



Implementation of Expert System for Diabetes Diseases using Naïve Bayes and Certainty Factor Methods

Muhammad Ilham Insani¹, Alamsyah², Anggyi Trisnawan Putra³

^{1,2,3}Computer Science Departement, Faculty of Mathematics and Natural Sciences,
Universitas Negeri Semarang, Indonesia

Email: ¹muhammadilhaminsani@gmail.com, ²alamsyah@mail.unnes.ac.id,

³anggyitrisnawan@mail.unnes.ac.id

Abstract

Expert Systems is a computer systems that has been entered the base knowledge and a set of rules used to solve problems like an expert. Methods that can be used in the expert systems which is Naïve Bayes and Certainty Factor. Naïve Bayes method can handle quantitative calculations and discrete data and only requires a little research data to estimate the parameters needed in the classification and Certainty Factor which is suitable for measuring something whether it is certain or not in diagnosing. Diabetes is one of the most frequent diseases suffered in Indonesia. The purpose of this research is implementation expert systems used Naïve Bayes and Certainty Factor in diagnosing diabetes and knowing the level of accuracy of the systems. Data that is used by researchers as much 100 data medical record, obtained from the medical record RSUD Bendan Kota Pekalongan. The variables used in this research is age, gender, the symptoms of the disease diabetes and result diagnose disease from expert. The accuracy rate of this system derived from the scenario distribution data 70 training data and 30 testing data that is equal to 100% according to the doctor's diagnosis.

Keywords: Expert Systems, Disease Diabetes, Naïve Bayes, Certainty Factor.

1. INTRODUCTION

Expert systems is part of high-level specialized software or high-level programming language, which is trying to duplicate the functionality of an expert in a particular field of expertise [1]. Usually expert system build to cope with the problems compound by performing reasoning using knowledge resembles the expert [2]. To overcome various problems, expert systems combine the knowledge base and inference engine [3]. The intent of the combining the knowledge base and the inference is adopted human knowledge into the computer, so the computer in solving problems such as commonly performed by an expert [4]. In its application the problems faced by expert system is not just a problem that rely on algorithm, but also problems that are difficult to understand. [5]. Although it is difficult to understand the expert system is trying to find the completion of satisfactory, that is settlement which is quite good so the job can be run although it's not optimal completion, in other words, expert system is a system-based software which makes or evaluates decisions based on rules

set in the software [6]. The goal of expert systems is not to replace the role of humans, but a variety of material of human knowledge which is presented in the form so that it can be in use by many people [7]. From some previous research of expert system give good result to finish problems that use complex data in the case of prediction or diagnosis, example diagnosis of diseases of the diabetes mellitus using Naïve Bayes method using 140 data obtained an accuracy rate of 100%, disease in cats using Naïve Bayes and Certainty Factor method with 25 data test obtained 80% accuracy rate, disease in goats using Naïve Bayes and Certainty Factor obtained accuracy rate of 86.80% [8-10].

Naïve Bayes form a certain probabilistic simple yang berdasarkan pada teorema Bayes in generally, inference Bayes especially with the assumption of the independence of the strong [11]. Process that is classification with method probability and statistic that is predict opportunities in the future based on the experience in earlier times and to look for degrees of believe using Certainty Factor who have clinical parameters to show the value of trust [12]. This is the factor of certainty in the contents by experts together with the rules and factor of certainty that is given by the user.

Diabetes is a disease of the metabolic characteristics hyperglikemia that occurs because of abnormalities of insulin secretion, insulin action or both. Diabetes is also accompanied by metabolic disorders of carbohydrates, fats, and proteins caused by a deficiency of the hormone insulin relative or absolute. [13]. When it is in let uncontrollable can occur complications of the metabolic and accurate complications of vascular long-term, better mikroangiopati and makroangiopati [14]. Diabetes needs in beware of because of the nature of the disease which is a chronic progressive number of patients who increasing and a lot of negative that caused.

Based on the above description, the purpose of this research is implementation expert system with method Naïve Bayes and Certainty Factor and to know accuracy in diagnose disease diabetes.

2. METHODS

2.1. Naïve Bayes

Naïve Bayes a classifier is simply based on theorem Bayes combine with “Naïve” which means each attribute is independent to calculate each class of decision calculate probability with the condition that the class is true, given vector object information [15]. The stage of the calculation Naive Bayes as follows:

- 1) Find the value of prior probability for each class calculate the average of each class.
- 2) Find the value of likelihood that is a process of calculating the probability of each attribute against the class, the possibility of emergence of a class when an attribute is selected.
- 3) Find the value of posterior that is result of calculation likelihood in the form of the probability of the attribute class, calculated to divert the possibility of

the attribute of the input with the class, in the process of this can be probability of the end.

Naïve Bayes method can handle quantitative calculations and discrete data and only requires a little research data to estimate the parameters needed in the classification.

2.2. Certainty Factor

Certainty Factor a method to accommodate questions from an expert in the delivery of the result of the analysis describe the level of confidence the expert to the problem being faced, [16]. Calculation Certainty Factor can be formulated in formula (1).

$$CF[H, E] = MB[H, E] - MD[H, E] \quad (1)$$

Explanation:

CF[H,E] : Certainty Factor from hypothesis H which is affected by symptoms (evidence) E. Biggest CF revolve between -1 until 1. value 1 show belief unconditional, while value -1 show disbelief unconditional.

MB[H,E] : measured of increased belief (measure of increased belief) about the hypothesis H that is influenced by symptoms of E

MD[H,E] : measured of increased disbelief (measure of increased disbelief) about the hypothesis H that is influenced by symptoms of E.

Certainty Factor in an expert system can measure something whether it is certain or not in diagnosing the disease and in calculating the certainty factor only uses two data so that the accuracy of the data can be maintained.

2.3. Data Collection

The data used in this study is the result of medical records of 100 patients data year of 2016 and 2017 who suffer from particular disease in whom type 1 diabetes, type 2, gestational, and type in take in RSUD Bendan Pekalongan, data that is used in the form of age, gender, symptoms experienced patients in RSUD Bendan and the doctor's diagnosis of the doctor responsible.

2.4. Interview

Interview done by internal disease specialist doctor who is the doctor in RSUD Bendan, from the results of the interviews produced the weights of each symptom in diabetes is show in Table 1.

Table 1. The value of the weight symptoms

No	Sympton Data	Type of disease diabetes				weight
		Type 1	Type 2	Gestasional	Special type	
1	Age < 20 year	V				0,8
2	Age > 20 year		V			0,8
3	Parents of patients with type 1 diabetes	V				0,5
4	Parents of patients with type 2 diabetes		V			0,5
5	Obesity > 20% from normal weight		V	V		0,6
6	In the pregnant			V		0,9
7	Post Radiation				V	0,9
8	Post chemotherapy				V	0,9
9	The onset of tingling(numbness) or pain in the hands and feet	V	V	V	V	0,4
10	Often feel thirsty (polidipsia)	V	V	V	V	0,7
11	Frequent urination (poliuris)		V	V	V	0,7
12	Excessive hunger (polifagia)		V	V	V	0,7
13	Fall weight without aclear reason.	V	V			0,4
14	High blood sugar levels > 140 mg/dl (high) Protein levels of Cpeptide low Insulin levels in the body high	V				0,9
15	High blood sugar levels > 140 mg/dl (high) Protein levels of Cpeptide high Insulin levels in the body low		V	V	V	0,9
16	Easy to feel tired	V		V	V	0,4
17	Breath smelling like fruit	V				0,8
18	The eyes less obviously		V		V	0,3

19	The onset of ulcers(sores) on the leg that never healed.	V	V	V	V	0,7
20	Nausea /vomiting/abdominal pain	V	V	V	V	0,3
21	Dry Mouth	V	V	V	V	0,2
22	Feel ichting	V	V	V	V	0,1
23	Previously never been sick a particular disease and consuming drugs that can distrust blood sugar levels, insulin as wellas protein in the body				V	0,8
24	The body feels weak/ often sleepy	V	V	V	V	0,1
25	Head feel dizzy	V	V	V	V	0,1

Explanation:

P1: Diabetes Type 1

P2: Diabetes Type 2

P3: Diabetes Gestasional

P4: Diabetes Special Type

2.5. Design Process

There is *flowchart* program shown in the Figure 1.

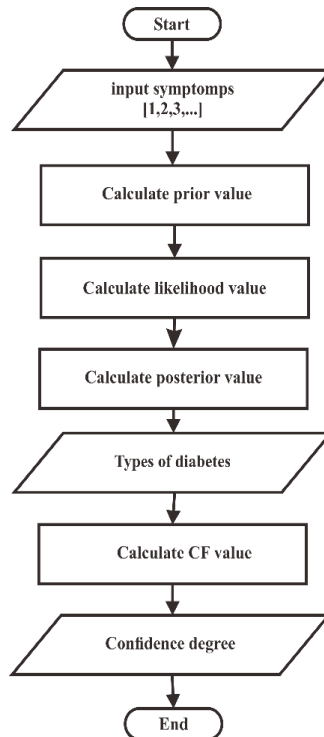


Figure 1. Flowchart program

3. RESULT AND DISCUSSION

3.1. Method

The stages of the calculation data at the start of calculation Naïve Bayes then continue the calculation Certainty Factor.

1) Naïve Bayes

Learning process model Naïve Bayes using 70 sample data training medical records of diabetic patients from RSUD Bendan. In this process consist of several sub processes as follow:

a) Calculation of the prior

A process where compare the many members of a class with the overall sample.

Example:

There are 10 patients disease diabetes type 1 on the data training so the process of calculation as follows.

$$\text{Prior} = \frac{10}{70} = 0,142$$

b) Calculation of the likelihood

A calculation of the probability value of each attribute against the class, the possibility of the emergence of a class when an attribute is selected, the calculation process for finding the likelihood as follows.

There are attributes G1 → Type 1 = 3 class

The overall type 1 = 10 class

then, Likelihood = $\frac{3}{10} = 0,3$

c) Calculation of the posterior

Based on the result calculation of likelihood in the form of probability attribute to the class, calculated by means of controlling the possible attributes of the input with the class, in this process obtained the probability of the end for the conclusion the results of the classification

The diagnosis with Naïve Bayes was calculated based on certainty factor as follows:

2) Certainty Factor

The calculation of value factor (Certainty Factor) from every input symptoms user. The class that was selected from the classification process is implementation on the Certainty Factor to find the value of certainty, this process is repeated for each of two symptoms of the input using the following formula:

$$CF_{1,2} = CF_1 + CF_2(1 - CF_1) \quad (2)$$

Example:

input user = G1 with CFuser = 0,8
 G2 with CFuser = 0,25

$$\begin{aligned} CF_{G1,G2} &= CF_{G1} + CF_{G2}(1 - CF_{G1}) \\ &= 0,8 + 0,75(1 - 0,8) \\ &= 0,8 + 0,75(0,2) \\ &= 0,8 + 0,15 \\ &= 0,95 \end{aligned}$$

Naïve Bayes process was obtained based on the calculation of the Certainty Factor.

3) The Result of Diagnosis

The user accepts the conclusion of a diagnosis was based illness carried out with the help of the value of certainty factors. These result have been compared with testing data that is taken based on the diagnosis of the reality of the doctor on duty at RSUD Bendan, and produce accuracy equal to 100% for the testing data in use, so the result is valid to be receive as the conclusion of the diagnosis the user

3.2. Testing

After testing the 100 data medical records of patients of the disease diabetes in RSUD Bendan Pekalongan divided into 2 that is 70 training data and 30 testing data then get the results by the accuracy of the 100% showed on Table 2.

Table 2. Result of the testing data

Amount of data <i>testing</i>	Data in accordance	Data does not match	Accuracy
30	30	-	100%

Based on the result of the data medical record used the system that has been made produce result 30 *testing* data accurate with the doctor's diagnosis.

4. CONCLUSION

Method of Naïve Bayes has 3 stages namely the research for prior, find the value of likelihood, and looking probabilitasi end of posterior after obtaining the classification results of the calculation Naïve Bayes, the class that was selected from the classification process Naïve Bayes implemented in the calculation of Certainty Factor to find the value of certainty. After calculation Certainty Factor it will get the value of the certainty of the class selected on the calculation of the Naïve Bayes. After implementation 100 medical records data taken from RSUD Bendan Pekalongan, divided into 2 that is 70 training data and 30 testing data, it can the level of accuracy of the results of the diagnosis system is much as 100% accurate in accordance with the diagnosis the expert.

5. REFERENCES

- [1] Hustinawaty & Aprianggi R. (2014). The Development of Web Based Expert System for Diagnosing Children Diseases Using PHP and MySQL. *International Journal of Computer Trends and Technology (IJCTT)*, 10(4): 197-202.
- [2] Naik, M.V. & Lokhanday S. (2012). Building a Legal Expert System for Legal Reasoning in Specific Domain-A Survey. *International Journal of Computer Science & Information Technology (IJCSIT)*, 4(5): 175.
- [3] Pramesti, A.A., Arifudin R., & Sugiharti E. (2016). Expert System for Determination of Type Lenses Glasses using Forward Chaining Method. *Scientific Journal of Informatics*, 3(2): 177.
- [4] Kusumadewi, S. (2003). *Artificial Intelligence (Teknik dan Aplikasinya)*. Yogyakarta: Graha Ilmu.
- [5] Setyabudi, W.U., Sugiharti E., & Arini F.Y. (2017). Expert System Diagnosis Dental Disease Using Certainty Factor Method. *Scientific Journal of Informatics*, 4(1): 44.

- [6] Josephine, M.S. & Jeyabalaraja V. (2012). Expert System and Knowledge Management for Software Developer in Software Companies. *International Journal of Information and Communication Technology Research*, 3(2): 243.
- [7] Muslim, M.A., Kurniawati I., & Sugiharti E. (2015). Expert System Diagnosis Chronic Kidney Disease Based on Mamdani Fuzzy Inference System. *Journal of Theoretical and Applied Information Technology*, 78(1): 70.
- [8] Prayoga, ND., Hidayat N., & Dewi RK. 2017. Sistem Diagnosis Penyakit Hati Menggunakan Metode Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2 (8): 2666-2671.
- [9] Ferdiansyah, W. R., Muflikhah, L., & Adinugroho. (2018). Sistem Pakar Diagnosis Penyakit Pada Kucing Menggunakan Metode Naive Bayes dan Certainty Factor. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(2): 451-458.
- [10] Suprayogi, AA., Hidayat, N., & Fanani, L. (2018). Sistem Pakar Diagnosis Penyakit Kucing Menggunakan Metode Naïve Bayes – Certainty Factor Berbasis Android. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(2): 650-658.
- [11] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*(3nd ed). Waltham: Morgan Kaufmann.
- [12] Kusriani., (2008). APLIKASI SISTEM PAKAR. *Menentukan Faktor Kepastian Penggunadengan Metode Kuantifikasi Pertanyaan*. Yogyakarta: Andi.
- [13] American Diabetes Association (ADA). (2014). Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*, 37(1): 2-4.
- [14] Darmono, dkk. (2007). *Naskah lengkap diabetes melitus di tinjau dari berbagai aspek penyakit dalam* . Semarang: Badan penerbit universitas Diponegoro
- [15] Olson & Delen. (2008). *Advanced Data Mining Techniques*. USA: Springer-Verlag Berlin Heidelberg.
- [16] Sutojo. (2010). *Kecerdasan Buatan*. Yogyakarta: Penerbit Andi.