http://infor.seaninstitute.org/index.php/infokum/index
JURNAL INFOKUM, Volume 8, No.2, JUNI 2020 ISSN 2302-9706

ANALYSIS OF SERVICE SATISFACTION LEVEL USING ROUGH SET ALGORITHM

Jon Hariono Sihotang
Mahasiswa Pascaarjana, STMIK Mikroskil. Jl M.H Thamrin No.140 Kel, Pusat Ps., Kec. Medan, Indonesia
Jon_hariono@gmail.com

Abstract
Data mining Is a technique that combines traditional data analysis techniques with algorithms for processing large amounts of data. Data mining can be used to perform data analysis and find important patterns in data. Data mining will be a benchmark or reference for making data mining processing decisions that can be done with the Rough Set method. Rough Set Method is one of the methods above that allows us to make decisions in hotel services because in this method there are formulations or stages of problem mechanics and a Result (decision) of a combination that may occur from the criteria above. From the results (decisions) derived from the processed data mining, it can be used as a reference for decision making. The Rought Set Method is a mathematical technique developed since 1980.
Keywords: data mining, roughest, service

1. Introduction

With the increasingly competitive conditions between companies, each company is racing to expand the market. The expectation of a direct market expansion is increased sales so that the company will have more consumers. However, there are several things that must be understood by the company as a producer, that the more consumers, the more difficult companies will recognize consumers carefully, especially about the likes or dislikes of the goods or services offered and the underlying reasons.

The facilities themselves are all things that are basic facilities and supporting facilities for the comfort and convenience that are deliberately provided by the hotel to be used, utilized, and enjoyed by guests during their stay at the hotel. In running its business, the hotel Suka Maju seeks to find out what are the needs and desires of consumers, including understanding consumer behavior and things that can provide a level of satisfaction to consumers. Competitors faced by the Hotel Suka Maju not only consist of companies that have the same facilities and services, but also from companies that have culinary facilities for tourism purposes. The number of customers is a very large influence on the survival of companies engaged in the sale of services because customer service companies are a source of income. The more company customers, the greater the revenue the company can achieve. Basically data mining is closely related to data analysis and software to look for patterns and similarities in data sets.

2. Literature Rivew

a. Data Mining

Data mining is the extraction of the most important and interesting information or patterns from existing data in a data warehouse. The existence of data mining is marked by the emergence of data problems at that time many companies have been collecting data for years, for example, the purchase data and sales data.[1]–[3]

In scientific journals data mining is also known as knowledge discovery in databases (KDD). Advances in technology in data storage media enables the collection and storage of large amounts of data can be done quickly and easily. Tools and data analysis techniques manually can no longer be used optimally in extracting information for large amounts of data. So we need a new technique that can answer these needs. Data Mining is one of the techniques that can be used because it has combined classical techniques with sophisticated algorithms such as Artificial Intelligence techniques to process data on a large scale. [4], [5]

Many terms are used to indicate the process of data mining. To be able to provide an understanding of Data Mining, here are some facts that occur as many organizations, both from the business and government world deal with a number of information and also the management of information databases and not including the need for a large scale data warehouse development. Often the data stored cannot be directly analyzed by standard statistical methods, this is due to some missing records or because of the qualitative rather than quantitative measures. The growth rate of database size is very fast, sometimes even the system administrator has problems to find out the relationship with the questions that arise, so there will be a distinct advantage if an organization has a way to explore the source of information in the form of a large database, so it can be known important information and also patterns that may be contained therein.

2.2 Teknik *Rough Set*

Teknik *Rough Set* merupakan sebuah teknik matematika yang di kembangkan oleh Pawlack pada tahun 1982 dan digunakan untuk analisis klasifikasi data dalam bentuk table (K. Thangavel, Qiang Shen, A. Pethalakshmi, 2006). Data yang digunakan biasanya data diskret. Tujuan dari analisis *Rough Set* adalah untuk mendapatkan perkiraan *rule* yang singkat dari suatu tabel. Hasil dari analis *rough set* dapat digunakan dalam proses *data mining* dan *knowledge discovery*. Teknik ini digunakan untuk menangani *masalah uncertainly, missing data, uncompleted, inconsistency data*, *imprecision dan vagueness* (tidak pasti, data hilang, tidak lengkap, tidak selaras, ketidaktepatan, ketidakjelasan).[6]–[8]

Teori ini memberikan pendekatan matematika baru untuk permasalahan dengan ketidakpastian yang tinggi. Teori ini menjadi dasar penting untuk kecerdasan buatan, pembelajaran mesin, perolehan informasi, analisis keputusan, *data mining*, sistem pakar, hingga pengenalan pola. Kelebihan teori ini adalah tidak diperlukannya *preliminary* dan juga informasi tambahan mengenai data dalam melakukan analisis suatu data. Tetapi teori *rough set* ini tidak dapat menyelesaikan permasalahan dengan atribut yang bernilai kontinu. Sedangkan yang ada dalam kasus di dunia nyata selalu mengandung variabel-variabel yang bernilai kontinu

Pendekatan *rough set* tampaknya menjadi dasar yang penting untuk AI dan ilmu kognitif, khususnya pada area *machine learning*, akuisisi pengetahuan, *decision analysis*, penemuan pengetahuan dari database, sistem pakar, penalaran induktif dan pengenalan pola. Teori *rough set* telah berhasil diterapkan dalam banyak masalah kehidupan nyata dalam kedokteran, farmakologi, teknik, perbankan, keuangan, analisis pasar, pengelolaan lingkungan dan lain-lain.[9]–[11]

Rough Set merupakan teknik yang efisien untuk *knowledge discovery in database* (KDD) proses dan *data mining*. Secara umum teori *rough set* telah digunakan dalam banyak aplikasi seperti *medicine, pharmacology, business, banking, engineering design, image processing* dan *decision analysis.*

Beberapa konsep dasar yang harus dilakukan untuk melakukan *knowledge discovery in database* (KDD) dengan teknik *rough set*, antara lain:

1. *Decision system*, representasikan data atau objek.
2. *Equivalence Class*, mengelompokkan objek-objek yang memiliki atribut kondisi yang sama.
3. *Discernibility Matrix / discernibility matrix modulo*, sekumpulan atribut yang berbeda antar objek.
4. *Reduction*, penyelesaian atribut minimal dari sekumpulan atribut kondisi dengan menggunakan *prime implicant* fungsi boolean.
5. *Generating Rules*, membangkitkan aturan-aturan *(rules*) dari pengetahuan yang didapat dalam proses ekstrak data.

Generating rules adalah suatu metode *rough set*untuk menghasilkan *rules/knowledge* berdasarkan *equivalence class* dan *reduct. Generating rules* dapat juga dikatakan sebagai suatu algoritma dari *data mining*, yang nantinya dari proses *generating rules*ini akan dihasilkan suatu *rules /knowledge* yang dapat di gunakan sebuah pengambilan keputusan.
Rough Set menawarkan dua bentuk repsentasi data yaitu *Information System* (IS) dan *Decision System*(DS). *Information System* adalah sebuah*Informating System (*IS) yang terdiri dari : IS ={U, A}, dimana U = {e1, e2,...en}dan A= {a1, a2,...,an} yang merupakan sekumpulan *example* dan *atribute* kondisi secara berurutan. Definisi di atas memperlihatkan bahwa sebuah *information system* terdiri dari sekumpulan *example*, seperti { e1, e2,...en }dan *attribute* kondisi, seperti { a1, a2, ...,an }.

2.2 Rough Set Technique

The Rough Set technique is a mathematical technique developed by Pawlak in 1982 and used for data classification analysis in tabular form. The data used is usually discrete data. The purpose of Rough Set analysis is to get a short rule estimate from a table. The results of the rough set analyst can be used in the process of data mining and knowledge discovery. This technique is used to deal with problems of uncertainly, missing data, uncompleted, inconsistency of data, imprecision, and vagueness (uncertainty, missing data, incomplete, incompatible, inaccuracy, obscurity).

This theory provides a new mathematical approach to problems with high uncertainty. This theory is an important basis for artificial intelligence, machine learning, information acquisition, decision analysis, data mining, expert systems, to pattern recognition. The strength of this theory is that there is no need for preliminary and additional information about data in analyzing data. But this rough set theory cannot solve problems with attributes that are of continuous value. While that exists in cases in the real world always contains variables that are continuous [12], [13]

The rough set approach seems to be an important basis for AI and cognitive science, especially in the area of machine learning, knowledge acquisition, decision analysis, knowledge discovery from databases, expert systems, inductive reasoning, and pattern recognition. The rough set theory has been successfully applied in many real-life problems in medicine, pharmacology, engineering, banking, finance, market analysis, environmental management, and others.
Rough Set is an efficient technique for knowledge discovery in database (KDD) processes and data mining. In general the rough set theory has been used in many applications such as medicine, pharmacology, business, banking, engineering design, image processing, and decision analysis. Some basic concepts that must be done to conduct knowledge discovery in database (KDD) with rough set techniques, including:

1. The decision system, represent data or objects.
2. Equivalence Class, grouping objects that have the same condition attribute.
3. Discernibility Matrix/discernibility matrix modulo, a set of different attributes between objects.
4. Reduction, the completion of a minimum attribute from a set of condition attributes using the prime implicant of the boolean function.
5. Generating Rules, generating rules (rules) from the knowledge gained in the data extraction process.

Generating rules is a rough set method for generating rules/knowledge based on equivalence classes and reducts. Generating rules can also be said as an algorithm of data mining, which later from the process of generating these rules will produce a rule/knowledge that can be used as a decision making. Rough Set offers two forms of data representation namely Information System (IS) and Decision System (DS). Information System is an Information System (IS) consisting of: IS = {U, A}, where U = {e1, e2, ... en} and A = {a1, a2, ..., an} which is a group example and attribute conditions in sequence.
The definition above shows that an information system consists of a set of examples, such as {e1, e2, ... en} and condition attributes, such as {a1, a2, ..., an}.

3. Research Methods

The steps involved in carrying out the activity are as follows:

1. Interview
   The interview is a direct question and answer to the relevant section to be able to provide information or information needed by the writer regarding the system that is running.
2. Observation (Observation)
   The method of observation (Observation) is one method of collecting data or facts that are quite effective. Observation is a direct observation that aims to obtain the information needed by making observations and records with direct observation to the agency or company.
3. Literature Study
   Literature is material or scientific sources that are commonly used to make a paper or other scientific activities. The literature used includes reference books, papers, journals, and internet documentation related to this research.
3. Analysis
   Efforts to systematically search and organize records based on the results of observations, and literature studies to enhance research understanding of the examined cases and present them as findings for others.

4. Results and Discussion

a. Data requirements

The data has been obtained will be simplified by using data transformation techniques based on the Interval Function Algorithm. According to the interval function algorithm, then for each variable the largest value is determined, the smallest value and range value, the number of classes and the interval value. From the results obtained, the data transformation process is carried out for complete data, as follows:

1. For Consumers:
   a. The greatest value (Xmax) = 112
   b. Smallest value (Xmin) = 8
   c. Value Range (Xrange) = 112 - 8 = 104
   d. Number of Classes (k) = 1 + 3.3 log (4) = 1 + (0.52) = 1.52
   e. Value Interval (Int) = 104 = 68.5 1.52

Data Transformation [Xmin + Int]:
The range between 30-62 is transformed into a number = 1
The range between 34 - 44 is transformed into a number = 2

2. For Age Groups:
   a. The greatest value (Xmax) = More than 40 years
   b. Smallest value (Xmin) = Less than 20 years
   c. Value Range (Xrange) = 40 - 20 = 20
   d. Number of Classes (k) = 1+ 3.3 log (4) = 1+ (0.52) = 1.52
   e. Value Interval (Int) = 20 = 13.2 1.52

Data Transformation [Xmin + Int]:
The range between 20-111 is transformed into a number = 1
The range between 112 - 187 is transformed into a number = 2
Based on the transformation data processing above, the transformation data obtained as table 1

Table 1 Data Using Data Transformation Techniques

No	Name	Jan	Feb	April	Mey	Result
1	A	1	2	2	1	Satisfied
2	B	1	1	3	1	Not satisfied
3	C	2	1	1	1	Not satisfied
4	D	3	2	1	2	Satisfied
5	E	2	2	1	3	Satisfied

a. Pembahasan

Berdasarkan data hasil transformasi yang telah didapatkan, maka data-data tersebut akan diolah sehingga suatu rules /knowledge yang dapat dipahami untuk pengambilan suatu keputusan. Dari data transformasi yang didapat pada tabel 4.3 maka dapat dilakukan proses pencarian knowledge seperti langkag-langkah ini :

1. *Discernibility Matrix*

In the Discernibility Matrix the condition variables consisting of Consumers, Amount, Age and Occupation, and decision variables consist of: Satisfied = 1 Dissatisfied = 2 Then each na, ma in the group in the form of equivalence class is simplified its name to EC1, EC2 , EC3, EC4 and EC5, so the results can be seen as in table 2

Table 2 Simplified Data Transformation

***Equivalence Class* (EC)**	**A**	**B**	**C**	**D**	**Decision (E)**
EC1	1	2	2	1	1
EC2	1	1	1	1	1
EC3	1	1	1	2	1
EC4	2	1	1	2	2
EC5	1	1	1	1	1

From the results of the transformation that has been simplified in table 3 then the data contained in each Equivalence Class is compared. In this comparison process, only the condition variables are considered, regardless of the decision variables. From this comparison process the Matrix Discernibility table is produced as in Table 3

Table 3 Matrix Discernibility

	EC1	**EC2**	**EC3**	**EC4**	**EC5**
EC1	**-**	**BC**	**D**	**D**	**ABC**
EC2	**BC**	**-**	**D**	**D**	**A**
EC3	**D**	**D**	**-**	**-**	**AD**

EC4	**D**	**D**	**-**	**-**	**AD**
EC5	**ABC**	**A**	**AD**	**AD**	**-**

2. *Discernibility Matriks Modulo D*

From the results of data processing with the Discernibility Matrix according to table 4, then the data is processed by means of the Modulo D Discernibility matrix, in this way the conditions and decision variables must be compared. So if the decision variables are also compared, the results will be like in table 4

Table 4. Modulo Matrix Discernibility D

	EC1	**EC2**	**EC3**	**EC4**	**EC5**
EC1	-	BC	D	-	-
EC2	BC	-	-	D	A
EC3	D	-	-	-	AD
EC4	-	D	-	-	-
EC5	-	A	AD	-	-

Note: If the value of the decision attribute is equal = empty decision attribute = Filled

3. Reduction

A reduct is a set of attributes that can produce the same classification as if all the attributes were used. While attributes that are not reducted are attributes that are not useful in the classification process. In this reduct process, the process of selecting the minimum variables from a set of condition variables is done using the Prime Implicant Boolean Function, in the following way:

EC1 = (BvC) ^ (D) ^ (BvC)
EC2 = (BvC) ^ (D) ^ (A)
EC3 = (D) ^ (A) v (D) ^ (D)
EC4 = (D) ^ (D)
EC5 = (D)

Table 5 Reduct

Class	CNF of Boolean Function	Implicant	Reduct
EC1	(BvC)^(BvC)	(D)	(D)
EC2	(BvC)^(D)^(A)	(BvD)^(CvD)^(A)	{B,C},
EC3	(D)^(A)^(D)^(D)	(DvA)	{D,A}
EC4	(D)	(D)	{D}

From the results of the reduct obtained, we get a rule / knowledge. Like the example above, the rules it gets are:

Rules:
D1➔E1
B1, C1➔E2, D1A2 ➔E2
D2, A1➔ E2
A2, D1➔E1
B1, C1➔E2, D1, A1➔ E2
D1➔E1EC1: IF D = 1 Then E = 1 IF Consumer = 30 Then Decision = Satisfied
EC2: IF B = 1, C = 1 Then E = 2, D = 1, A = 1 Then E = 2
IF Age = 65 & Consumer = 53 Then Decision = Not Satisfied
IF Age = 38 & Occupation = 40 Then Decision = Not Satisfied
EC3: IF D = 2, A = 1 Then E = 2 IF Work = 50 Then Decision = Not Satisfied
EC4: IF D = 2 Then E = 2 IF Work = 65 Then decision = Not satisfied
EC5: IF A = 2, D = 1 Then E = 2 IF Consumer = 120, = 50 Then Decision = Satisfied

4. Conclusions

The use of the Rough Set method is very helpful in estimating consumer needs and satisfaction that must be fulfilled in the coming period.

Reference

[1] S. Džeroski, "Data Mining," in *Encyclopedia of Ecology, Five-Volume Set*, 2008.
[2] R. Sowmya and K. R. Suneetha, "Data Mining with Big Data," in *Proceedings of 2017 11th International Conference on Intelligent Systems and Control, ISCO 2017*, 2017.
[3] C. C. Aggarwal and C. X. Zhai, *Mining text data*. 2013.
[4] R. Feldman and I. Dagan, "Knowledge Discovery in Textual Databases (KDT)," *Int. Conf. Knowl. Discov. Data Min.*, 1995.
[5] S. Spiegel, J. Gaebler, A. Lommatzsch, E. De Luca, and S. Albayrak, "Pattern Recognition and Classification for Multivariate Time Series Categories and Subject Descriptors," in *KDD*, 2011.
[6] A. Nakamura, S. Tsumoto, H. Tanaka, and S. Kobayashi, "Rough Set Theory and Its Applications," *J. Japanese Soc. Artif. Intell.*, 1996.
[7] Z. Pawlak, "Rough set approach to knowledge-based decision support," *Eur. J. Oper. Res.*, 1997.
[8] R. W. Swiniarski and A. Skowron, "Rough set methods in feature selection and recognition," *Pattern Recognit. Lett.*, 2003.
[9] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. 2016.
[10] D. Mining and T. Mining, "A Programmers Guide to Data Mining," *Text*, 2017.
[11] A. Peña-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," *Expert Systems with Applications*. 2014.
[12] K. Thangavel and A. Pethalakshmi, "Dimensionality reduction based on rough set theory: A review," *Applied Soft Computing Journal*. 2009.
[13] J. Liang, F. Wang, C. Dang, and Y. Qian, "A group incremental approach to feature selection applying rough set technique," *IEEE Trans. Knowl. Data Eng.*, 2014.