Model Expert System for Diagnosis of Covid-19 Using Naïve Bayes Classifier

D Silahudin¹, Henderi², A Holdin³

¹,³Master of Information Technology University of Raharja Indonesia
²Correspondence Author, Department of Informatics Engineering, University of Raharja

E-mail: henderi@raharja.info²

Abstract. This paper offers an expert system model for COVID-19 diagnosis as an effort to overcome the spread of COVID-19 in Indonesia. The expert system model was built using the Naive Bayes Classifier method. Model development is carried out with preliminary research stages, data collection, analysis, model design, implementation, and testing. The data used to build and test the model comes from the health department and the acceleration of the Covid-19 countermeasure group in Indonesia. The model was developed with a unified modeling language and a prototyping approach. Tests show that the developed COVID-19 diagnosis system expert model can diagnose COVID-19 based on the symptoms inputted by the user into the system. The application of the model produced in this study helps assist doctors in diagnosing COVID-19.

1. Introduction

A novel coronavirus which is label as Covid-19, caused an outbreak in the city of Wuhan, China, and it has further spread to other parts of China and many other countries in the world [1]. The common signs of infection include respiratory symptoms, fever, coughing, shortness of breath, and difficulty breathing. In more severe cases, the infection can cause pneumonia, acute respiratory syndrome, kidney failure, and even death. And technology is increasingly becoming a massive part of today's healthcare scenario. Technology has changed the way how patients communicate with the doctor and not only that, but also healthcare is administered [2]. The early prediction and evaluation of disease severity are extremely important for patient prognosis. But in this paper, we propose to investigate the Naïve Bayes event model for an expert system to diagnose Covid-19 because it has not been considered for this problem before. In response to the outbreak, we summarize the current knowledge of Covid-19 and compare it with previous experiences of the SARS outbreak in Hong Kong studying effective measures to control the Covid-19 epidemic [3].

In Indonesia, until March 17, 2020, there were 134 confirmed cases spread in eight provinces, namely Bali, Banten, DKI Jakarta, West Java, Central Java, West Kalimantan, North Sulawesi, and Yogyakarta [4]. In Indonesia, medical personnel who are experts in the field of Covid-19 disease are
still limited, both in terms of number and working time. In resolving the Covid-19 disease attack in various provinces and districts in Indonesia.

So, an expert system for diagnosis Covid-19 disease is needed. It is expected to help the doctor to overcome the problem by providing a good solution. The process of making this expert system the Bayes Classifier theorem certainty method where the method is based on the initial conditions where the initial conditions are conditions of existing phenomena then subject to predetermined rules and then the greatest truth value is taken to determine the conclusions and solutions of the phenomena mentioned earlier.

But the expert system research is about detecting Covid-19 disease, in contrast to the expert system using the Bayes Classifier theorem is to diagnose diseases in humans and the symptoms that cause these diseases. Based on the description above, the author is interested in making an expert system in the hope that making this expert system work with the system and benefit the community.

The system can assist the doctor's work in dealing with special patients who have Covid-19 disease. Now, the computer-aided diagnostic system exists has a significant role in them. A faster and accurate system is necessary. So, in this research, we present our model an expert system for diagnosis of the Covid-19 epidemic in Indonesia.

The rest of this paper is organized as follows. Section 2 presents a general overview of related work about expert systems in medical or disease, Naïve Bayes method in related research, and writing state of the art. Section 3, explains about research method that including the research framework and the Naïve Bayes Classifier attribute. In Section 4, presents about results and discusses experimental results. Finally, a general conclusion of this work is presented in Section 5.

2. Related work.

One of the major issues that need to be considered when talking about an expert system model problem is the fixed method of the model that developed. But, the other issues like fields and models also important to be considering. So, there are some related work about the application for medical fields, the expert system in the medical field, Naïve Bayes method, Covid-19, and so on.

Including the research about Naïve Bayes Classifiers for authorships [5], artificial intelligence in the retina and beyond ocular disease [6], diagnosis of coronavirus disease [7], and predict the Covid-19 epidemic [1]. In this group, Naïve Bayes method for an expert system model for diagnosis the Covid-19 has not covered and discussed.

The other research made reported that computer-aided diagnostic system exist has a significant role in them [8]. It used the classification and feature selection method for the medical aided system. The other journal published about retrieval-based diagnostic Aid [9], development and use of a clinical decision support system [10], and research about diagnostic for supporting primary health care system [2]. In this research group also has not discussed an expert system for diagnosis Covid-19 using Naïve Bayes Classifier. So, our research has state of the art in the field and the methods in the system. It was about an expert system for the diagnosis of Covid-19 using Naïve Bayes Classifier

3. Research Methods

3.1. Research Framework

For the research to be more directed, a research framework is needed as an illustration of the research conducted. The research framework used is the flow of software development, because this research produces a system that can be used by the community. The research framework has six stages. It is preliminary research, data collection, analysis, design, implementation, and testing.

Stages of research carried out starting from preliminary research that is identifying problems by visiting the website www.who.int to look for common problems that occur about Covid-19 disease. After the problem is found, data collection is done by finding information from interviews with doctors to obtain valid knowledge. Then the data is analyzed and processed by using the Bayes Classifier method. The design carried out is to make use case diagrams as an illustration of the
relationship of actors with the system. The resulting diagram is translated into program form using the Java programming language and MySQL database. Finally, testing is done online so that everyone can consult with an expert system.

3.2. Naive Bayes Classifier

Naive Bayes Classifier is a probability classification based on the Bayes Theorem. Bayes’ theorem will be combined with “Naïve” which means that each attribute or variable is independent. Naïve Bayes Classifier can be trained efficiently in supervised learning [11].

\[
P(H|X) \quad \text{: The probability of hypothesis } H \text{ is based on condition } X \quad \text{(posterior probability)}
\]

\[
P(H) \quad \text{: Hypothesis probability } H \quad \text{(prior probability)}
\]

\[
P(X|H) \quad \text{: Probability } X \text{ is based on the conditions in hypothesis } H
\]

\[
P(X) \quad \text{: Probability } X
\]

Characteristics of Naive Bayes Classifier [11]:

1) Toughness faces irrelevant attributes
2) Attributes that correlate can degrade classification performance.
3) Naïve Bayes because the assumption of independence of these attributes is gone.

Naïve Bayes classifier calculation: 

\[
\text{nc} = \text{the number of records in the learning data v = Vj} \\
\text{a = ai ;p = } 1/\text{many types of classes / diseases}. \\
\text{m = number of parameters / symptoms; n = number of records in the learning data. Equation 2 solved [9]:}
\]

1) Determine the value nc for each class.
2) Then calculate the value \( P(\text{ai | Vj}) \) and calculate values \( P(\text{Vj}) \).

\[
\text{VMAP} = \arg \max Vj \in V P(\text{Vj}) \pi P(\text{ai | Vj }) \quad \text{......... (3)}
\]

\[
P(\text{Ai | Vj }) = \frac{nc+m_p}{n+m} \quad \text{......... (2)}
\]

3) Calculate \( P(\text{ai | Vj }) \times P(\text{Vj}) \) for each v.
4) Determine the classification results, namely v which has the greatest multiplication results

4. Results And Discussion

The problem discussed in this study is about Covid-19 disease. Data requirements in an expert system are data that is used in identifying problems as knowledge acquisition. The following are the data obtained from interviews with experts, regarding data on symptoms and types of diseases as well as the rules set forth addressed in Table 1, Table 2, and Table 3 [7]. The rules in Table 3 are developed using a similar approach in the other research about rules-based [12]. The rule base constructed in this research also can be used to construct of patient information track model [13].

Some symptoms of Covid-19 in Table 2 adopted from systematic disorder [14]. Systematic disorder has used in similar research about SAR but it also may use in the research about Covid-19 because both epidemic has some the same symptoms. In Covid-19, the systematic disorders including are fever, cough, and fatigue. Then, these rules are implemented to application likes in other research [15], and expert clinical decision support system [16].
Table 1. Emerging Symptoms of Every Disease

| #  | Disease Code | Types of Diseases |
|----|--------------|-------------------|
| H1 | P1           | PDP               |
| H2 | P2           | ODP               |
| H3 | P3           | OTG               |

From the 3 tables above as a reference for taking data samples in testing the calculation of manual Naïve Bayes Classifier. The test aims to determine the level of accuracy of the results of the calculation of the system diagnosis in this research which is calculated manually.

Table 2. Examples of Disease Symptoms

| #  | Symptom Code | The Symptoms                                                                 |
|----|--------------|-------------------------------------------------------------------------------|
| 1  | G1           | Fever or history of fever                                                    |
| 2  | G2           | Symptoms and signs of respiratory distress (cough, cold, sore throat, fatigue)|
| 3  | G3           | Severe Pneumonia or Acute Respiratory Infections (ARI)                       |
| 4  | G4           | There are no other causes based on convincing clinical descriptions          |
| 5  | G5           | In the last 14 days before the symptoms have a history of travel or living    |
|    |              | abroad who reported local transmission                                       |
| 6  | G6           | In the last 14 days before the symptoms have a history of travel or stay in   |
|    |              | the local transmission area in Indonesia                                     |
| 7  | G7           | Contact with Coronavirus Disease 2019 (COVID19) confirmation cases In the     |
|    |              | last 14 Days before symptoms                                                 |

Table 3. The rules

| #Rule | IF (Condition) | Then (Output) |
|-------|----------------|---------------|
| R1    | G1,G2,G3,G4,G5 | P1            |
| R2    | G1, G4, G6    | P2            |
| R3    | G7            | P3            |

Examples of calculations using the Naïve Bayes Classifier classification can be applied to those experiencing symptoms g2 = Fever or history of fever, g3 = Severe Pneumonia or Acute Respiratory Infections (ARI), g6 = In the last 14 days before the symptoms have a history of travel or stay in the local transmission area in Indonesia, g7 = Contact with Covid-19 confirmation cases in the last 14 days before symptoms. The steps for calculating the Naïve Bayes Classifier are as follows.

1. Determine the value of nc for each class
   If a symptom is included in a class of disease, then nc will be 1, if not then 0. Known:
   a) Disease 1 PDP
      Value of symptoms for each class (n) = 1
      The value of symptoms is divided by many classes of disease.
      (p) = 1/3 = 0.333
      Total symptoms (m) = 7
      g2=1, g3=1, g6=1, g7=0
   b) Disease 2 PDP
      Value of symptoms for each class (n) = 1
      The value of symptoms is divided by many classes of disease.
      (p) = 1/3 = 0.333
      Total symptoms (m) = 7
      g2=1, g3=1, g6=1, g7=0
   c) Disease 3 OTG
      Value of symptoms for each class (n) = 1
      The value of symptoms is divided by many classes of disease.
      (p) = 1/3 = 0.333
      Total symptoms (m) = 7
      g2=0, g3=0, g6=0, g7=1

2. Calculating values P(ai|vj) and calculate value P(vj).
   At this stage, the probability value of 1 symptom will be calculated. Starting from the first disease class, namely PDP, then ODP, then OTG. The PDP category is symbolized by the variable (P1). To calculate the value of the probability, then it is used, (ai | vj) and P (vj) then:
   a) Disease 1 PDP
      (g2|p1) = (1+7+0.3)/(1+7)
   b) Disease 2 ODP
      (g2|p2) = (0+7+0.3)/(1+7)
   c) Disease 3 PDP
      (g2|p3) = (0+7+0.3)/(1+7)
3. Calculate $P(a_i | v_j) \times P(v_j)$ for each grade-$v$

a) PDP 2st Covid-19 Disease.
\[
= [P(P1) \times P(G|P2P1) \times P(G3|P1) \times P(G6|P1) \times P(G7|P1)]
= 0.333 \times 1.038 \times 1.038 \times 1.038 \times 0.913 = 0.340
\]

b) ODP 2nd Covid-19 Disease.
\[
= [P(P2) \times P(G2|P2) \times P(G3|P2) \times P(G6|P2) \times P(G7|P2)]
= 0.333 \times 0.913 \times 1.038 \times 1.038 \times 0.913 = 0.299
\]

c) OTG 3rd Covid-19 Disease.
\[
= [P(P3) \times P(G2|P3) \times P(G3|P3) \times P(G6|P3) \times P(G7|P3)]
= 0.333 \times 0.913 \times 0.913 \times 0.913 \times 1.038 = 0.263
\]

4. Determine the results of the classification namely $v$ which has the greatest multiplication value.

| Disease | Score V |
|---------|---------|
| PDP     | 0.3400  |
| ODP     | 0.2991  |
| OTG     | 0.2631  |

The largest $v$ value is 0.3400. It can be concluded that the user belongs to the PDP category.

4.1. Design and Implementation

The model of an expert system for diagnosis of Covid-19 was developed using the use case diagram. It was illustrated about the relationship between the actor and the system [17]. In this expert system, the actor is Manager of Expert System and Member. The implementation of model an expert system for diagnosis Covid-19 program in this research using Java programming and MySQL Server for database. The results of the expert system program, tested online by advising patients to consult with the expert system, to get an initial diagnosis of Covid-19 disease suffered by the patient.
Figure 3 and 4 illustrates one of the expert system interfaces. Figure 4 is a consultation page, where on this page the patient enters the symptoms he feels. The results from this research in Figure 4 and Figure 5 are like similar research to diagnose Covid-19 earlier and to improve its treatment by applying medical technology [18].

The following figure provides the results of the consultation carried out by the patient. The results from an expert system model for diagnosis Covid-19 in Figure 5 is an early diagnosis thus increasing the survival rate and preventing demolition caused due to Covid-19 like a similar research about diagnosis [9]. So, this research is a part of the research study which addresses the design of a decision support system for diagnosing the severity of SAD is recommended [10], and in general it using the concept of decision support system [19].

This paper demonstrated the simple expert system model, rule-based, and the dataset of newly diagnosed cases in Indonesia for diagnosis of Covid-19 using Naïve Bayes method. This research linkage between the data from the patients and the rule-based, which result in the system expert pages (Figure 5). The result is a little answer which is in the other paper about the Covid-19 [1], and it can be used by the government in Indonesia to make some decisions about Covid-19.

5. Conclusion

This research has proposed a model expert system for the diagnosis of Covid-19 using Naïve Bayes Classifier. It is useful and can help the community in recognizing Covid-19 disease, especially Government in Indonesia. The application of the expert system in this research provided the information on early symptoms of Covid-19 disease in patients previously unknown to the general public. It can tracing and elucidate the results of the diagnosis and provide better. The testing conducted by patients, the expert system for diagnostic Covid-19 results obtained from a disease based on facts.

6. References

[1] Roda W.C., Varughese M.B., Donglin Han., and Michael Y. Li. Why is it difficult to accurately predict the COVID-19 epidemic? Infectious Disease Modelling 5 (2020) 271-281.
[2] Kidwai B. and Nadesh RK. Design and Development of Diagnostic Chabot for supporting Primary Health Care System. Procedia Computer Science 167 (2020) 75-84.
[3] KOMPAS.COM, 2020. Update: Add 17 Patients, There Are 134 Positive Cases of Coronavirus. KOMPAS.COM. Access on 2 April 2020.
[4] S. Law, A.W Leung, and C. Xu Severe acute respiratory syndrome (SARS) and coronavirus disease-2019 (COVID-19): From causes to preventions in Hong Kong, International Journal of Infectious Diseases 94 (2020) 156–163.
[5] Altheneyen A.S., and Menai M.E.B. Naïve Bayes Classifiers for authorship attribution of Arabic texts. Journal of King Saud University -- Computer and Information Sciences (2014) 26, 473–484.
[6] Schmidt-Erfurth S., Sadeghipour A., Gerendas B.S., Waldstein S.M., and Bogunovic H. Artificial intelligence in retina. Progress in Retina and Eye Research 67 (2018) 1-29.
[7] W. Wiguna and D. Riana. (2020) Diagnosis Of Coronavirus Disease 2019 (Covid-19) Surveillance Using C4.5 Algorithm Jurnal PILAR Nusa Mandiri Vol. 16, No. 1.
[8] Tuba E., Strumberger I., Bezdan T., Bacanin N., and Tuba M., Classification and Feature Selection Method for Medical Datasets by Brain Storm Optimization Algorithm and Support Vector Machine. Procedia Computer Science 162 (2019) 307-315.
[9] Singh L., Janghel R.R., and Sahu S.P. Designing a Retrieval-Based Diagnostic Aid using Effective Feature to Classify Skin Lesion in Dermoscopic Image. Procedia Computer Science 167 (2020) 2172-2180.

[10] Sina Fathi, Maryam Ahmadi, Behrouz Birashk, Afsaneh Dehnad. Development and use of a clinical decision support system for the diagnosis of social, anxiety disorder. Computer Methods and Programs in Biomedicine (2020).

[11] Han J., Kamber M., and Pei J. Data Mining Concepts and Techniques. 2012. Morgan Kaufmann Publishers, Elsevier Inc. All right reserved.

[12] Henderi, Aini Q., Santos N.P.L., Faturahman A., and Rahardja U., 2020. A Proposed Gamification Framework for Smart Attendance System Using Rule Base. Jour of adv research in dynamical & control systems, vol. 12, Issue-02, 2020

[13] Shakhovska N., Fedushko S., Gregus ml M., Melnykova N., Shvorob I, and Syerov Y. Big Data analysis in development of personalized medical system. Procedia Computer Science 160 (2019) 229-234.

[14] Law S., Leung A.W., and Chuanshan Xu. Severe acute respiratory syndrome (SARS) and coronavirus disease-2019 (COVID-19): From causes to preventions in Hong Kong. International Journal of Infection Disease 94 (2020) 156-163.

[15] Henderi, Aini Q., Srienggini A.D. and Khoirunisa A. Rule Base Expert System for Supporting Assessment of Learning Outcomes. International Journal of Advanced Trends in Computer Science and Engineering Vol. 9, No.1.2, 2020. pp. 266-271.

[16] H. Emrana, Z. Shahid and H. Rokibul. An Expert Clinical Decision Support System to Predict Disease Using Classification Techniques. 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE). Bangladesh. pp 396-400.

[17] Henderi. 2010. Object oriented modeling with unified modeling language. Raharja Enrichment Centre (REC), STMIK Raharja.

[18] Bai L., Yang D., Wang X., et all. Chinese experts' consensus on the Internet of Thing-aided diagnosis and treatment of coronavirus disease 2019 (COVID-19). Clinical eHealth 2 (2020) 7-15.

[19] Asidik, I., Kusrini, Henderi. Decision support system model of teacher recruitment using algorithm C4.5 and Fuzzy Tahani. Journal of Physics: Conference Series, vol:1140 issue: 1, 2018-12-21.