODEL

The flow work of this research divided into six steps such as design of knowledge representation, prototyping system, validation expert, testing of program, results.

3.1 Design of knowledge representation

a. Design of knowledge representation model that was employed on this system based on the production rule using pattern of IF – THEN. Each lungs' symptom has determined the weight value (confidence factor) that was defined by the domain expert within range 0....1. This value represents the confidence value of each lungs' symptom which causing particular diseases.
b. The lungs' disease diagnosis adopted a forward chaining inference. This system allows users to select the symptoms of the infected kids' lung. The user could be selected the symptoms textual statement and sample image in the expert system. The system will process the choice of users and give the evaluation results of the infected kids' lung.

3.2 System Prototyping

The expert system, for diagnosing lung disease, consists of lung disease diagnosis, knowledge-based, inference engine and database. The expert system architecture can be shown in Figure 2. In research we adopted PHP program language with MySQL as database. Inference engine contains thought mechanism and system reasoning pattern used by an expert. This mechanism will analyze a symptom selected by the user for producing the conclusion of diagnosis result. This inference system used forward chaining method.

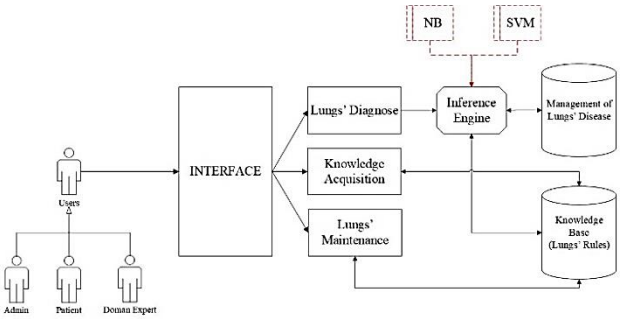

Figure 1. The framework of Proposed Model

3.2.1 Naïve Bayes Classifier

Given the test description of the document d of an opinion represented by the vector $< w_1, w_2, \ldots\ldots, w_m >$, to classify the document d, MNB is defined as:

$$C_{MNB^{(d)}} = P(c) \prod_{i=1}^{n} P(w_i|c)^{fi}$$

where, $P(c)$ is a prior probability that a document d belongs to class c, n is a number of the features, $P(w_i|c)$ is the conditional probability that a word w_i occurring in the class c, w_i is the word feature occurred in d, fi is the number of frequency count of a word w_i in reporting d, and $C_{MNB^{(d)}}$ is the class label of d predicted by the classifier [9].

3.2.2Support Vector Machine (SVM)

SVM is a machine learning technique which is used in prediction, classification, and regression. The ith opinion in SVM trained with all of the opinions in the ith class with the positive labels, then all other opinions with negative and neutral labels. Given l as training data $(x_i, y_i), \ldots, (x_n, y_n)$, where $x_i \in R^l$ and $Y_i \in \{1,2 \ldots\ldots, c)$ describe an opinion class of x_i. To classify the document x_i, SVM is defined as:

$$\min_{w_m \in H, b \in R^k, \xi \in R^{l\,x\,k}} \frac{1}{2}\sum_{m=1}^{k} w_m^T w_m + C \sum_{i=1}^{l} \sum_{t \neq y_i} \xi_{i,t} \quad (2)$$

$$subject\ to\ w_{y_i}^T \varphi(x_i) + b_{yi} \geq w_{y_i}^T \varphi(x_i) + b_t + 2 - \xi_{i,t} \quad (3)$$

$$\xi_{i,t} \geq 0$$
$$i = 1, \ldots, l, t \in \{1, \ldots, k\} \backslash y_i$$

where, in training the opinions data X_i is illustrated to highest dimensional space by the function φ and C is presenting the penalty parameter. Minimizing $\frac{1}{2}\sum_{m=1}^{k} w_m^T w_m$ describes we shall like to maximize $\frac{2}{||w^i||}$ the margins between three groups of opinions data. If the data training is not linear distinguishable, there is penalty term $C \sum_{i=1}^{l} \sum_{t \neq y_i} \xi_{i,t}$ that can be reduced the total number of training error. For the summary, the concept of SVM is finding for the balance between the rule term of $\frac{1}{2}\sum_{m=1}^{k} w_m^T w_m$ and training errors [11–13].

3.3 Validation expert

The expert system validation is the stage which the experts re-examine the design of a system that has been developed by the authors. Validation of the prototype expert system is done by two senior doctors of clinic at Indonesia.

IV. RESULTS

The results of the research we have done can be seen from the sub-chapters which will be explained in detail below.

4.1 Diagnosis Results

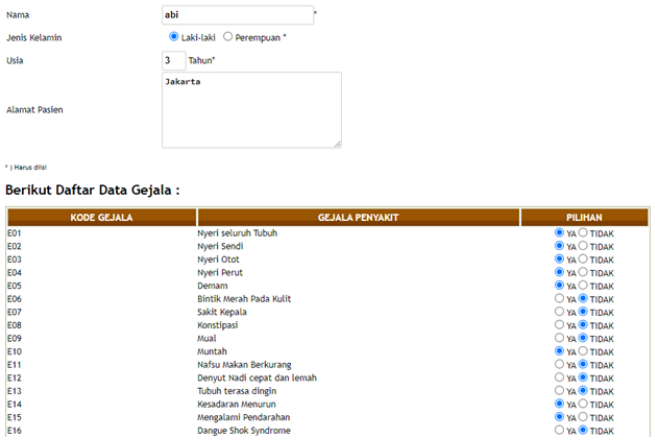

Figure 2. The diagnosis results

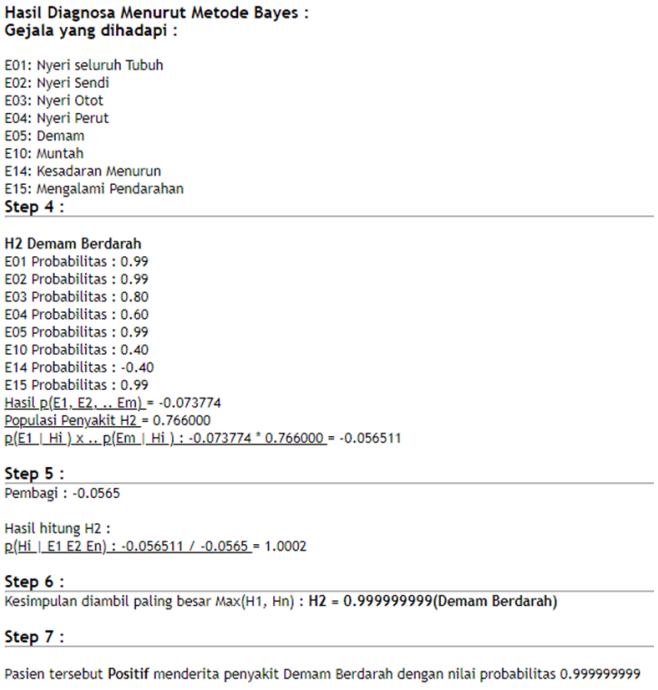

Figure 3. The NB diagnosis results

4.2 Classification Results

This section will explain the results of the research that has been done in evaluating kids' lung diseases. The results of data classification carried out by NB and SVM classifiers are then calculated using several techniques, such as: precision, recall, F1 and accuracy.

TABLE I

DATA CLASSIFICATION RESULTS FROM SVM

Fold (#)	Accuracy (%)	Recall (%)	Precision (%)	F1 (%)
1	**80.05**	**66.41**	71.90	69.05
2	**77.49**	**66.00**	72.79	69.23
3	**78.97**	**71.43**	72.41	71.92
4	**82.82**	**72.03**	79.23	75.46
5	**78.72**	**66.67**	77.04	71.48
6	**78.21**	**67.36**	71.85	69.53
7	**80.77**	**79.26**	69.48	74.05
8	**78.72**	**66.91**	71.54	69.14
9	**77.95**	**79.03**	62.03	69.50
10	**81.28**	**75.52**	73.97	74.74
Average	**79.50**	**71.06**	72.22	71.41

From the results obtained from the two methods, it can be seen that NB is the best method of classifying the data in this study. The accuracy value obtained from NB is 80.32% while SVM is 79.5% with a difference of 8.2%. From the fold (#) value that has been done that the smallest result of NB is in the second iteration with a value of 77.49%, and the highest value is 82.82 in the 4th iteration. Meanwhile, the smallest value in SVM is in the 2nd iteration with a value of 78.77% and the highest value is 82.05% in the 4th iteration. The Recall value at NB is 76.66% and the value at SVM is 71.06%. with a difference in value of 5.6 Meanwhile, the value of fold (#) in NB displays the lowest result in the 7th iteration with a value of 75.49%, while in SVM is 66% in the second iteration.

TABLE 2

DATA CLASSIFICATION RESULTS FROM NB

Fold (#)	Accuracy (%)	Recall (%)	Precision (%)	F1 (%)
1	**81.33**	**76.00**	**81.82**	**73.06**
2	**78.77**	**76.53**	**83.09**	**73.14**
3	**81.79**	**77.65**	**84.14**	**77.46**
4	**82.05**	**76.90**	**83.85**	**75.69**
5	**79.74**	**76.57**	**86.67**	**74.76**
6	**77.69**	**76.48**	**77.78**	**70.71**
7	**82.31**	**75.49**	**83.12**	**78.77**
8	**80.00**	**76.71**	**78.46**	**72.34**
9	**79.23**	**77.28**	**77.85**	**75.23**
10	**80.26**	**76.99**	**82.88**	**75.86**
Average	**80.32**	**76.66**	**81.96**	**74.70**

If seen in the 2nd iteration NB displays the results of 76.53%, then there is a difference of 10.53%, while in the 7th iteration SVM displays the results of 79.26% with a value of 3.77% difference from NB. In the precision section, NB displays the results of 81.96% and SVM 72.22% with a difference in value of 9.74%. Whereas in section F1, NB displays the data classification results of 74.7% and SVM displays the results of 71.41% with a difference in value of 3.29%. based on the results of data classification that has been done by both methods, the biggest difference in value occurs in precision.

V. CONCLUSION

In this paper, an empirical study was conducted to evaluate the application of expert system to detect the kids' lung based on Naïve Bayes (NB) and Support Vector Machine (SVM). Based on Naïve Bayes is the best algorithm to classify the users' opinions in this study, the accuracy reported is 80.32% and 79.50% of SVM classifier. To The findings of this study reveal that NB classifier outperformed than SVM classifier in evaluating the kids' lung disease.

VI. REFERENCES

[1]. Sadikin M, Fanany MI, Basaruddin T (2016) A New Data Representation Based on Training Data Characteristics to Extract Drug Name Entity in Medical Text. Comput Intell Neurosci. https://doi.org/10.1155/2016/3483528

[2]. Sadikin M (2017) Mining relation extraction based on pattern learning approach. Indones J Electr Eng Comput Sci. https://doi.org/10.11591/ijeecs.v6.i1.pp50-57

[3]. Triana YS (2018) Monte Carlo Simulation for Modified Parametric of Sample Selection Models Through Fuzzy Approach. In: IOP Conference Series: Materials Science and Engineering

[4]. Kurniawan R, Yanti N, Ahmad Nazri MZ, Zulvandri (2015) Expert systems for self-diagnosing of eye diseases using Naïve Bayes. In: Proceedings - 2014 International Conference on Advanced Informatics: Concept, Theory and Application, ICAICTA 2014

[5]. de Carvalho Filho AO, Silva AC, de Paiva AC, et al (2017) Lung-Nodule Classification Based on Computed Tomography Using Taxonomic Diversity Indexes and an SVM. J Signal Process Syst. https://doi.org/10.1007/s11265-016-1134-5

[6]. Naqi SM, Sharif M, Yasmin M (2018) Multistage segmentation model and SVM-ensemble for precise lung nodule detection. Int J Comput Assist Radiol Surg. https://doi.org/10.1007/s11548-018-1715-9

[7]. Wang S, Jiang L, Li C (2015) Adapting naive Bayes tree for text classification. Knowl Inf Syst 44:77–89. https://doi.org/10.1007/s10115-014-0746-y

[8]. Balamurugan AA, Rajaram R, Pramala S, et al (2011) NB+: An improved Naïve Bayesian algorithm. Knowledge-Based Syst 24:563–569. https://doi.org/10.1016/j.knosys.2010.09.007

[9]. Jiang L, Zhang L, Yu L, Wang D (2019) Class-specific attribute weighted naive Bayes. Pattern Recognit 88:321–330. https://doi.org/10.1016/j.patcog.2018.11.032

[10]. Huang H, Wei X, Zhou Y (2018) Twin support vector machines: A survey. Neurocomputing 300:34–43. https://doi.org/10.1016/j.neucom.2018.01.093

[11]. Weston J, Watkins C (1999) Support Vector Machines for Multi-Class Pattern Recognition. Proc 7th Eur Symp Artif Neural Networks

[12]. Hsu CW, Lin CJ (2002) A comparison of methods for multiclass support vector machines. IEEE Trans Neural Networks 13:415–425. https://doi.org/10.1109/72.991427

[13]. He X, Wang Z, Jin C, et al (2012) A simplified multi-class support vector machine with reduced dual

optimization. Pattern Recognit Lett 33:71–82. https://doi.org/10.1016/j.patrec.2011.09.035

Cite this article as :

Sulis Sandiwarno, "Developing an Expert System Application to Detect Childs' Lung Disease", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 6, pp. 285-290, November-December 2020. Available at doi : https://doi.org/10.32628/CSEIT206657
Journal URL : http://ijsrcseit.com/CSEIT206657