actors_and_its_dominant_criteria_affecting_cholesterol_level.pdf

by

Submission date: 17-Apr-2023 06:08AM (UTC+0700)
Submission ID: 2066307670
File name: actors_and_its_dominant_criteria_affecting_cholesterol_level.pdf (116.78K)
Word count: 2548
Character count: 12842

Decision Support Model in Determining Factors and Its Dominant Criteria Affecting Cholesterol Level Based on Rough-Regression

Riswan Efendi[1,2](✉) and Mustafa Mat Deris[1]

[1] Faculty of Computer Science, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Johor, Malaysia
{riswan, mmustafa}@uthm.edu.my
[2] Mathematics Department, Faculty of Science and Technology, State Islamic University of Sultan Syarif Kasim Riau, 28294 Panam, Pekanbaru, Indonesia

Abstract. The statistical regression models have been frequently used to explain the causal relationship between exogenous factors and the cholesterol level of patients. While, the dominant criteria for each exogenous factor which give impact to the cholesterol level are not yet investigated by previous studies. In this paper, we are interested to introduce a decision making model based on rough-regression in handling the significant contribution between the dominant criteria, exogenous and endogenous factors, respectively. The result showed the proposed model is able to investigate the dominant criteria and factors affecting cholesterol level patients. This model may help the counterparts in the decision making.

Keywords: Rough-regression · Decision making · Dominant criteria Cholesterol level

1 Introduction

Monitoring of cholesterol level is very essential activity for patients to continue their life. Besides that, this monitoring is also help the counterparts in providing information for decision making and health budget planning. Many models have been presented to investigate the factors (variables), such as, blood pressure [1], sleeping hour [2, 3], weight or obesity [4] and calorie level [5] which affect the patient cholesterol level. However, the dominant criteria (category) for each factor which give the significant impact to the cholesterol level is still issue and not easy to handle.

In the previous work, rough set model has been implemented to medical diseases, especially in prediction and management the diabetes mellitus [6]. In this paper, we introduce a decision making model to solve the issue and problem using rough-regression model (RRM). This model is very appropriate to handle the relationship among the qualitative variables with categorical data. The detail of theories, proposed model, and implementation are discussed in Sects. 2, 3 and 4, respectively.

R. Ghazali et al. (eds.), *Recent Advances on Soft Computing and Data Mining*,
Advances in Intelligent Systems and Computing 700,
https://doi.org/10.1007/978-3-319-72550-5_24

2 The Basic Theories

2.1 Rough Set Theory

The rough set theory has been proposed by Pawlak in 1982 [7], this theory has been well divided by researchers into information systems, indiscernibility relation, set approximations, rough clustering, and others. An information system $S = (U, \Omega, V_q, f_q)$ consists of [8–12]:

U: a nonempty, finite set called the universe;
Ω: a nonempty, finite set of attributes;
$\Omega = C \cup D$, in which C is a finite set of condition attributes and D is a finite set of decision attributes;
for each $q \in \Omega$, V_q is called the domain of q;
f_q: an information function $f_q : U \to V_q$.

Objects can be interpreted as cases, states, processes, patients and observations. Attributes can be assumed as features, variables, and characteristic information. A special case of information systems called decision table or attribute-value table is applied in the following analysis. In a decision table, the row and column correspond to objects and attributes, respectively. The starting point of rough set theory is the indiscernibility relation, generated by information about objects of interest. Let $S = (U, \Omega, V_q, f_q)$ be an information system, then any subset B of A determines a binary (equivalence) relation IND(B) on U, which will be called B-indiscernibility relation, and is defined as follows:

$$IND(B) = \{(x, y) \in U^2 : \forall a \in B, a(x) = a(y)\}, \tag{1}$$

where $a(x)$ denotes the value of attribute a for element x in U. The collection of all equivalence classes determined by IND(B), denoted by U/B. An equivalence class of U/B, containing x, is denoted by $[X]_B$. In rough set theory, an equivalence class is the basic concepts of our knowledge. The indiscernibility relation will be used next to define approximations, basic concepts of rough set theory.

Let $S = (U, \Omega, V_q, f_q)$ be an information system and let $B \subseteq A$ and $X \subseteq U$. We can approximate X using only the information contained in B by constructing the B-lower and B-upper approximations of X. Both approximations are denoted as:

$$\underline{B}(X) = \{\{x \in U | [x]_B \subseteq X\}\}, \tag{2}$$

and

$$\overline{B}(X) = \{x \in U | [x]_B \cap X \neq \emptyset\}, \tag{3}$$

where $[x]_B$ is an equivalence class containing x. While, the difference between both approximations and its accuracy can be written:

$$\mathrm{BND}(X) = \underline{B}(X) - \overline{B}(X), \tag{4}$$

$$\alpha(X) = \frac{\underline{B}(X)}{\overline{B}(X)}, \tag{5}$$

Moreover, the dependency attributes is formulated as [13]:

$$k = \frac{\sum_{x \in U/D} |\underline{C}(X)|}{|U|}; C, D \subseteq A \wedge C \cap D = \emptyset. \tag{6}$$

The maximum value of k can be interpreted as a dominant attribute or criteria.

2.2 Regression Model

A simple linear regression can be written as [14]:

$$Y = \beta_0 + \beta_1 X + e, \tag{7}$$

where Y is a dependent (endogenous) variable, X is an independent (exogenous) variable, β_0 and β_1 are intercept and slope, while e is error of model. The algorithm in building Eq. (7) can be explained by following steps:

Step 1: Check correlation between Y and X.
Step 2: Estimate intercept and slope, respectively using ordinary least square (OLS) method.
Step 3: Verify the significance X and Y using ANOVA and F-test.
Step 4: Verify the significance intercept and slope using t-test.
Step 5: Verify the normality error.

3 Proposed Decision Making Model

Let we have a multiple regression equation as follows [14]:

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \cdots + a_n X_n + e, \tag{8}$$

Based on Eq. (8), $X_1, \ldots, X_n$ are independent variables which affect to dependent variable, Y. While, $a_0, \ldots, a_n$ are estimated parameters (coefficients). In this equation, the main objective to determine the significant independent variables and its coefficients by following steps given in Sect. 2.2. We assume all independent variables affect to the dependent variable. In the previous studies, there is no approach can be implemented to investigate the dominant criteria (decisive condition) from each independent variable which give the maximum dependency to dependent variable. In this paper, we are interested to implement rough set approximations into regression model for investigating the dominant criteria from each independent variable by following steps:

Step 1: Transform real values of independent and dependent variables into categorical values (data) as presented in Table 1.

Table 1. Categorical data of exogenous and endogenous factors

ID	X_1	X_2	...	X_n	Y
P_1	Low	Low	...	Very rarely	Low
P_2	Low	Moderate	...	Very rarely	High
P_3	Very low	Low	...	Rarely	Medium
P_4	Many	Very low	...	Sometimes	High
...	...	...	...	...	...
P_n	Many many	Sometimes	...	Very rarely	High

Step 2: Based on Eqs. (2), (3), (5) and Table 1, determine the accuracy approximation as follows:

$$\text{Let } U/Y = U/\{\text{Low cholesterol, Medium cholesterol, High cholesterol}\}$$
$$= \left[\{R_{1,\ldots,}R_{n-3}\},\ldots,\{R_{3,\ldots,}R_{n-5}\}\right].$$

$$\underline{B}(X_1,\ldots,X_n) = \{R_2, R_7,\ldots,R_{n-3},R_{n-1}\},$$

$$\overline{B}(X_1,\ldots,X_n) = \{R_7, R_{n-7},\ldots,R_{n-5},R_{n-1}\},$$

Then, accuracy approximation can be written as:

$$\alpha(X_1,\ldots,X_n) = \frac{\underline{B}(X_1,\ldots,X_n)}{\overline{B}(X_1,\ldots,X_n)}. \tag{9}$$

Step 3: Based on Step 2, determine universe of decision as follows:

$$U/D = U/Y = \left\{\{R_{1,\ldots,}R_{n-3}\},\ldots,\{R_{1,\ldots,}R_{n-5}\}\right\}, \tag{10}$$

Step 4: Determine universe of each criteria (category) and variable (factors) as follows:
For X_1,

$$U/X_1(\text{Criteria} - 1) = \{\{R_1,\ldots,R_{n-3}\},\ldots,\{R_4,\ldots,R_{n-5}\}\},$$
$$\ldots$$
$$U/X_1(\text{Criteria} - p) = \{\{R_3,\ldots,R_{n-2}\},\ldots,\{R_7,\ldots,R_{n-4}\}\},$$

For X_2,

$$U/X_2(\text{Criteria} - 1) = \{\{R_4,\ldots,R_{n-1}\},\ldots,\{R_7,\ldots,R_{n-8}\}\},$$
$$\ldots$$
$$U/X_2(\text{Criteria} - p) = \{\{R_1,\ldots,R_{n-2}\},\ldots,\{R_4,\ldots,R_{n-4}\}\},$$
$$\ldots$$

For X_n,

$$U/X_n(\text{Criteria}-1) = \{\{R_1,\ldots,R_{n-3}\},\ldots,\{R_4,\ldots,R_{n-5}\}\},$$
$$\cdots$$
$$U/X_n(\text{Criteria}-p) = \{\{R_1,\ldots,R_{n-3}\},\ldots,\{R_4,\ldots,R_{n-5}\}\},$$

Step 5: Based on Eq. (6), determine dominant criteria for each variable or factor as follows:
For X_1,

$$\text{Criteria}-1 \to k_1^Y = \frac{U/X_1(\text{Criteria}-1)}{U/Y},$$
$$\cdots$$
$$\text{Criteria}-p \to k_p^Y = \frac{U/X_1(\text{Criteria}-p)}{U/Y}.$$

For X_2,

$$\text{Criteria}-1 \to k_1^Y = \frac{U/X_2(\text{Criteria}-1)}{U/Y},$$
$$\cdots$$
$$\text{Criteria}-p \to k_p^Y = \frac{U/X_2(\text{Criteria}-p)}{U/Y}.$$
$$\cdots$$

For X_n,

$$\text{Criteria}-1 \to k_1^Y = \frac{U/X_n(\text{Criteria}-1)}{U/Y},$$
$$\cdots$$
$$\text{Criteria}-p \to k_p^Y = \frac{U/X_n(\text{Criteria}-p)}{U/Y}.$$

Step 6: Based on Step 5, determine the max-value (dominant) for each category (criteria) and variable as follows:
For X_1,

$$\text{Dominant criteria}(X_1) = \max\left\{\left(k_1^Y\right),\ldots,k_p^Y\right\}, \tag{11}$$

For X_2,

$$\text{Dominant criteria}(X_2) = \max\left\{\left(k_1^Y\right),\ldots,k_p^Y\right\}, \tag{12}$$
$$\cdots$$

For X_n,

$$\text{Dominant criteria}(X_n) = \max\left\{\left(k_1^Y\right),\ldots,k_p^Y\right\}, \tag{13}$$

4 Implementation

Based on steps given in Sect. 2.2 and using row data from 20 patients, a linear relationship (regression model) between sleeping hour (X_1), weight (X_2), calorie level (X_3), blood pressure (X_4) and cholesterol level (Y) can be written mathematically:

$$Y = 71.791 + 3.012X_1 - 0.460X_2 + 0.002X_3 + 0.119X_4. \tag{14}$$

Based on Eq. (14), the coefficient (slope) of sleeping hour is highest among them, so that this variable give the biggest impact to cholesterol level if compared with weight, calorie level and blood pressure. On the other hand, we do not know which criteria (condition) from all exogenous factors or variables really affect to the patient cholesterol level. By using proposed model in Sect. 3, we can investigate the dominant condition or criteria from each exogenous variable by following steps:

Step 1: Transform numerical values of exogenous and endogenous variables into categorical values (data) as presented in Table 2.

Table 2. Transformation numerical into categorical values of variables

Patient ID	X_1	X_2	X_3	X_4	Y
P_1	Not normal	Small	L_2	Normal	High level
P_2	Not normal	Small	L_1	Normal	High level
P_3	Normal	Medium	L_3	Pre-H	High level
...	...	...	...	...	...
P_{20}	Normal	Big	L_3	Pre-H	High level

Step 2: Based on Table 2, determine universe of decision as follows:

$$\begin{aligned} U/Y &= \{\text{High level cholesterol}\}, \\ U/Y &= \{P_1, \ldots, P_{20}\}. \end{aligned} \tag{15}$$

Step 3: Determine universe of each criteria (category) and variable (factors) as follows:
For X_1,

$$\begin{aligned} &U/(\text{Not normal}) = \{P_1, P_2, P_8, P_{11}, P_{13}, P_{18}\}, \\ &U/(\text{Normal}) = \{P_3, P_5, P_6, P_9, P_{12}, P_{14}, P_{15}, P_{16}, P_{17}, P_{20}\}, \\ &U/(\text{Over sleep}) = \{P_4, P_7, P_{10}, P_{19}\}. \end{aligned}$$

For X_2,

$$U/(\text{Small}) = \{P_1, P_2, P_6, P_{11}\},$$
$$U/(\text{Medium}) = \{P_3, P_5, P_8, P_9, P_{12}, P_{13}, P_{18}\},$$
$$U/(\text{Big}) = \{P_4, P_7, P_{10}, P_{14}, P_{15}, P_{16}, P_{17}, P_{19}, P_{20}\}.$$

For X_3,

$$U/(\text{L1}) = \{P_2, P_8, P_{13}, P_{17}, P_{18}\},$$
$$U/(\text{L2}) = \{P_1, P_9, P_{11}, P_{12}\},$$
$$U/(\text{L3}) = \{P_3, P_{15}, P_{16}, P_{19}, P_{20}\},$$
$$U/(\text{L4}) = \{P_2, P_8, P_{13}, P_{17}, P_{18}\}.$$

For X_4,

$$U/(\text{Low}) = \{P_3, P_5, P_8, P_9, P_{12}, P_{13}, P_{18}\},$$
$$U/(\text{Normal}) = \{P_{11}, P_{13}\},$$
$$U/(\text{Pre}-\text{H}) = \{P_3, P_4, P_5, P_9, P_{10}, P_{18}, P_{19}, P_{20}\},$$
$$U/(\text{H}-1) = \{P_7\}.$$

Step 4: Based on Eq. (6), determine dominant criteria for each variable or factor as follows:
For X_1,

$$\text{Not normal} \rightarrow k_1^Y = \frac{6}{20}, \text{Normal} \rightarrow k_2^Y = \frac{10}{20}, \text{Over sleep} \rightarrow k_3^Y = \frac{4}{20}.$$

For X_2,

$$\text{Small} \rightarrow k_1^Y = \frac{4}{20}, \text{Medium} \rightarrow k_2^Y = \frac{7}{20}, \text{Big} \rightarrow k_3^Y = \frac{9}{20}.$$

For X_3,

$$\text{L1} \rightarrow k_1^Y = \frac{5}{20}, \text{L2} \rightarrow k_2^Y = \frac{5}{20}, \text{L3} \rightarrow k_3^Y = \frac{5}{20}, \text{L4} \rightarrow k_3^Y = \frac{5}{20}.$$

For X_4,

$$\text{Low} \rightarrow k_1^Y = \frac{2}{20}, \text{Normal} \rightarrow k_2^Y = \frac{9}{20}, \text{Pre}-\text{H} \rightarrow k_3^Y = \frac{8}{20}, \text{H1} \rightarrow k_3^Y = \frac{1}{20}.$$

Step 5: Based on Step 4, determine the max-value (dominant) for each category (criteria) and variable as follows:
For X_1,

$$\text{Dominant criteria of } X_1 = \max\left\{\frac{6}{20}, \frac{10}{20}, \frac{4}{20}\right\} = \frac{10}{20}. \quad (16)$$

For X_2,

$$\text{Dominant criteria of } X_1 = \max\left\{\frac{4}{20}, \frac{7}{20}, \frac{9}{20}\right\} = \frac{9}{20}.$$

For X_3,

$$\text{Dominant criteria of } X_3 = \max\left\{\frac{5}{20}, \frac{5}{20}, \frac{5}{20}, \frac{5}{20}\right\} = \text{all}. \quad (17)$$

For X_4,

$$\text{Dominant criteria of } X_4 = \max\left\{\frac{2}{20}, \frac{9}{20}, \frac{8}{20}, \frac{1}{20}\right\} = \frac{9}{20}. \quad (18)$$

Step 6: Based on Step 5, provide a decision making table for the dominant criteria (category) and variables affecting cholesterol level as presented in Table 3.

Table 3. Decision making for dominant criteria and factors

Variable/Impact	Criteria	Y (High level cholesterol)
X_1 (+)	Normal	10/20
X_2 (-)	Big	9/20
X_3 (+)	All criteria	5/20
X_4 (+)	Normal	9/20

Based on Table 3, the dominant criteria and factors which affect to cholesterol level (Y) are the sleeping hour (X_1) with positive impact and normal criteria (7–9 h) per day, the weight (X_2) with negative impact and big criteria (63–69 kg), the calorie level (X_3) with positive impact and all criteria (2235–3455 Cal) and blood pressure (X_4) with positive impact and normal criteria (90–110 mm/Hg). While, for other criteria are not dominant. In this case, most of condition cholesterol patients fall in the normal criteria, but we assume that the life style and the food consumption are also very significant in influencing their cholesterol level.

5 Conclusion

In this paper, we propose a procedure to investigate the dominant criteria (category) of each independent variable which influence the dependent variable. This proposed procedure has been examined to evaluate the dominant criteria and factors affecting student achievement. While this procedure is no yet discussed in the previous studies. Interestingly, the proposed procedure is also help the decision makers in determining the dependency variables to the dominant criteria precisely. Based on our perspective, rough-regression model (RRM) is suggested to handle the qualitative variables (data) in the social-economics research domains.

Acknowledgements. This study is supported by Research, Innovation, Commercialization, and Consultancy Management Office (ORICC) at Universiti Tun Hussein Onn Malaysia (UTHM) and in part by Contract Research Grant Vot. U689.

References

1. Sakurai, M., Stamler, J., Miura, K., Brown, I.J., Nakagawa, H., Elliot, P., Ueshima, H., Chan, Q., Tzoulaki, I., Dyer, A.R., Okayama, A., Zhao, L.: Relationship of dietary cholesterol to blood pressure: the intermap study. J. Hypertens. **29**, 222–228 (2011)
2. Wolk, R., Somers, V.K.: Sleep and the metabolic syndrome. Exp. Physiol. **92**, 67–78 (2007)
3. Kaneita, Y., Uchiyama, M., Yoshiike, N., Ohida, T.: Associations of usual sleep duration with serum lipid and lipoprotein levels. Sleep **31**, 645–652 (2008)
4. Miettinen, T.A.: Cholesterol production in obesity. Circulation **64**, 842–850 (1971)
5. Ueshima, H., Iida, M., Shimamoto, T., Konishi, M., Tanigaki, M., Doi, M., Nakashini, N., Takayama, Y., Ozawa, H., Komachi, Y.: Dietary intake and serum total cholesterol level: their relationship to different lifestyles in several Japanese populations. Circulation **66**, 519–526 (1982)
6. Ali, R., Hussain, J., Siddiqi, M.H., Hussain, M., Lee, S.: A hybrid rough set reasoning model for prediction and management of Diabetes Mellitus. Sensors **15**, 15921–15951 (2015)
7. Pawlak, Z.: Rough sets. Int. J. Compt. Inf. Sci. **11**, 341–356 (1982)
8. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publisher Dordrecht (1991)
9. Hampton, J.: Rough set theory: the basics (part 1). J. Compt. Intel. Finance **5**, 25–29 (1997)
10. Hampton, J.: Rough set theory: the basics (part 1). J. Compt. Intel. Finance **6**, 40–42 (1998)
11. Hampton, J.: Rough set theory: the basics (part 1). J. Compt. Intel. Finance **6**, 35–37 (1998)
12. Tay, F.E.H., Shen, L.: Economic and financial using rough sets model. Eur. J. Oper. Res. **141**, 641–659 (2002)
13. Herawan, T., Deris, M.M., Abawajy, H.: A rough set approach for selecting clustering attribute. Knowl. Based Syst. **23**, 220–231 (2010)
14. Wooldridge, M.: Introductory Econometrics a Modern Approach, 3rd edn. Thomson, South Western, USA (2006)

actors_and_its_dominant_criteria_affecting_cholesterol_level....

ORIGINALITY REPORT

20%	%	20%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Francis E.H. Tay, Lixiang Shen. "Economic and financial prediction using rough sets model", European Journal of Operational Research, 2002 Publication	5%
2	Riswan Efendi, Noor Azah Samsudin, Mustafa Mat Deris, Yip Guan Ting. "Flu Diagnosis System Using Jaccard Index and Rough Set Approaches", Journal of Physics: Conference Series, 2018 Publication	3%
3	"Recent Advances on Soft Computing and Data Mining", Springer Nature, 2017 Publication	3%
4	Riswan Efendi, Novi Yanti, Alex Wenda, Susnaningsih Mu'at, Noor Azah Samsudin, Mustafa Mat Deris. "Dominant Criteria and Its Factor Affecting Student Achievement Based on Rough-Regression Model", 2018 2nd International Conference on Informatics and Computational Sciences (ICICoS), 2018 Publication	3%
5	Rasyidah, Riswan Efendi, Nazri Mohd. Nawi, Mustafa Mat Deris, S. M. Aqil Burney. "Cleansing of Inconsistent Sample in Linear Regression Model based on Rough Sets Theory", Systems and Soft Computing, 2022 Publication	2%
6	Komorowski, J.. "Modelling prognostic power of cardiac tests using rough sets", Artificial	2%

Intelligence In Medicine, 199902
Publication

7	Rd. Rohmat Saedudin, Hairulnizam Mahdin, Shahreen Kasim, Edi Sutoyo, Iwan Tri Riyadi Yanto, Rohayanti Hassan. "Chapter 8 A Relative Tolerance Relation of Rough Set for Incomplete Information Systems", Springer Science and Business Media LLC, 2018 Publication	2%
8	Rasyidah, Riswan Efendi, Nazri Mohd. Nawi, Herdyan Maulana, Lisya Chairani. "Chapter 30 Customer's Behavior in Purchase Decision of Textile Materials: Rough-Regression Model", Springer Science and Business Media LLC, 2022 Publication	2%

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		