

Accuracy of Algorithm C4.5 to Study Data Mining Against Selection of Contraception

by Fakultas Teknik

Submission date: 19-Jan-2024 02:38PM (UTC+0700)

Submission ID: 2273774956

File name: Prosiding_6.pdf (214.97K)

Word count: 2087

Character count: 11152

Accuracy of Algorithm C4.5 to Study Data Mining Against Selection of Contraception



Des Suryani, Ause Labellapansa and Eka Marsela

Abstract Family planning is an attempt to infuse or plan the number and spacing of pregnancy using contraception. Contraception is a method or tool used to prevent pregnancy. Arrangements can be made with the use of contraceptives such as pills, spirals, implants. Patients may consult with the midwife or physician in choosing a contraceptive that is suitable or convenient to use. But in reality there are still many patients, especially patients at Clinic Pratama Hasanah Pekanbaru, who still hesitate in choosing contraceptives in accordance with the patient's body condition. For that, writer was interested to do study data of contraception usage to patient at Clinic Pratama Hasanah Pekanbaru to get pattern of decision tree and its level of accuracy. Based on this, it is necessary to evaluate the data collection of contraceptive usage to determine the pattern of contraceptive selection. The data mining process uses the classification method with the Decision Tree C4.5 algorithm. The attributes used in determining the pattern of selection of contraceptives consist of nine regular attributes, i.e., age, term of usage, menstrual cycle, just married, just give birth, breastfeeding, already having offspring, health problems, and have more than four children, whereas the attribute label/class that is contraceptive used. This study resulted in recognition of contraceptive pattern selection with accuracy of 93.15% (excellent classification). With this level of accuracy, it will help the midwife to direct the patient in the selection of contraceptives based on the resulting pattern.

Keywords Data mining · Decision Tree C4.5 · Classification · Contraception

Suryani (✉) · A. Labellapansa · E. Marsela
Department of Information Technology, Universitas Islam Riau,
Pekanbaru, Indonesia
e-mail: des.suryani@eng.uir.ac.id

© Springer Nature Singapore Pte Ltd. 2018
R. Saian and M. A. Abbas (eds.), *Proceedings of the Second International Conference on the Future of ASEAN (ICoFA) 2017 – Volume 2*,
https://doi.org/10.1007/978-981-10-8471-3_95

1 Introduction

Information technology is growing so rapidly and touching on all sides of life. In order for human activities more easily and efficiently, many researchers do collaboration between computer science disciplines and other disciplines. One of the collaborations took place between computer science and medical science to choose a contraceptive tool for family planning (FP) acceptors.

Contraception is a method or tool used to prevent pregnancy. In other words, KB is planning the number of families. Arrangements can be made with the use of contraceptives such as pills, syringes, spirals. Patients can consult a physician and then choose a contraceptive device that is suitable or convenient to use.

However, despite consultation with a doctor, the choice of methods or contraceptives is not an easy thing. This is because effects that affect the body will not be known before the contraceptive is used. Each method has advantages and disadvantages. Despite the advantages and disadvantages, all contraceptives are available; women still find it difficult to control fertility safely, effectively, with acceptable methods. It is not surprising that women feel that the use of contraception is sometimes problematic and forced to choose contraceptives that do not match the adverse consequences of or not using the FP method at all.

In Clinic Pratama Hasanah Pekanbaru, the patient in choosing contraceptives especially with consultation with midwife or nurse who will identify by giving question to patient.

Based on the above background, the authors are interested to analyze the data of contraceptive device selection to Clinic Pratama Hasanah Pekanbaru patients to get the pattern of decision tree and the level of accuracy. The method used is to apply the method of classification with algorithm C4.5. With a better level of accuracy, it will help midwives and the public to know the appropriate contraceptive and appropriate body condition.

2 The C4.5 Algorithm for Building Decision Trees

Planning in having children is one of the important things in a family. To anticipate the occurrence of pregnancy, couples usually do family planning (FP) program. Some contraceptives used for the family planning program include pills, syringes, spirals. But determining contraceptives is not an easy task.

Usually, couples should consult a doctor- or midwife-related contraceptive to be used. However, this does not mean this way without risk, because doctors do the diagnosis based on the knowledge they have. In one condition, a midwife or doctor may be mistaken when suggesting an appropriate contraceptive for a woman. As a result, women feel that the use of contraception is sometimes problematic because they have to choose a method that does not match.

The **selection of** contraceptives in Clinic Pratama Hasanah Pekanbaru is done by way of the society must fill the registration form, then nurse will identify by giving question to patient. From the results of this identification, the doctor will advise contraceptives that are considered appropriate to the condition of the patient. Although the midwife is an expert in this field, in reality they also have limited memory and stamina work. This could have made a mistake in identifying the appropriate contraceptive device. But until now, there are still many people who do not fit with the method or contraceptives because the situation and condition of the body of each individual will continue to change.

Nugroho (2014) have used the k-nearest neighbor algorithm in the classification method. This research successfully applied k-nearest neighbor algorithm to calculate and give result of a recommendation class to choose contraception method.

Trisnawarman and Erlysa (2007) designed a decision support system that can assist in determining appropriate methods or devices of contraception and in accordance with the situation and condition of the body of each user.

Kamaludin (2012) has created a decision support system for choosing contraceptives using the simple additive weighting method where Kamaludin uses age, cost, side effect, duration, tool efficacy, menstrual history, potential complications, health benefits, and health status.

Differences with the research made are to dig the data in a training data in the form of contraceptive device selection data with age attributes, duration of use, menstrual period, new marriage, new childbirth, breastfeeding, having offspring, health problems, children more than four, and label attributes of contraceptive devices while the results of analysis in the form of contraceptive pattern selection.

In general, the C4.5 algorithm for building decision trees is as follows:

1. Select an attribute as root
2. Create a branch for each value
3. For the case in the branch
4. Repeat the process for each branch until all the cases on the branch have the same class.

An attribute is selected as a root based on the highest gain value of the attributes. To calculate the gain used, the formula is shown in Eq. (1):

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \quad (1)$$

While the calculation of entropy value can be seen in the following Eq. (2):

$$\text{Entropy}(S) = \sum_1^n -p_i * \log_2 p_i \quad (2)$$

where

S The Set of Cases

A Features

N Number of partitions S

P The proportion of S_i to S

3 Process Modeling and Knowledge Outcomes

Data of clinical contraceptive use Pratama Hasanah Pekanbaru in 2015 and 2016. Data performed data cleaning. Then the data is done to determine the selection of data to be analyzed. Selection result data is data that has been validated. In this case amounted to 146 records such as Table 1. The data is continued to the process of data transformation; that is, the data is converted into the appropriate format for processing in data mining. The result of data transformation process (dataset) is tested using C4.5 algorithm through RapidMiner software. The results of this test obtained the decision tree as in Fig. 2.

After experimenting with data processing using Decision Tree C4.5 algorithm to data of contraceptive used tool counted 146 record with nine regular attributes, i.e., age, term of usage, menstrual cycle, just married, just give birth, breastfeeding, already having offspring, health problems, and having more than four children, whereas the attribute label/class that is contraceptive used.

3.1 Modeling Process

Data mining process to data set is processed by using algorithm C4.5 through application of RapidMiner Studio version 7.3 as in Fig. 1.

3.2 Knowledge Outcomes

Based on the modeling process is generated knowledge in the form of a pattern of decision tree which can be seen in Fig. 2.

Table 1 Number of patients in selection of contraceptives

No.	Contraceptive used	Number of patients
1	Spiral	47
2	Pill	59
3	Implant	26
4	Sterile	14

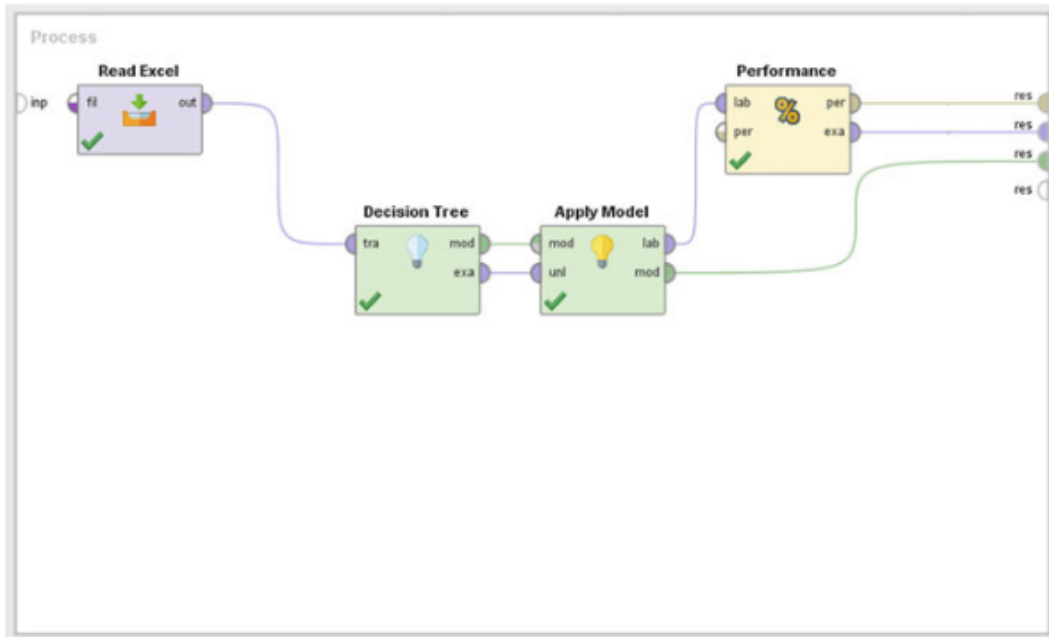


Fig. 1 Applying the model

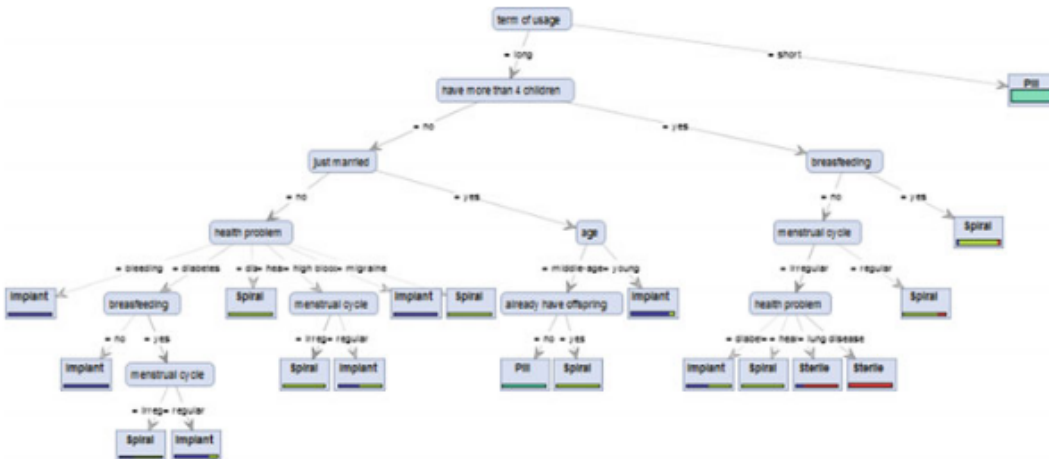


Fig. 2 Tree resulting from C4.5 algorithm

The decision tree in Fig. 2 can be spelled out in the form of rule as follows:

```

term of usage = long
| have more than 4 children = no
| | just married = no
| | | health problem = bleeding: Implant {Implant = 4, Pill = 0, Spiral = 0, Sterile = 0}
| | | health problem = diabetes
| | | breastfeeding = no: Implant {Implant = 2, Pill = 0, Spiral = 0, Sterile = 0}

```



```

| | | | breastfeeding = yes
| | | | menstrual cycle = irregular: Spiral {Implant = 1, Pill = 0, Spiral = 2,
Sterile = 0}
| | | | menstrual cycle = regular: Implant {Implant = 4, Pill = 0, Spiral = 1,
Sterile = 0}
| | | health problem = dizzy: Spiral {Implant = 0, Pill = 0, Spiral = 5, Sterile = 0}
| | | health problem = heart disease
| | | | menstrual cycle = irregular: Spiral {Implant = 0, Pill = 0, Spiral = 5,
Sterile = 0}
| | | | menstrual cycle = regular: Implant {Implant = 1, Pill = 0, Spiral = 1,
Sterile = 0}
| | | health problem = high blood pressure: Implant {Implant = 3, Pill = 0,
Spiral = 0, Sterile = 0}
| | | health problem = migraine: Spiral {Implant = 0, Pill = 0, Spiral = 2,
Sterile = 0}
| | just married = yes
| | | age = middle-aged
| | | | already have offspring = no: Pill {Implant = 0, Pill = 6, Spiral = 0,
Sterile = 0}
| | | | already have offspring = yes: Spiral {Implant = 0, Pill = 0, Spiral = 6,
Sterile = 0}
| | | age = young: Implant {Implant = 8, Pill = 0, Spiral = 1, Sterile = 0}
| have more than 4 children = yes
| | breastfeeding = no
| | | menstrual cycle = irregular
| | | | health problem = diabetes: Implant {Implant = 1, Pill = 0, Spiral = 1,
Sterile = 0}
| | | | health problem = dizzy: Spiral {Implant = 0, Pill = 0, Spiral = 2,
Sterile = 0}
| | | | health problem = heart disease: Sterile {Implant = 1, Pill = 0, Spiral = 0,
Sterile = 4}
| | | | health problem = lung disease: Sterile {Implant = 0, Pill = 0, Spiral = 0,
Sterile = 8}
| | | menstrual cycle = regular: Spiral {Implant = 0, Pill = 0, Spiral = 4,
Sterile = 1}
| | breastfeeding = yes: Spiral {Implant = 1, Pill = 0, Spiral = 16, Sterile = 1}
term of usage = short: Pill {Implant = 0, Pill = 53, Spiral = 1, Sterile = 0}

```

3.3 The Accuracy Level

The accuracy level obtained from the 146 train data records used in the Decision Tree C4.5 modeling can be seen in Fig. 3.

accuracy: 93.15%

	true Implant	true Pill	true Spiral	true Sterile	class precision
pred. Implant	23	0	4	0	85.19%
pred. Pill	0	59	1	0	98.33%
pred. Spiral	2	0	42	2	91.30%
pred. Sterile	1	0	0	12	92.31%
class recall	88.46%	100.00%	89.36%	85.71%	

Fig. 3 Performance vector

Based on Fig. 3, out of 146 record numbers there are 23 correctly predicted records of choosing Implant, 59 records are predicted to correctly select Pill, 42 records are predicted to correctly choose Spiral, and 12 records are correctly predicted to choose Sterile with accuracy = $(23 + 59 + 42 + 12)/146 = 93.15\%$. So that, the result gives accuracy rate obtained from data processing, i.e., 93.15% which can be classified excellent.

4 Conclusion

Based on the results of the study and modeling process using Decision Tree C4.5 algorithm conducted on the data of contraceptive use in Clinic Pratama Hasanah Pekanbaru can be drawn conclusion as follows:

1. Data selection of contraceptives can be used as training data to generate knowledge by using the method of classification.
2. The results of training data study of contraceptive usage tools can form knowledge in the form of decision tree models that transform data into decision trees that represent the rules.
3. Accuracy rate obtained from data processing reaches 93.15% which can be classified excellent so that this result can be a reference for midwife in advised patient in choosing contraception.

Acknowledgements This research was partially supported by Universitas Islam Riau.

References

- Kamaluddin (2012). Sistem Pendukung Keputusan Dalam Pemilihan Alternatif Alat Kontrasepsi Menggunakan Simple Additive Weighting. *Jurnal Jurusan Teknik Informatika UIN SGD Bandung*.

- Nugroho, C. G., Nugroho, D., & Fitriasih, S. H. (2014). Sistem Pendukung Keputusan Untuk Pemilihan Metode Kontrasepsi Pada Pasangan Usia Subur Dengan Algoritma K-Nearest Neighbour (KNN). *Jurnal Ilmiah SINUS*. ISSN:1693-1173.
- Trisnawarman D., Erlisa W. (2007). Sistem Penunjang Keputusan Pemilihan Metode/Alat Kontrasepsi. *Gematika. Jurnal Manajemen Informatika*, 9(1).

Accuracy of Algorithm C4.5 to Study Data Mining Against Selection of Contraception

ORIGINALITY REPORT

13%

SIMILARITY INDEX

11%

INTERNET SOURCES

12%

PUBLICATIONS

7%

STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

5%

★ repository.uir.ac.id

Internet Source

Exclude quotes On

Exclude bibliography On

Exclude matches < 1%