Usage Pattern Exploration of Effective Contraception Tool

Endang Wahyu Handamari

Department of Mathematics, University of Brawijaya, Malang, Indonesia

Abstract. Determination of contraception tool used by acceptors to support the Family Planning ("Keluarga Berencana") is a problematic. In choosing contraception tool, the acceptor must consider several factors, namely health factor, partner factor, and contraceptive method. Each contraception tool which is used has its advantages or disadvantages. Although it has been considering the advantages and disadvantages, it is still difficult to control fertility safely and effectively. So, an acceptor change a contraception tool used more than once. In order acceptors get the appropriate contraception tool then the patterns of changing in the usage of effective methods or contraception tool is determined. One of the methods that can be used to look the patterns of changing in the usage of contraception tool is *data mining* i.e. an extraction of large amounts of data. A pattern is said to be interesting if the pattern is not trivial, implicit, previously unknown, and useful. The patterns presented should be easy to understand, can be applied to data that will be predicted with a certain degree, useful, and new. The early stage before applying data mining is using *k nearest neighbors* algorithm to determine the factors shortest distance selecting the contraception tool. The next step is applying data mining to usage changing data of method or contraception tool of family planning acceptors which is expected to dig up information related to acceptor behavior pattern in using the method or contraception tool. Furthermore, from the formed pattern, it can be used in decision making regarding the usage of effective contraception tool. The results is the *k nearest neighbors* by using the Euclidean distance can be used to determine the similarity of attributes owned by the acceptors of Family Planning to the training data is already available. Based on available training data, it can be determined the usage pattern of contraception tool with the concept of data mining, where the acceptors of Family Planning are given a recommendation if the pattern is on the training data pattern. Conversely, if the pattern is none match, then the system does not provide recommendations of contraception tool which should be used.

Keywords: Contraception, Acceptors, Data Mining, k Nearest Neighbors

Abstrak. *Penentuan metode atau alat kontrasepsi oleh akseptor untuk mendukung Keluarga Berencana adalah suatu masalah. Dalam memilih alat kontrasepsi, akseptor harus mempertimbangkan faktor kesehatan, faktor pasangan, dan metode kontrasepsi. Setiap alat kontrasepsi yang digunakan memiliki kelebihan atau kekurangan, tetapi masih sulit untuk mengontrol kesuburan dengan aman dan efektif. Karenanya, akseptor mengubah alat kontrasepsi yang digunakan lebih dari satu kali. Agar akseptor mendapatkan alat kontrasepsi yang tepat maka pola perubahan dalam penggunaan alat kontrasepsi yang efektif ditentukan. Salah satu alat yang dapat digunakan untuk mencari pola perubahan dalam penggunaan alat kontrasepsi adalah penambangan data, yaitu ekstraksi pola yang menarik dari data berukuran besar. Suatu pola dikatakan menarik jika polanya berharga, tersirat, belum pernah diketahui, dan bermanfaat. Pola-pola yang disajikan harus mudah dipahami, dapat diterapkan pada data yang akan diprediksi dengan derajat tertentu, bermanfaat, dan baru. Tahap awal adalah menggunakan algoritma k-nearest neighbors untuk menentukan faktor jarak terpendek memilih alat kontrasepsi.*

*Corresponding author at: University of Brawijaya, Malang, Indonesia

E-mail address: ewahyu-math@ub.ac.id

Selanjutnya, melakukan penambangan data pada data perubahan penggunaan alat kontrasepsi akseptor KB yang diharapkan menggali informasi terkait pola perilaku akseptor dalam menggunakan alat kontrasepsi. Dari pola yang terbentuk, dapat digunakan dalam pengambilan keputusan mengenai penggunaan alat kontrasepsi yang efektif. Diperoleh hasil bahwa k-nearest neighbors dengan menggunakan jarak Euclidean dapat digunakan untuk mengetahui kesamaan atribut yang dimiliki oleh akseptor KB dengan data percobaan yang sudah tersedia. Dari data percobaan yang tersedia, dapat ditentukan pola penggunaan alat kontrasepsi dengan konsep data mining, di mana akseptor KB diberi rekomendasi jika pola tersebut ada pada pola data pelatihan. Sebaliknya, jika polanya tidak cocok, maka sistem tidak memberikan rekomendasi alat kontrasepsi yang sebaiknya digunakan.

1.   Introduction

Population problem in Indonesia which has a positive growth rate is a national problem in achieving development goals. These conditions will make problems in efforts to increase and the equitable distribution of public welfare. The growth rate is likely to increase so that if it is not controlled, it is predicted that in 2035 the population of Indonesia amounted to 305.6 million people. Therefore, it is necessary to do a cooperation and commitment of the government and society in an effort to reduce the rate of birth. The cooperation is realized in the form of development programs KKB (Kependudukan dan Keluarga Berencana). This KKB program was initiated since 1951 and has been experiencing growth. In 1970, the government formed Badan Koordinasi Keluarga Berencana Nasional (BKKBN). The BKKBN program includes thinning pregnancy by using contraceptive methods and creating economic prosperity for the society through the efforts of controlling and planning the population.

The contraception tool used by acceptors especially women should consider several factors. It concerns with the implications that may occur can not be known before usage. Therefore, it makes the acceptors of Family Planning possible to change contraception tool they use. An acceptor may change the contraception tool more than once until the acceptor feel comfortable and not problematic. Factors that are often being consideration in choosing a contraception tool are health factor, partner factor and factor of contraception methods. Medical history, among them are the menstrual cycle, degenerative diseases, blood pressure are included to health factor. Income level, frequency of intercourse, are some examples of partner factor. While the cost of contraception tool usage and security/risk is one example of contraception method factor. Research about classification of choosing potential contraception tool (has a big enough opportunity for success rate) using methods *Naive Bayes classification* has been done which after that implemented in the form of an application program [1]. With the same acceptors data, research on comparative methods Naive Bayes Classifier with *Backpropagation* Neural Network algorithm has been done [2].

In this research, the pattern of usage changing of effective contraception tool based on the primary data of Family Planning acceptors. The ascceptors which becomes the research object are the acceptors that change the contraception tool more than once. The method used in searching the pattern is *k-Nearest Neighbors (kNN)* algorithm to determine the shortest distance of several determining contraception tool factors and *Generalized Sequential Pattern* algorithm which is one of data mining algorithms.

2.   Materials and Methods

2.1.   Data Source

The data used in this study is the data taken from the acceptors of Family Planning in the sub-district of Jagir, Surabaya municipality.

2.2.   Data Analysis

In the data analysis, the required data are:

a)   Eleven factors (attributes) that affect acceptors in determining contraception tool. The first factor is the partner factor which contains: wife age, frequency of intercourse, acceptors educational level, acceptors job. The second facto is the health factors include: health status which includes weight, height, blood pressure (*systole* and *diastole*), the frequency of menstruation, and hereditary disease. The third factor is the factor of contraceptive method which is the cost incurred. While contraception tools that are used include condom, IUD (*Intra Uterine Device*), pills, Family Planning injection, implants and MOW.

b)   The last using of contraception tool.

c)   The pattern of usage changing of contraception tool.

2.3.   The Stages of Research

The stages used in this research includes:

a)   Identifying the problem and formulation of the problem.

b)   Collection the secondary data of family Planning acceptors in the Jagir sub-district, Surabaya municipality, which is the data related to eleven attributes and the patterns of usage changing of contraception tool.

c)   Data processing of questionnaires result from eleven attributes by finding the shortest distance of training data using *k Nearest Neighbor* method. *k-Nearest Neighbors (kNN)* algorithm is an algorithm that has the toughness in managing training document in large quantities and has a lot of *noise*. The process of *kNN* algorithm is to choose a training document of *k (k-values)* that is located closest to tested document. Training document category that are most appeared/found in some nearby *k*, which is later selected as a predicted category to the tested document[3].

d)   Determining for Family Planning acceptors pattern based on subsequence of *k* selected data through *sequence pattern mining*. *Sequence pattern mining* is used to find data that have a sequence, such as a sequence of usage changing of contraception tool. The process of *sequence pattern mining* occurs if given a number of *sequence*, where each *sequence* contains some elements and some elements consist of items may also be given a value of *minimum support*. Sequential pattern exploration is finding all *subsequence* whose frequency of occurrence is greater than the *minimum support* [4]. The completion of the sequential problems can be done by several methods such as the *Generalizes Sequential Pattern* [5].

3. Results and Discussion

3.1. Retrieval of Acceptor Data

The sampling for the research object is done on secondary data consists of data of Family Planning acceptors of Jagir sub-district, Surabaya municipality. Data obtained in the form of quantitative data and qualitative data. The data includes qualitative and quantitative data are given in Table 1 below.

Table 1 Qualitative data and quantitative data from the questionnaire results of family planning acceptors and contraception tool

No.	Qualitative Data	No.	Quantitative Data	No.	Contraception
1.	Degenerative disease	1.	Menstrual cycle	1.	Condom
2.	Highest education acceptor	2.	Age	2.	IUD
		3.	Weight	3.	Pills
3.	Highest education of husband's acceptor	4.	Height	4.	Family Planning Injection
		5.	Blood pressure (diastolic)		
4.	Acceptor's job	6.	Blood pressure (systolic)	5.	Implant
		7.	Frequency of intercourse	6.	MOW

The sourcecode for determining *k* closest distance of acceptors data with eleven factors of data using *k Nearest Neighbors (kNN)* as follows:

The function to calculate the euclidean distance between the two private data

```
double distance (double [] x, double [] y) {
        double d = 0.0;
        for (int i = 0; i <11; i ++) {
            d = d + (x [i] - y [i]) * (x [i] - y
[i]);
        }
        return Math.sqrt (d);
}
```

Figure 1 Syntax of the function to calculate

Followed by sorting the data index based on the shortest distance and the result is stored in an *array*. The final step is to use *data mining* techniques to look at the pattern of the test data to training data taken with the *k* shortest distances. If the test data pattern is in *k* training data pattern for the first time, then the system will provide final recommendations of contraception tool. If the pattern of test data is not in the *k* training data, then the system cannot give recommendations of contraception tool which should be selected. For example :

Given following test data with 11 attributes and the pattern of contraception tool usage is 35. Next, given 36 training data, by using Nearest Neighbor method, it is obtained Euclidean distance of test data to 36 training data as follows:

Table 2 Euclidean distance of test data to 36 training data

Number of the *i*-th Training Data	Distance between the Test Data to the Number of the *i*-th Training Data	Number of the *i*-th Training Data	Distance between the Test Data to the Number of the *i*-th Training Data
0.	0.910	18.	1.025
1.	1.456	19.	1.030
2.	1.230	20.	1.141
3.	1.141	21.	1.216
4.	0.972	22.	1.046
5.	1.133	23.	1.247
6.	1.043	24.	1.104
7.	1.011	25.	1.176
8.	1.080	26.	1.097
9.	1.077	27.	0.994
10.	1.042	28.	0.956
11.	0.848	29.	1.199
12.	1.049	30.	1.098
13.	0.871	31.	1.009
14.	1.211	32.	1.172
15.	0.829	33.	0.905
16.	0.995	34.	1.152
17.	1.112	35.	1.069

Table 3 10 nearest records after ordered from the smallest

No.	Record Order	Euclidean Distance to the Test Data	Consumption Pattern	True/False
1	15	0.829	563	False
2	11	0.848	645	False
3	13	0.871	635	True
4	33	0.905	345	True
5	0	0.910	365	True
6	28	0.956	345	True
7	4	0.972	415	False
8	27	0.994	516	False
9	16	0.995	361	False
10	31	1.009	365	True

Therefore, the test data that contains eleven attributes and contraception tool pattern 35 is given a recommendation to use contraception tool, namely Implant (see Table 1), because it has a short Euclidean distance and suitable pattern for the first time.

REFERENCES

[1]   S. Abusini, "Klasifikasi Pemakaian Alat Kontrasepsi yang Potensional dengan Metode Naïve Bayes Classifier", *Laporan Penelitian BOPTN DIKTI*, 2013.

[2]   E. Wahyu, K. Andawaningtyas and S. Abusini, "Comparison of Naïve Bayes Classifier and Back Propagation Method to Determine Suitable Contraception", *Proceeding of the First International Conference on Science (ICOS – 1)*, 2014.

[3]   R. Al-Shalabi, G. Kanaan and M. Gharaibeh, "Arabic Text Categorization using kNN Algorithm", *Proceedings of the 4th International Multi-conference on Computer Science and Information Technology*, Amman, Jordan, pp. 1-9, 2006.

[4]   R. Agrawal and R. Srikant, "Fast Algorithms for Mining Assocoation Rules", *Proceeding of the 20th International Conference on Very Large Databases*, Chile, 1994.

[5]   R. Srikant and R. Agrawal, "Mining Quantitative Association Rules in Large Relational Tables", *Association for Computing Machinery (ACM) SIGMOD Record*, vol. 25, no. 2, pp. 1-12, Jun. 1996.