nts_in_investigating_of_financial_wellbeing_attributes_using.pdf

by

Submission date: 17-Apr-2023 06:08AM (UTC+0700)
Submission ID: 2066307705
File name: nts_in_investigating_of_financial_wellbeing_attributes_using.pdf (784.91K)
Word count: 2650
Character count: 14613

Removing Unclassified Elements in Investigating of Financial Wellbeing Attributes Using Rough-Regression Model

Riswan Efendi[1]
[1]Mathematics Department, Universitas Islam Negeri Sultan Syarif Kasim Riau
28293 Pekanbaru, Indonesia
riswan.efendi@uin-suska.ac.id

Susnaningsih Mu'at[2], Nelsy Arisandi[2]
[2]Faculty of Economics and Social Science, Universitas Islam Negeri Sultan Syarif Kasim Riau
28293 Pekanbaru, Indonesia
susnaningsih@uin-suska.ac.id; nelsi_arisandy@yahoo.co.id

Noor Azah Samsudin[3]
[3]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia
86400 Batu Pahat, Malaysia
azah@uthm.edu.my

ABSTRACT
In economics research survey, the causal relationship between independent and dependent attributes has been frequently investigated by using regression linear models. However, not easy to achieve the high R-square value between both attributes if there are too many unclassified elements in data sets. This paper presents removing unclassified elements in conventional regression model using rough sets approximation. The proposed model is address to handle the unclassified academic staffs in data set which less contribution for supporting financial wellbeing decision. The result showed that number of unclassified staff has a positive effect to increase coefficient determination (R-square) value in the regression model. In this case study, the financial wellbeing of academic staff is significantly influenced by two different attributes, namely, financial behavior and financial stress. It also may help decision makers or universities management in improving their staff in financial wellness and wellbeing.

CCS Concepts
• **Applied computing→ Multi-criterion optimization and decision-making**

• **Applied computing→Forecasting**

Keywords
Rough-regression; unclassified element; financial wellbeing; financial stress; financial behavior

1. INTRODUCTION
In economic-finance applications, the rough sets theory as a new knowledge discovery tool provide many advantages, which inspired many scholars to do research in adapting this theory to the business domain [1]. So far they have achieved many promising results. The rough sets model has the following advantages [2][3]. For example, it is based on the original data only and does not need any external information, unlike probability in statistics or grade of membership in the fuzzy set theory [4]. The rough set theory is a appropriate tool for analyzing not only numerical attributes but also non-numerical (categorical) ones. It bridges important facts uncertainty in data and expresses them in the natural language of decision rules. The set of decision rules derived by the rough set gives a generalized description of the knowledge contained in the financial information tables, eliminating any redundancy typical of the original data. The decision rules obtained are based on data reduction and unclassified element elimination. In the above review of rough sets model for economic and financial prediction, the rough sets model is a promising alternative method to conventional methods [5].

Recently, rough-regression model has been implemented to medical and also production data [6][7][8][9]. However, we found some rooms can be improved. In this paper, we introduce the importance of removing unclassified elements using rough sets approximation for better regression model performance. Proposed model will be implemented in investigating of financial wellbeing of university staff at Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia. This model is very appropriate to increase the R-square of linear regression model by removing unclassified elements in data set. The detail of theories, proposed model, and implementation are discussed in Sections 2, 3 and 4, respectively.

2. ROUGH SETS AND REGRESSION MODELS

2.1 Rough Set and Applications
The rough set theory was initiated by Pawlak in 1982 [1], there are some components in this theory, such as, information systems, indiscernibility relation, set approximations, rough clustering, and others. An information system $S = (U, \Omega, V_q, f_q)$ consists of a nonempty finite set called the universe, finite set attributes, and others [1][2].

A cases, states, processes, patients and observations are interpreted as objects in rough sets. While, the features, variables, and characteristic information are denoted as attributes. A decision table is used for a special case of information systems or attribute-value table is applied in the following analysis. In a decision table, the row and column correspond to objects and attributes, respectively [1][2]. Based on [2], the indiscernibility,

ICSCA '19, February 19–21, 2019, Penang, Malaysia

ACM ISBN 978-1-4503-6573-4/19/02…$15.00
https://doi.org/10.1145/3316615.3316651

lower-upper approximations, and boundary region can be formulated by following Equations (1) – (5) accordingly.

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

where $a(x)$ denotes the value of attribute a for element x in U. The collection of all equivalence classes determined by $\mathrm{IND}(B)$, 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:

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

and

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

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

$$\mathrm{BND}(X) = \underline{B}(X) - \bar{B}(X), \quad (4)$$

$$\alpha(X) = \frac{\underline{B}(X)}{\bar{B}(X)}, \quad (5)$$

Moreover, the dependency attributes is formulated as [10]:

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

The maximum value of k can be interpreted as a dominant attribute or criteria. Majority, rough sets theories can be implemented to investigate the dependency between conditional attributes and decision attribute for various domains, such as, cognitive sciences and medical data.

2.2 Regression Models and Its Applications

A simple linear regression can be written as [11]:

$$Y = \beta_0 + \beta_1 X + e, \quad (7)$$

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 [6][7]:

Step 1: Check correlation between Y and X.

Step 2: Estimate intercept and slope, respectively using ordinary least square (OLS) method.

Step 3: Evaluate the significance X and Y using R-square value and ANOVA and F-test

Step 4: Evaluate the significance intercept and slope using t-test.

Step 5: Evaluate the normality error.

In the real world application, the regression models are widely used to measure the causal effect between independent and dependent variables in various domains, such as, sciences, medical, education, economics, others.

3. REMOVING UNCLASSIFIED ELEMENTS USING ROUGH SETS AND REGRESSION MODEL

Eq. (7) can be implemented and generalized to investigate the financial wellbeing of academic staff and its attributes as follows:

$$FWB = a_0 + a_1 FB + a_2 FS \quad (8)$$

Based on Eq. (8), FB is financial behavior of staff, and FS is financial stress as independent attributes which affect to financial wellbeing (FWB) of academic staff. While, $a_0, \ldots, a_3 = 2$ are estimated parameters (coefficients). In this equation, the main objective to determine the significant independent variables and its coefficients by following steps given in Section 2.2. We assume all independent variables affect to the dependent variable. Motivated by previous studies [6][7][8][9], the effect of removing unclassified patients do not yet considered in building rough-regression model. In this paper, we are interested to measure the relationship between numbers of unclassified element toward R-square values of regression model. In this case, the financial wellbeing and its attributes is used to examine the proposed idea by following steps in Cross Industry Standard Process for Data Mining (CRISP-DM) which describes the life cycle of a data mining project in form of six different phases, such as, business (any area) understanding, data understanding, data preparation, modeling, evaluation and deployment [12].

Step 1: Transform numerical to categorical values of financial wellbeing attributes as presented in Table 1.

TABLE I. TRANSFORMATION DATA AND VALUES

Staff Code	Financial Literacy		…	Financial Behavior	
	Num.	Cat.	…	Num.	Cat.
S1	1	Good	…	15	High
S2	2	Very Good	…	15	High
S3	1	Good	…	14	High
…	…	…	…	…	…
Sn	0	Poor	…	9	Low

Step 2: Based on Table 1, determine regression equation (before data reduction) by using the steps given in Section 2.2.

Step 3: Based on Table 1 and Equations (2)-(4), determine lower approximation, upper approximation and boundary region (BR) between conditional attributes and decision attribute.

Step 4: Based on Step 3, eliminate unclassified elements or objects or students [13].

Step 5: Based on Step 4, construct new regression equation (after data reduction) by following number unclassified student accordingly.

Step 6: Based on Step 5, evaluate R-Square of regression models.

Step 7: Determine the best regression models based on Step 6.

4. IMPLEMENTATION

By following Steps 1-7 in Section 3, the proposed idea is examined to the real data set which has been collected from 50 academic staffs in UIN Suska Riau, Indonesia.

Step 1: Construct regression equation before data reduction as shown in Fig 1.

Regression Analysis: FWB versus FB and FS
(before removing of unclassified staff)

```
The regression equation is
       FWB = 9.14 + 0.014 FB + 0.681 FS

Predictor        Coef   SE Coef     T      P
Constant        9.137     2.454  3.72  0.001
FB             0.0138    0.6384  0.02  0.983
FS             0.6808    0.8338  0.82  0.418

         S = 1.74805    R-Sq = 2.4%
```

Fig. 1 Regression model before data reduction

Fig. 1 shows that R-square value of regression model is very low before data reduction. On the other words, the contribution of FB (financial behavior) and FS (financial stress) is 2.4% to financial wellbeing of academic staff level. Otherwise, affecting by any attribute which not mentioned in this study.

Step 2: Determine unclassified staffs using data reduction procedure [4] as presented in Table 2.

TABLE II. DATA REDUCTION BASED ON FWB LEVEL

Staff code	FB	FS	FWB
S10	High	Low	High
S3	Moderate	High	High
S13	Low	Moderate	High
S20	High	High	High
S34	Low	Low	High
⋮	⋮	⋮	⋮
S46	High	Moderate	Medium
S50	Moderate	Moderate	Medium
⋮	⋮	⋮	⋮
S22	Low	Moderate	Low
⋮	⋮	⋮	⋮
S25	High	Low	Low

Based on Table 2, there are three unclassified staffs with codes, S3, S34, and S22. They are should be removed from the data set, because their information cannot be used to support the decision attribute (financial wellbeing) generally. Moreover, we can investigate the effect of number unclassified element to R-square value of regression models by following Steps 3-6.

Step 3: Construct a new regression model with removing 1 unclassified staff as illustrated in Fig. 2.

Regression Analysis: FWB versus FB and FS
(1 unclassified staff removed)

```
The regression equation is
       FWB = 9.12 + 0.102 FB + 0.723 FS
Predictor        Coef   SE Coef     T      P
Constant        9.120     2.454  3.72  0.001
FB              0.102    0.6384  0.02  0.983
FS              0.723    0.8338  0.82  0.418
         S = 1.74805    R-Sq = 20.4%
```

Fig. 2 Regression model after data reduction I

Based on Fig.2, the FB and FS attributes contribute to financial literacy (TFL) is 20.4% after removing 1 unclassified element from data set, namely, S3. Moreover, the removing of 2 unclassified elements and its R-square, namely, S3 and S34 is shown in Fig. 3.

Regression Analysis: FWB versus FB and FS
(2 unclassified staffs removed)

```
The regression equation is
       FWB = 8.49 + 0.110 FB + 0.902 FS
Predictor        Coef   SE Coef     T      P
Constant        8.493     2.287  3.71  0.001
FB             0.1100    0.5931  0.19  0.854
FS             0.9020    0.7772  1.16  0.252
         S = 1.62149    R-Sq = 30.5%
```

Fig. 3 Regression model after data reduction II

Based on Fig.3, 30.5% of FB and FS contribute to financial literacy (TFL) is 30.5% after removing S3 and S34 from data set. Moreover, this percentage can be increased by removing of unclassified elements, such as, S3, S34 and S22. Thus, a new regression model can be reconstructed and shown in Fig. 4.

Regression Analysis: FWB versus FB, FS
(3 unclassified students removed)

```
The regression equation is
       FWB = 6.73 + 0.008 FB + 1.45 FS
Predictor        Coef   SE Coef     T      P
Constant        6.726     2.300  2.92  0.005
FB             0.0085    0.5662  0.01  0.098
FS             1.4547    0.7755  1.88  0.047
         S = 1.54356    R-Sq = 41.1%
```

Fig. 4 Regression model after data reduction III

From Fig.4, the R-square of regression model is significantly increased become 41.1% after removing three unclassified staffs, namely, S3, S34 and S22 from data set. Based on Figures 2 - 4, the number of unclassified element removed against R-square values can be illustrated in Fig.5.

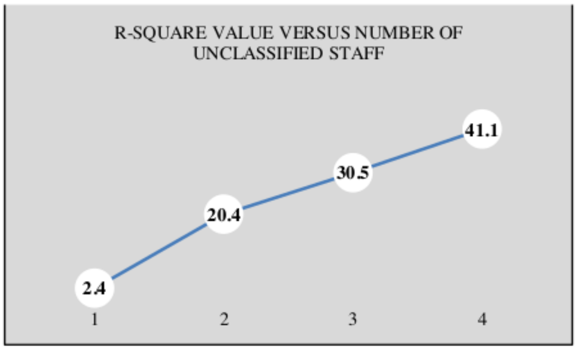

Fig. 5 R-square value against number unclassified staff

Based on Fig. 5, the data elimination strategy is able to improve the performance of regression model significantly. It can be proven by increasing coefficient determination (R-square percentage).

Step 4: Evaluate the best regression model based on Step 3. We can claimed that regression model in Fig.4 is the best model if compared with Figures 1-3 because this model has the higher R-square also. In this case, we proposed the new strategy to improve the contribution of independent variables or conditional attributes to purchase decision without removing independent variables, but elimination of unclassified elements. Moreover, these elements can be denoted as "*vague information in data sets*". The information may positively influence the model building and its accuracy.

The proposed rough-regression model is one of appropriate model in investigating of categorical data, especially for conditional attributes which have high uncertainty in measurement.

5. CONCLUSION

In this paper, we consider an appropriate strategy in improving the conventional regression models with rough sets approximation. In this strategy, the unclassified elements are removed from data set based on data reduction. Its contribution is very significant to increase the dependency between conditional attributes and decision attribute through R-square value. This proposed strategy has been examined to evaluate the effect of removing unclassified students against coefficient determination. Interestingly, the best rough-regression model is obtained after removing of three unclassified staff with codes, S3, S34 and S22. In this paper, the data elimination strategy has a very significant contribution to improve the performance of conventional regression models.

6. ACKNOWLEDGEMENT

This study was funded by DIKTIS-Kemenag, Indonesia under research collaborative grant, 2018.

7. REFERENCES

[1] Pawlak Z.: Rough sets. Int. J. Compt. Inf. Science. **11**, 341-356 (1982)

[2] Tay, F.E.H., Shen, L.: Economic and financial using rough sets model. European J. Operation Research. **141**, 641-659 (2002)

[3] Polkowski, L., Skowron, A.: Rough Sets in Knowledge Discovery 1: Methodology and Applications; Rough Sets in Knowledge Discovery 2: Application, Case Studies, and Sofware Systems. Physica-Verlag, Wurzburg (1998)

[4] Shen, L., Tay, F.E.H., Qu, L., Shen, Y.: Fault Diagnosis using Rough Sets Theory. Computers in Industry (**43**), 61-72 (2000)

[5] Krusinska, E., Slowinski, R., Stefanowski, J.: Discriminant versus Rough Set Approach to Vague Data Analysis. Applied Stochastic Models and Data Analysis (**8**), 43-56 (1992)

[6] Efendi, R., Deris, M. M.: Decision Support Model in Determining Factors and Its Dominant Criteria Affecting Cholesterol Level Based on Rough-Regression. Recent Advances on Soft Computing and Data Mining, 243-251 (2018)

[7] Efendi, R., Samsudin, N.A., Deris, M.M.: Medipre: Medical Diagnosis Prediction using Rough-Regression Approximation. ACM Proceeding on High Compilation, Computing and Communications, 35-39 (2018)

[8] Efendi, R., Samsudin, N. A., Deris, M. M, and Ting Y. G: Flu Diagnosis System Using Jaccard Index and Rough Sets Approaches. Journal of Physic, Conference Series (2018)

[9] Rasyidah, Nawi, N. M and Efendi, R: Rough-Regression Model for Investigating Product Attributes and Purchase Decision. IEEE Xplore Proceeding on Computer and Communication Engineering, (2018).

[10] Herawan, T., Deris, M. M., Abawajy, H.: A Rough Sets Approach for Selecting Clustering Attribute. Knowledge-Based Systems. **23**, 220-231 (2010)

[11] Wooldridge, M.: Introductory econometrics a modern approach. Third Ed. Thomson, South Western, USA (2006)

[12] Chapman, P., Clinton, J., Khabaza, T., Reinartz, T and Wirth, R. The CRISP-DM Process Model. August (2000)

[13] Rissino, S., Torres, G. L. Rough set theory-fundamental concepts, principals, data extraction, and applications, Julio Ponce and Adam Karahoca (Ed), Data Mining and Knowledge Discovery in Real Life App. Inform (2009) 35-58.

nts_in_investigating_of_financial_wellbeing_attributes_using....

ORIGINALITY REPORT

24%	%	24%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	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	7%
2	Cherry R. Gumiran, Joel M. Gumiran. "Applying Design Science Research in the Development of Human Resource Record Management System with Predictive Analysis through Pointing System", Proceedings of the 2019 8th International Conference on Software and Computer Applications - ICSCA '19, 2019 Publication	4%
3	Ching-Hsue Cheng. "A NEW E-LEARNING ACHIEVEMENT EVALUATION MODEL BASED ON ROUGH SET AND SIMILARITY FILTER", Computational Intelligence, 05/2011 Publication	3%

4	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%
5	"Rough Set Theory: A True Landmark in Data Analysis", Springer Science and Business Media LLC, 2009 Publication	2%
6	Kweku-Muata Osei-Bryson. "Organization-Ontology Based Framework for Implementing the Business Understanding Phase of Data Mining Projects", Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008), 01/2008 Publication	2%
7	ASHWANI KUMAR, D. P. AGRAWAL, S. D. JOSHI. "ADVERTISING DATA ANALYSIS USING ROUGH SETS MODEL", International Journal of Information Technology & Decision Making, 2011 Publication	2%
8	"Recent Advances on Soft Computing and Data Mining", Springer Science and Business Media LLC, 2017 Publication	2%

9	Laisheng Wang. "Fuzzy Comprehensive Evaluation Model based on Rough Set Theory", 2006 5th IEEE International Conference on Cognitive Informatics, 07/2006 Publication	2%

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		