urchase_Decision_of_Textile_Materials_Rough-Regression_Model.pdf

by

Customer's Behavior in Purchase Decision of Textile Materials: Rough-Regression Model

Rasyidah[1,2], Riswan Efendi[3](✉), Nazri Mohd. Nawi[1], Herdyan Maulana[4], and Lisya Chairani[5]

[1] Faculty of Computer Science, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Johor, Malaysia
[2] Information Technology Department, Politeknik Negeri Padang, 25164 Padang, Indonesia
[3] Mathematics Department, Faculty of Science and Mathematics, Univerisiti Pendidikan Sultan Idris, 35900 Tanjung Malim, Perak, Malaysia
riswanefendi@fsmt.upsi.edu.my
[4] Faculty of Psychology, Universitas Negeri Jakarta, 12980 Jakarta Selatan, Indonesia
[5] Faculty of Psychology, Universitas Islam Negeri Sultan Syarif Kasim Riau, 28293 Pekanbaru, Indonesia

Abstract. Ordinary linear regression models have been widely implemented to measure the causal relationship between exogenous factors and purchase decision. While inconsistent samples are not yet investigated by previous studies through these models. In this paper, we are interested to discuss a purchase decision model by ordinary regression using data reduction strategy of rough sets in handling these samples type. The primary data and information were collected from 265 random customers of the textile stores in Padang City, Indonesia. The results showed regression model has better performance after data reduction in selecting purchase's factors. In this case, the customer's psychology is the most significant factor related to decision making in purchasing of textile material whether before and after data reduction. The proposed rough-regression model is very appropriate to support categorical data analysis with many uncertainty factors, especially in cross-sectional survey research types.

Keywords: Rough-regression · Purchase decision · Textile material · Customer's psychology · Data reduction

1 Introduction

Various Multiple Regression Models (MRM) presented to explain the causal relationship between independent (exogenous) variables (attributes) and dependent (endogenous) variable based on mathematical equations. Its applications can also be found in various problem domains, such as, education [1–3], economics-finance [4], customer behavior [5, 6]. The regression models can help the decision makers in providing information on interrelationship variables (attributes), the causal relationship equation and prediction planning for future data [5].

R. Ghazali et al. (Eds.): SCDM 2022, LNNS 457, pp. 302–310, 2022.
https://doi.org/10.1007/978-3-031-00828-3_30

Some previous studies [7–10, 18, 19], have been discussed the implementations of rough sets and regression models in various domains for categorical data analysis. The comparison studies between MRM and RST have been initiated by previous researchers. For example, the performance both models have been compared in property transaction for mass appraisal problems. RST produces comparable findings to MRM but outperforms MRM in terms of out-of-sample valuation accuracy. RST may be particularly effective in markets where econometric modeling is not possible due to a lack of quality and quantity of property market data sources. In 2009, the attribute selection in marketing for Indian cosmetic industries based on MRM and RST has been also compared where the rule generated through RST can an "expert" which may be referred to in future strategic decision making [7]. While MRM model needs many assumptions to obtain the good results. It can be coped the similarity components between MRM and RSM on decision making purposes. However, both studies do not discuss more how to integrate RST into MRM in providing the better decisions.

In early 2018, some studies introduced RST into MRM through lower-upper approximations in determining dominant criteria for each significant factor which contributed to dependent factor in the regression equation. Moreover, the data reduction strategy of RST was adopted to eliminate the inconsistent samples in improving the performance of RSM, such as, adjusted R-square, intercept-slopes and mean square error of training and testing data [8]. Further study, the dominant criteria and data elimination strategy were also applied to student achievement data set and its factors [9]. While RST and MRM also used to investigate the significant attribute which influence the purchase decision based on customer's preference [10]. Additionally, Rough-Regression Model (RRM) was also implemented to investigate the significant factors of financial wellbeing. However, these RRM were integrated using the real case studies data without supporting by simulation results. Therefore, the continue study should be considered to scope the capability of RRM in categorical data analysis. Motivated by previous works, we are interested to further study of RRM through simulation and implementation in a case study of purchase decision based on customer's preferences. The main target of this simulation is to evaluate the effect of data reduction into three different aspects such as, adjusted R-square, slope-intercepts and mean square error of regression model.

2 The Basic Theories and Methodology

2.1 Rough Set Theory (RST)

It was first proposed by mathematician Zdzislaw Pawlak in the early 1980s as a mathematical tool for dealing with the ambiguous and imprecise. Rough Set Theory (RST) is related to Fuzzy Set Theory (FST), except the uncertainty and imprecision are indicated by a set's boundary region, rather than partial membership, as in FST. Interior and closure topological procedures known approximations can be used to define the rough set idea in a very generic way [11, 12]. Comparisons of the definitions of classical sets, fuzzy sets, and rough sets are fascinating. A classic set is a fundamental concept that can be defined intuitively or axiomatically. The fuzzy membership function, which incorporates sophisticated mathematical structures, numbers, and functions, is used to define a fuzzy set. Rough set is defined through topological operations known as approximations, which necessitates the use of advanced mathematical ideas.

An information system, often known as an information table (Table 1), is a table that contains objects (rows) and characteristics (columns). It's used in the form of data that Rough Set will use, where each object has a specific number of attributes. These objects are specified in line with the data table's format, with rows serving as analysis objects and columns serving as attributes.

Table 1. Example of object's information and attributes

Patient code	Conditional attribute			Decision attribute
	Headache	Vomiting	Temperature	Viral illness
P1	No	Yes	High	Yes
P2	Yes	No	Very high	Yes
P3	Yes	Yes	Very high	Yes
P4	Normal	Yes	Normal	Normal
P5	Yes	Normal	High	Normal
P6	Normal	Yes	Very high	Yes

The IR (indiscernibility relation) is a key notion in RST. It's a relationship between two or more items in which all of the values are the same in relation to a subset of the characteristics being analyzed. IR is an equivalence relation that considers all identical things in a collection to be elementary [12]. Based on Table 1, The set appears to be made up of conditional qualities that are closely related to the patients' symptoms, such as headaches, vomiting, and temperature. When Table 1 is split down, the set of items or patients P2, P3, and P5 is indistinguishable in terms of headache attribute. In terms of the vomiting attribute, the set involving P1, P3, and P4 is indistinguishable. P2 has a viral infection, whereas P5 does not; yet, conditional features such as headache, vomiting, and temperature make them indistinguishable. As a result, P2 and P5 are components of a patient's set of unresolved symptoms [20].

The indiscernibility relation is designed to express the fact that it is unable to discern some things using the available information due to a lack of knowledge. Approximations are another essential term in rough sets theory, as they are linked to the meaning of the topological operations of the approximation. The interior and closure operations in a topology generated by the indiscernibility relation are the lower and upper approximations of a set. Below is presented and described the types of approximations that are used in RST.

The rough sets algorithm can be derived by following steps given:

Step 1: Provide the information table or information system include the relation between conditional and decision attributes based on data sets given.
Step 2: Determine the relation between two objects or more (indiscernibility relation).
Step 3: Determine lower-upper approximations and boundary region.

Step 4: Remove the inconsistent information from information table.
Step 5: Generate the decision rules.

2.2 Multiple Regression Model (MRM)

Regression models are commonly employed to answer scientific questions about the relationships between a set of variables. Linear regression models, in particular, demonstrate how one or more explanatory factors can explain a portion of the natural individual-to-individual variance in a continuous response variable. The mean values of a response variable (denoted by y) vary as a linear function of a set of explanatory variables and x as independent variables in linear regression. The foundation for multiple linear regression is the same as for simple linear regression, and the four fundamental assumptions given for simple linear regression must also be valid for multiple linear regression. The following is a mathematical representation of a multiple linear regression model:

$$y = b_0 + b_1x_1 + \ldots + b_nx_n + e. \tag{1}$$

From Eq. (1), y is a response or dependent variable, $(x_1, x_2, \ldots, x_n)$ are explanatory or independent variables, $b_0, \ldots, b_n$ are coefficient model, while n is a number of observation. In this equation, the linear relationship between dependent and independent variables can be investigated statistically by using assumptions. Additionally, the knowledge delivery from features until prediction on regression model can be illustrated in Fig. 1.

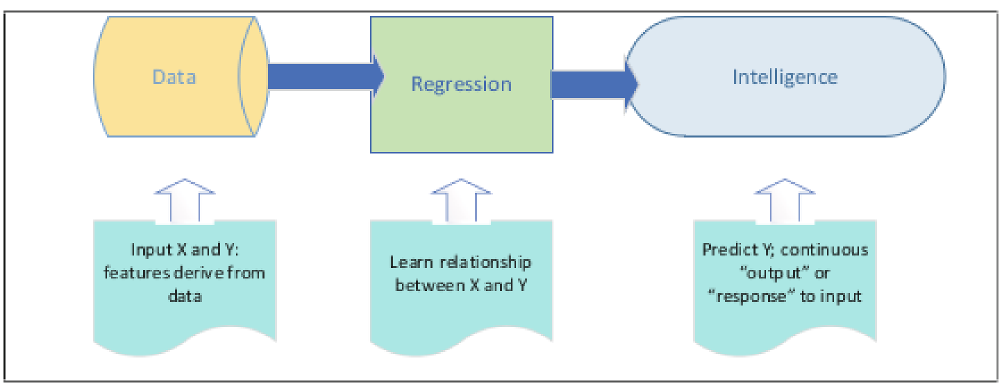

Fig. 1. Regression and its knowledge delivery process

Based on both sections (Sect. 2.2 and 2.3), it can be highlighted that RST and MRM have some similarities in investigating relationship between dependent (decision attribute) and independent variables (conditional attributes). Moreover, this relationship will be explored and applied for decision making purposes such as planning, prediction, and classification. Thus, both approaches have a potential to be combined, especially in removing of inconsistent elements in the data sets. Therefore, the performance of data analysis could be improved.

3 Implementation on Purchase Decision

In this section, some independent factors are selected and considered which contribute to purchase decision of textile materials such as demography of respondent, culture, social, family, personality, and psychology of respondent [5]. Integrated RST-MRM is implemented to determine the significant factors according to the following steps as given in Sect. 3:

Step 1: Select the independent factors using MRM (without data reduction). For our case study, the output is presented in Table 2.

Table 2. MRM for purchase decision of textile material (before reduction)

Independent factor	Coefficient	t-calculated	Sig
Intercept	0.890	3.382	0.001
Culture (x_1)	0.118	2.228	0.027
Social (x_2)	0.102	−1.632	0.104
Personality (x_3)	0154	1.935	0.055
Psychology (x_4)	0.622	8.213	0.000
Income (x_5)	0.025	0.696	0.487

Table 2 showed respondent's culture and psychology are significant factors which contribute to their purchase decision if compared with social, personality and income. While the contribution of these factors could be further investigated by data reduction strategy of RST. This strategy is used to remove the inconsistent samples and factors in the data set, respectively.

Step 2: Transform all values into categorical data types. For our case study, the transformation data is presented in Table 3.

Table 3. Transformation numerical values into categorical type

Cust. code	x_1; (A_1)	x_2; (A_2)	$x_{3;}(A_3)$	x_4; (A_4)	$x_{5;}(A_5)$	y; (D)
C1	agree	agree	agree	Agree	Very low	agree
C2	agree	agree	agree	Doubt	low	doubt
…	…	…	…	…	…	…
C186	Very agree	doubt	agree	Agree	low	agree

Step 3: Provide indiscernibility relation for conditional attributes (A_i) and purchase decision (D) as presented in Table 4.

Table 4. Indiscernibility relation among conditional attributes and purchase decision

Cust. code	x_1; (A_1)	x_2; (A_2)	x_3; (A_3)	x_4; (A_4)	x_5; (A_5)	y; (D)
C1	agree	agree	agree	Agree	very low	agree
C6	agree	agree	agree	Agree	very low	agree
C22	agree	agree	agree	Agree	very low	agree
…	…	…	…	…	…	…
C186	Very agree	doubt	agree	Agree	low	agree

Step 4: Based on Table 4, eliminate the inconsistent customers/respondent using lower-upper approximation approaches of RST.

In this case study, 87 inconsistent respondents have been eliminated from the data set. While 99 data will be used to rebuild the regression model and generated rules for decision making. In general, these respondents have considered the same perception or opinion for each conditional attribute, but difference in choosing criteria of purchase decision. Therefore, they are very vague in the information system and should be removed from the data set. Most of perception are very similar from respondents C25, C56, C120 and C121 for each attribute. However, they are different on considering the purchase decision, namely "Doubt and Disagree". One of the most difficult aspects of decision analysis' explanation and recommendation duties is inconsistency of data. We distinguish two types of information inconsistencies: the first is caused by indiscernibility of objects described by attributes defined on nominal or ordinal scales, and the second is caused by violations of the dominance principle among attributes defined on preference ordered ordinal or cardinal scales, i.e. criteria. In this study, we look about how a new RST-based technique handles these two types of discrepancies. When this theory is combined with inductive learning techniques, the ordinary regression model is improved. The influence of numerical attribute scales and preference-ordered criteria scales on the syntax of induced decision rules is studied in depth.
Step 5: Rebuild MRM after data elimination ($\widehat{PD}_{\mathrm{AE}}$) for purchase decision. Mathematically, it can be written as below:

$$\widehat{PD}_{\mathrm{AE}} = 0.556 + 0.064\,Cult - 0.08\,Soc + 0.358\,Pers + 0.641\,Psy - 0.016\,Inc. \quad (2)$$

Step 6: Compare the performance MRM before and after data reduction by considering some aspects as presented in Table 5.

Based on Table 5, all aspects are improved significantly, especially the percentage of adjusted R-square value is increased sharply from 48% to 62%. This value is also supported by simulation result in Sect. 3.1. In this case, we state that the inconsistent samples (respondents) may directly influence the performance of the conventional regression model as well. It can be proven by other aspects such as intercept-slopes model as justified in Sect. 3.2. Some slopes values are up and down after data reduction, because we consider the consistent samples only. In this table, the accuracy of testing data after data

Table 5. Comparison MRM before-after data reduction for textile material data

Aspect	MRM (before elimination)	MRM (after elimination)	Remark
Intercept	+0.89	+0.56	Down
Slope x_1	+0.12	+0.06	Down
Slope x_2	−0.10	−0.08	Up
Slope x_3	+0.15	+0.36	Up
Slope x_4	+0.62	+0.64	Up
Slope x_5	−0.03	−0.16	Down
Significant factors	Culture and Psychology of customer	Psychology of customer	
R-square	0.48	0.62	Up

reduction is higher than before (Sect. 3.3). Besides, the number of significant factors is also changed after the data reduction process.

In this case, a very significant factor is psychology (x_4) of the respondent. Since the majority respondent is an early adult (17–23 years old), who do not have their salary or income. But they are very independent in determining their decision in purchasing materials. Besides, they have been born as Gen Y (millennial generation). Buying decision making is a result of numerous psychological factors. These factors consist of cognitive and emotional aspects [13]. The interplay among these aspects is associated with the personality, trust, and motivational states that people engage when they are selecting, purchasing, using and disposing of products or services [14, 15]. Purchasing and using products are more than just economic-objective activities but rather include a wide spectrum of mental processes that involved people's psychological capacity in identifying needs, generating ideas, and deciding post-purchase behaviour. These mental processes help people to satisfy what they think is important. Consumer's purchase behaviour is reflecting their personality, motivation, and level of trust toward the product/service [16]. Personality refers to a set of behaviour characteristic reflecting people's way of thinking (cognitive) and feeling (emotion). These characteristics are evolved with biological and environmental factors [17]. While motivation is a psychological aversion, reflects in a desire and reasons for people to act or behave in a particular way.

It may come from external and internal sources. Internal motivation is strongly associated with how people feel (emotion) about something. Every individual is having his own cognitive and emotional preferences which may be associated with the way he purchases and uses the products [15]. At the end, both factors developed into a stable cognitive schema that people used when they choose, buy, and use the product. This mental schema is defined as a trust, which refers to one's faith and hope in something [17]. In a buying behaviour, trust is a person's consistent pattern of responses toward a product/service. Trust is a part of the cognitive process that developed based on one's past experience about the products. All of these psychological components are playing important roles in individual purchase behaviour. These components regulate the way people consider something as important, useful, and valuable. It also helps them to dispose appropriate emotional response toward a product/service. Therefore, the data reduction strategy is not only capable to eliminate the inconsistent samples but also applicable to remove the insignificant attributes or factors in the data sets [18].

4 Conclusion

In this case study, the psychology is the most significant factor which influence purchase decision based on customer's preferences. This factor always appears before and after data reduction considered. While other independent factors such as, culture, social, personality and income also affect to customer's decision, but they could not be considered statistically. In this case, these attributes are not easy to be measured accurately, because perception or preference items by people. Sometimes to recollect data is also big issue because related with budget, human resources, and other instruments. Therefore, the data analysis or modelling can be resolved by data reduction strategy. Our contribution is addressed to help the companies or industries in customized services and targeted marketing strategies based on customer's preferences. Besides, it also to help the survey research with many uncertainty variables or factors. They do not need to recollect data if obtained unsatisfied results. By proposed rough-regression model, this problem can be solved, especially improvement of R-square value and number of significant independent variables.

Acknowledgments. This research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) through Tier 1 vot. H938.

References

1. Yaxian, W.: Analysis on the influence of education level on the strategy of rural revitalization based on the stepwise regression analysis method. In: 2nd (ICEKIM) International Conference on Education, Knowledge and Information Management, pp. 859–862 (2021)
2. Ranjan, R., Ranjan, J.: Predicting parameters importance in higher education using data analysis : a regression model. In: 2nd International Conference on Computational Intelligence & Communication Technology, pp. 605–608 (2016)
3. Lopez-Martin, C., Ulloa-Cazarez, R.L., Garcia-Floriano, A.: Support vector regression for predicting the productivity of higher education graduate students from individually developed software projects. Int. J. Inst. Eng. Technol. **11**, 265–270 (2017)
4. Lei, L., Yao, G.: Regression analysis of education and social conflict willingness of migrant workers under the social and economic form, pp. 661–664 (2020)
5. Therasa, C., Banu, C.V., Gayathri, R., Gopalakrishnan, S., Manikandan, S.: Linear regression model fit for internal marketing and customer relationship management, pp. 927–932 (2017)
6. Qi, W., Luo, X., Liu, X.: Discrete choice model of customer behavior and empirical study, pp. 5677–5682 (2016)
7. Kurniawan, H., Agustin F., Yusrizal, Ummi, K.: Implementation data mining in prediction of sales chips with rough set method. In: 6th (CITSM) International Conference on Cyber and IT Service Management (2018)
8. Efendi, R., Samsudin, N.A., Deris, M.M.: Medipre: Medical diagnosis prediction using rough-regression approximation. In: Proceeding HP3C, pp. 35–39 (2018)

9. Efendi, R., Yanti, N., Wenda, A., Mu'at, S., Samsudin, N.A., Deris, M.M.: Dominant criteria and its factor effecting student achievement based on rough-regression model. In: Proceeding 2nd ICICoS, pp. 29–32 (2018)
10. Rasyidah, Nawi, N.M., Efendi, R.: Rough-regression model for investigating product attributes and purchase decision. In: Proceeding 7th ICCCE-IEEE Xplore, Kuala Lumpur, Malaysia, pp. 264–267 (2018)
11. Pawlak, Z.: Rough sets. Int. J. Compt. Inf. Science. **11**, 341–356 (1982)
12. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publisher, Dordrecht (1991)
13. Li, W., Chen, W., Ma, C.: Application of attribute reduction algorithm based on rough set in financial management of commercial banks. In: 13th ICNC-FSKD, pp. 2383–2390 (2017)
14. Orji, M.G., Sabo, B., Abubakar, M.Y., Usman, A.D.: Impact of personality factors on consumer buying behaviour towards textile materials in south eastern Nigeria. Int. J. Bus. Econ. Res. **6**(1), 7–18 (2017)
15. Chen, H.-T., Lin, Y.-T.: A study of the relationships among sensory experience, emotion, and buying behavior in coffeehouse chains. Serv. Bus. **12**(3), 551–573 (2017). https://doi.org/10.1007/s11628-017-0354-5
16. Mathai, S.T., Shanthaamani, V.: Materialism: its impact on impulse buying behaviour among the retail customers of Cochin. Int. J. Manag. Res. Rev. **6**(5), 647 (2016)
17. Fahmi, M., Panjaitan, N.A., Habibie, I., Siregar, A.Y., Amarullah, G., Sunjaya, D.K.: Does your neighborhood protect you from being depressed? A study on social trust and depression in Indonesia. BMC Public Health **19**(1), 1371 (2019)
18. Sahid, D.S.S., Efendi, R., Putra, E.H., Wahyudi, M.: Categorizing attributes in identifying learning style using rough set theory. Int. J. Adv. Comput. Sci. Appl. **11**(1), 292–298 (2020)
19. Efendi, R., Sahid, D.S.S., Putra, E.H., Deris, M.M., Annisa, N.G., Sari, I.M.: Healthy diet food decision using rough-chi-squared goodness. In: Jizat, J.A.M., et al. (eds.) Advances in Robotic, Automation and Data Analytics. Advances in Intelligent Systems and Computing, vol. 1350, pp. 296–308. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-70917-4_28
20. Rissino, S., Torres, G.L.: Rough set theory-fundamental concepts, principals, data extraction, and applications. In: Ponce, J., Karahoca, A. (eds.) Data Mining and Knowledge Discovery in Real Life Application Information, pp. 35–58 (2009)

urchase_Decision_of_Textile_Materials_Rough-Regression_Model.pdf

ORIGINALITY REPORT

8%	%	8%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Silvia Rissino, Germano Lambert-Torres. "Chapter 3 Rough Set Theory — Fundamental Concepts, Principals, Data Extraction, and Applications", IntechOpen, 2009 Publication	4%
2	W Gata, H B Novitasari, R Nurfalah, R Hernawati, M J Shidiq. "Analysis of Regression Algorithm to Predict Administration, Production, and Delivery to Accuracy of Delivery of Products in Cosmetic Industry", IOP Conference Series: Materials Science and Engineering, 2019 Publication	2%
3	Riswan Efendi, Susnaningsih Mu'at, Voni Apriana Dewi, Nelsi Arisandy, Noor Azah Samsudin, Dadang Syarif Sihabudin Sahid. "Rough-Regression for Categorical Data Prediction based on Case Study", 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), 2019 Publication	2%

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		