EKSAKTA
Journal of Sciences and Data Analysis

E-ISSN: 2720-9326
P-ISSN: 2716-0459
DOI: 10.20885/EKSAKTA.vol4.iss2.art2

Research Article

Analysis of the Character of the Insured by Policy Holders of Private Insurance Companies in Indonesia with the Rough Set Method

Ameilea Chealsea Ekaputrie[1], Achmad Fauzan[1] *

[1] Statistics Department, Faculty of Mathematics and Natural Sciences Universitas Islam Indonesia

[*] *Corresponding author: achmadfauzan@uii.ac.id*

Received: 18 November 2022; Accepted: 12 May 2023; Published: 19 May 2023

Abstract: Insurance provides services in the form of services in facilitating a disaster that may occur based on a history of problems that have occurred. One of the private insurance companies in Indonesia has provided an insurance service program to guarantee employee welfare for policy-holding companies. The purpose of this research is to find out the pattern of association rules from the character of the insured by the policyholder so that it is expected to be able to make valuable information as input or consideration for the policymakers of private insurance companies in Indonesia. This study uses the rough set method as one of the efficient data analysis methods or techniques in database mining or knowledge discovery in relational databases. Rough sets provide algorithms to quickly and easily find hidden patterns in data. The results of the pattern of association rules for the character of the insured by the policyholder have formed as many as 14 rules. The certainty value is intended as a proportionate evaluation amount in order to find out the record of insurance claims that can be chosen by the insured. At the same time, the coverage value is intended as an evaluation amount to produce a decision for the insured to submit a claim record. The probable percentage of all events that are most recommended is seen from the highest coverage value related to record indicators, namely for administration by 45.9%, inspection by 44.1%, others by 40.8%, operation by 48.8%, financing by 46.3%, maintenance by 42%, and compensation of 65.9%.

Keywords: Association Rules, Claim, Rough Set.

Introduction

Business in the insurance sector is a challenge that is different from other industrial fields. Insurance provides services in facilitating a disaster that may occur using forecasting knowledge based on a history of problems that have occurred. Insurance services are also increasingly varied in line with developments in the community's need for risk management and investment management, which are increasingly inseparable from personal life and business activities [1].

One of the private insurance companies in Indonesia has provided an insurance service program to guarantee employee welfare for policy-holding companies. Through group insurance products that offer risk coverage related to health, life, and personal accidents. The main programs providing services for risks are inpatient, outpatient, birth care, and eye care.

The sequence of activities in utilizing insurance facilities by the insured to the insurer is called filing an insurance claim. The main asset owned by every company is business data. Insurance companies have valuable information to determine policies and decisions in doing business, namely insurance claim filing data. Therefore, the success factor of a company in planning a business strategy can be identified through the quality of its data processing.

The Data mining technique is a process to explore knowledge hidden in the database to be used as valuable information [2]. A rough set is an efficient data analysis method or technique in database mining

or knowledge discovery in relational databases [3]. The advantage of the rough set method is that it provides an algorithm for finding hidden patterns in data quickly and easily [4].

Several previous studies related to the rough set method include research by Tamara [5] which applies the rough set method to classify customer data that has the potential to open deposits. Jualiansa *et al.*, [6] implement this method in identifying the level of damage to computer laboratory equipment. Indriani [7] achieves a rough set in determining the purchase of an Android smartphone by consumers, while Sihombing [8] implements the rough set method in predicting the impact of landslides in the province of North Sumatra. Jeprianto [9] researched to obtain rules for selecting applicants for assistance with the house of worship facilities in the Pringsewu district government. Novianti [10] makes predictions regarding the number of new student applicants at the Sinar Harapan private SMK, while Samaray [11] makes predictions related to the results of learning activities. In addition, this research method is often compared with other classification methods, such as the a priori algorithm in the research of Beny Irawan [12] and Jaka Nugraha [3].

Analysis of the insured's character by the policyholder plays an essential role in supporting the policy of the insurer company. If there is a high risk in the future, considering the current conditions and so on, life will coexist with the pandemic. The main objective is to find out the pattern of association rules from the character of the insured by the policyholder so that it is expected to be able to make valuable information as input or consideration for the policymakers of private insurance companies in Indonesia. Based on this, this study focuses on determining the pattern of association rules from the character of the insured by private insurance company policyholders in Indonesia.

Materials and Methods

Materials

The population in this study is claim data for the insured policyholder at a private insurance company in Indonesia for March 2010 to February 2022. The sample used is claim data for the character of the selected insured, namely gender, class, type, and claim record. The sample is determined through the selection of variables that are expected to represent the character of the insured in determining the pattern of filing insurance claims. The data used is 67,531 data. Table 1 is an indicator of the data variables used in the study.

Table 1. Data Indicator

Variable	Indicator
Gender ($\boldsymbol{X_1}$)	Male (M) and Female (F)
Class ($\boldsymbol{X_2}$)	plan A and plan B.
Type ($\boldsymbol{X_3}$)	Hospitalization (H), Outpatient Care (OC), Birth Care (BC), and Eye Care (EC).
Claim record ($\boldsymbol{Y}$)	administration, inspection, others, operations, financing, maintenance, and compensation.

Based on Table 1, sex data is used to determine the gender of the insured, class to determine the level of the facility group owned by the insured, and type to determine the facilities used by the insured. For the insured who has plan A class facilities, of course, it will be different compared to plan B class facilities. Meanwhile, claims records are a grouping of various notes on submitting claims the insured gives to the insurer.

Rough Set Method

The rough set method was first developed by Zdizislaw Pawlak in the early 1980s [13]. The stages in the rough set analysis begin with preparing all the variables consisting of condition attributes and decision attributes. This stage is used to determine the pattern of association rules from the character of the insured by private insurance company policyholders in Indonesia.

The second stage is the data reduction process. Data reduction can be made by eliminating one of the conditional attributes without losing the actual value because redundant attributes will not affect the

E-ISSN: 2720-9326
P-ISSN: 2716-0459

classification results if removed, resulting in a classification set. While the process of generating rules will occur if certain conditions are met [14].

Each row of data in a decision table is a single decision rule. Therefore, the decision table is obtained by uniting all existing data rows based on the similarity of the condition and decision values [15]. The number of association rules obtained can be seen on the rules page as the probability of all the most suggested events. The following is a settlement scheme using the rough set algorithm based on Nofriansyah and Nurcahyo [15]. The rough set algorithm is presented in Figure 1.

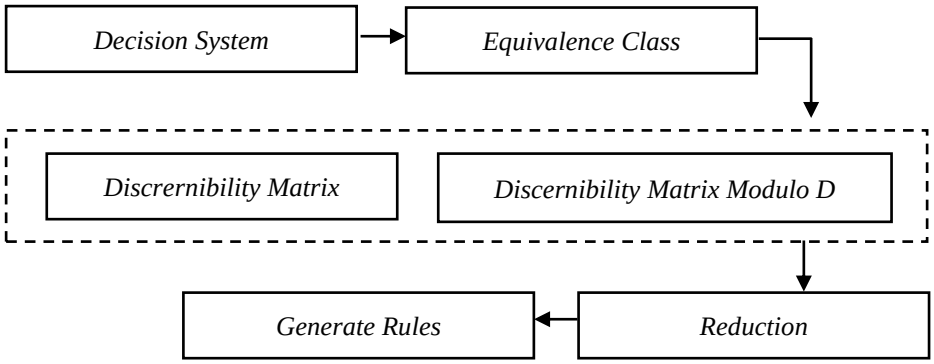

Figure 1. Rough set algorithm

The language of the decision rule is needed to read the data table obtained from the rules page. Data can be read with implications if $\emptyset$ then ψ. $\emptyset$ is the symbol for the condition attribute, and the symbol ψ represents the decision attribute.

Decision-making requires certainty and coverage values from the data used. Certainty factors are the frequencies of ψ_s in $\emptyset$, while Coverage factors are the frequencies of $\emptyset_s$ in ψ. If the decision rule $\emptyset \rightarrow \psi$ n determine a decision in terms of conditions, it can be written if $\pi(\emptyset \mid \psi) = 1$, then the rule is called "certain". If a decision rule $\emptyset \rightarrow \psi$ not determine decision making in terms of conditions that can be written as $0 < \pi(\emptyset \mid \psi) < 1$, then the rule is called "uncertain" [16]. Certainty and Coverage calculations are presented in equations (1) and (2) sequentially.

$$\pi(\psi|\emptyset) = \frac{\text{all cases } \emptyset \text{ and } \psi}{\text{all cases } \emptyset} \quad (1)$$

Certainty determines the comparison ratio of objects that meet the antecedent and fulfills the conclusion to only objects that satisfy the antecedent [17].

$$\pi(\emptyset|\psi) = \frac{\text{all cases } \emptyset \text{ and } \psi}{\text{all cases } \psi} \quad (2)$$

Coverage is the comparison ratio of objects that complete the rule conclusion and also fulfill the antecedents of objects that only fulfill the conclusion.

Result and Discussion

Result of Qualitative Analysis

Descriptive analysis using indicator data of the insured claim record by policyholders of private insurance companies in Indonesia can be described visually, as shown in Figure 2.

E-ISSN: 2720-9326
P-ISSN: 2716-0459

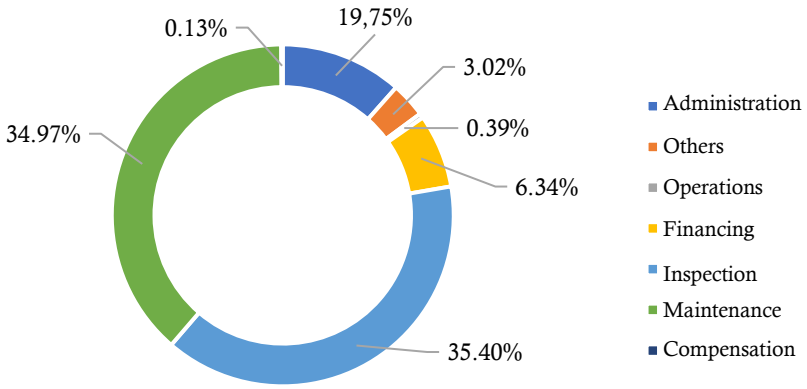

Figure 2. Percentage of claim record indicators

Figure 2 describes the percentage comparison to the state of each indicator in the claim record. The largest percentage, the light blue part, occupies an inspection indicator of 35.40%. Approximately 34.97% filed claims with private insurance companies for treatment indicators, which became the second largest indicator after inspection indicators. While the smallest percentage figure, namely the gray part, occupies a compensation indicator of 0.13%. So from these results, we can determine the classification with association rules using the following rough set method.

Rough set Analysis Result

Rough set analysis is used for variables with variable indicator levels in the set. The character value of the insured participant is based on gender, class, type of treatment, and record of claims with indicators of claims records for inspection, treatment, administration, financing, others, operations, and compensation. The results of the rough set analysis are presented in Table 2**.**

Table 2. Result of association rules

Number	Rules	Certainty	Coverage
1	GENDER (M) AND CLASS (plan A) AND TYPE (H) => CLAIM RECORD (operations) OR CLAIM RECORD (maintenance) OR CLAIM RECORD (others) OR CLAIM RECORD (administration) OR CLAIM RECORD (inspection) OR CLAIM RECORD (financing) OR CLAIM RECORD (compensation)	0.0441, 0.4223, 0.2851, 0.1708, 0.0508, 0.0063, 0.0203	0.4883, 0.0523, 0.4085, 0.0376, 0.0062, 0.0043, 0.6590
2	GENDER (M) AND CLASS (plan A) AND TYPE (H) => CLAIM RECORD (operations) OR CLAIM RECORD (maintenance) OR CLAIM RECORD (others) OR CLAIM RECORD (administration) OR CLAIM RECORD (inspection) OR CLAIM RECORD (financing) OR CLAIM RECORD (compensation)	0.0512, 0.4161, 0.2880, 0.1777, 0.0488, 0.0056, 0.0122	0.4883, 0.0444, 0.3557, 0.0339, 0.0051, 0.0034, 0.3409
3	GENDER (F) AND CLASS (planA) AND TYPE(H) => CLAIM RECORD (operations) OR CLAIM RECORD (maintenance) OR CLAIM RECORD (others) OR CLAIM RECORD (administration) OR CLAIM RECORD (inspection)	0.0759, 0.3544, 0.2658, 0.2025, 0.1012	0.0232, 0.0012, 0.0105, 0.0012, 0.0003
4	GENDER (M) AND CLASS (plan A) AND TYPE (OC) => CLAIM RECORD (maintenance) OR CLAIM RECORD (others) OR CLAIM RECORD (administration) OR CLAIM RECORD (inspection) OR CLAIM RECORD (financing)	0.3496, 0.0051, 0.2142, 0.3707, 0.0602	0.4205, 0.0718, 0.4599, 0.4419, 0.4052

Number	Rules	Certainty	Coverage
5	GENDER (F) AND CLASS (plan A) AND TYPE (OC) => CLAIM RECORD (maintenance) OR CLAIM RECORD (others) OR CLAIM RECORD (administration) OR CLAIM RECORD (inspection) OR CLAIM RECORD (financing)	0.3456, 0.0050, 0.2075, 0.3721, 0.0696	0.4113, 0.0688, 0.4410, 0.4389, 0.4636
6	GENDER (F) AND CLASS (plan B) AND TYPE (OC) => CLAIM RECORD (maintenance) OR CLAIM RECORD (others) OR CLAIM RECORD (administration) OR CLAIM RECORD (inspection) OR CLAIM RECORD (financing)	0.3255, 0.0063, 0.2100, 0.3616, 0.0964	0.0223, 0.0050, 0.0256, 0.0245, 0.0369
7	GENDER (M) AND CLASS (plan B) AND TYPE (OC) => CLAIM RECORD (maintenance) OR CLAIM RECORD (administration) OR CLAIM RECORD (inspection) OR CLAIM RECORD (financing)	0.4074, 0.1851, 0.3703, 0.0370	0.0004, 0.0003, 0.0004, 0.0002
8	GENDER (F) AND CLASS (plan A) AND TYPE (BC) => CLAIM RECORD (maintenance) OR CLAIM RECORD (others) OR CLAIM RECORD (inspection) OR CLAIM RECORD (financing)	0.0232, 0.0650, 0.7679, 0.1437	0.0023, 0.0773, 0.0782, 0.0827
9	GENDER (F) AND CLASS (plan B) AND TYPE (BC) => CLAIM RECORD (maintenance) OR CLAIM RECORD (others) OR CLAIM RECORD (inspection) OR CLAIM RECORD (financing)	0.0190, 0.0285, 0.8285, 0.1238	0.00008, 0.0015, 0.0037, 0.0031
10	GENDER (M) AND CLASS (plan A) AND TYPE (BC) => CLAIM RECORD (others) OR CLAIM RECORD (inspection) OR CLAIM RECORD (financing)	0.1, 0.8, 0.1	0.00050, 0.00034, 0.00024
11	GENDER (M) AND CLASS (plan A) AND TYPE (EC) => CLAIM RECORD (maintenance)	1.0	0.02100
12	GENDER (F) AND CLASS (plan A) AND TYPE (EC) => CLAIM RECORD (maintenance)	1.0	0.01830
13	GENDER (F) AND CLASS (plan B) AND TYPE (EC) => CLAIM RECORD (maintenance)	1.0	0.00543
14	GENDER (M) AND CLASS (plan B) AND TYPE (EC) => CLAIM RECORD (maintenance)	1.0	0.00008

Table 2 shows the results of the character of the insured by the policyholder in as many as 14 patterns of association rules. Table 2 can provide input for stakeholders if they want to determine policies related to filing claims.

As for the acquisition of a certain value as an evaluation amount in proportion to finding out the record of insurance claims that can be chosen by the insured. The first rule has a specific value of 0.0441, which means that as much as 4.4% of claims occur with the association rule "if the insured is male, has to plan A, chooses hospitalization, then the insured can claim insurance records with surgery for a weight of 4.4% or (weight of the proportion of other indicators in the first rule)".

While the coverage value is an evaluation measure to produce a decision for the insured to file a claim record. The first rule with a coverage value of 0.4883, which means that as much as 48.8% results in a claim record decision that occurs with the association rule "chance of 48.8% of operations (other opportunities in the first rule) will result in a decision by the insured male policyholder, has a class plan A, and choose hospitalization" There is an interesting thing in rule 10, where the male insured has class plan A and the type of treatment is birth care. So, it can be seen that the logarithm with the privileges of

class plan A can benefit from the type of treatment used by the insured's spouse, namely the wife of the insured policyholder.

Based on Table 2, it can also be seen the probability of all events that are most recommended for private insurance companies in Indonesia related to the claim record indicator, looking at the highest coverage value obtained as follows:

1. Records of claims related to administration have the greatest chance of 45.9%, where if the insured is male, has class plan A, and chooses outpatient care.
2. Claim records related to inspection have the greatest chance of 44.1%, where if the insured is male, has class plan A, and chooses outpatient care.
3. Claim records related to others have the greatest chance of 40.8%, where if the insured is male, has class plan A, and chooses hospitalization.
4. Records of surgery-related claims have the greatest chance of 48.8%, where if the insured is male and female, has class plan A, and chooses hospitalization.
5. Records of claims related to financing have the greatest chance of 46.3%, where if the insured is female, has class plan A, and chooses outpatient care.
6. Records of claims related to treatment have the greatest chance of 42%, where if the insured is male, has class plan A, and chooses outpatient care.
7. Records of claims related to compensation have a chance of 65.9%, where if the insured is male, has class plan A, and chooses hospitalization.

Conclusion

Based on the research objectives, it was obtained as many as 14 rules. The certainty value is intended as a proportionate evaluation amount to find out the record of insurance claims that can be chosen by the insured. At the same time, the coverage value is intended as an evaluation amount to produce a decision for the insured to submit a claim record. The probable percentage of all events that are most recommended sees the highest coverage value related to record indicators, namely for administration by 45.9%, inspection by 44.1%, others by 40.8%, operation by 48.8%, financing by 46.3%, maintenance by 42%, and compensation of 65.9%.

Acknowledgment

The author thanks to Statistics Department, Faculty of Mathematics and Natural Science, Universitas Islam Indonesia which has assisted in the research process.

References

[1] Kementrian Pertahanan, Penetapan Peraturan Pemerintah Pengganti Undang – Undang Nomor 1 Tahun 2013 Tentang Perubahan Kedua Atas Undang-Undang Nomor 24 Tahun 2003 Tentang Mahkamah Konstitusi Menjadi Undang-Undang. Indonesia; 2014.

[2] Junarto, Datamining Berbasis Web Menggunakan Mahasiswa S-1 Pada Universitas Darma Persada Mahasiswa S-1 Pada Universitas Darma Persada, Peranc. Dan Pembuatan Apl. Datamining Berbas. Web Menggunakan Algoritm. C4.5 Untuk Memprediksi Kelulusan Mahasiswa S-1 Pada Universitas Darma Persada; 2015.

[3] J. Nugraha, M. Muhajir, and R. Febrian, Perbandingan rough set dan algoritma apriori untuk sistem rekomendasi peprpustakaan, Unisda Journal of Mathematics and Computer Science, 4(2) (2018) 25–31.

[4] J. Komorowski, Z. Pawlak, L. Polkowski, and A. Skowron, Rough sets: a tutorial, Norway, 2000.

[5] L. A. V. Tamara, Klasifikasi data nasabah yang berpotensi membuka simpanan deposito mengggunakan algoritma rough set, Universitas Sanata Dharma; 2018.

[6] H. Juliansa, S. Defit, and S. Sumijan, Identifikaasi Tingkat Kerusakan Peralatan Laboratorium Komputer Menggunakan Metode Rough Set, Jurnal RESTI (Rekayasa Sist. dan Teknol. Informasi), 2(1) (2018) 410–415.

[7] U. Indriani, Penerapan Metode Rough Set Dalam Menentukan, Jurnal Teknik Informatika Kaputama, 2(1) (2018) 85–92.

[8] P. Sihombing, Implementasi Metode Rough Set Dalam Memprediksi Dampak Tanah Longsor (Studi Kasus Badan Penanggulangan Bencana Daerah (BPBD) Provinsi Sumatera Utara, Jurnal Riset Komputer (JURIKOM) 6(4) (2019) 407–415.
[9] Jeprianto and A. R. Aziz, Implementasi algoritma rough set dan Naive Bayes untuk mendapatkan rule dalam menyeleksi pemohon bantuan fasilitas rumah ibadah (studi kasus: pemerintah kabupaten Pringsewu), Jurnal Teknologi Komputer dan Sistem Informasi, 3(93) (2020) 74-83.
[10] S. Novianti and P. M. Hasugian, Implementasi Algoritma Rough Set Untuk Memprediksi Jumlah Pendaftar Siswa Baru Pada SMK Swasta Sinar Harapan, Jurnal Teknik Informatika UNIKA Santo Thomas, 6 (2021) 248–259.
[11] S. Samaray, Implementasi Algoritma Rough Set dengan Software Rosetta untuk Prediksi Hasil Belajar, Jurnal Eksplora Informatika, 11(1) (2022) 57–66.
[12] B. Irawan, Analisis kinerja metode rough set dan algoritma apriori dalam identifikasi pola penyakit demam tifoid, Universitas Sumatera Utara; 2016.
[13] Z. Pawlak, Some issues on rough sets, Lect. Notes Comput. Sci., 3100 (2004)1–58.
[14] Z. Pawlak, Rough set theory and its applications, J. Telecommun. Inf. Technol, 3 (2002) 7-10.
[15] D. Nofrinsyah and G. W. Nucahyo, Algoritma data mining dan pengujian. Deepublish, 2015.
[16] I. A. Anastasya, Penerapan metode If-Then rules dari rough set theory pada kecelakaan di lokasi pertambangan, Universitas Islam Indonesia; 2010.
[17] R. Soelaiman, W. Anggraeni, and E. Setiawan, Penerapan rough set quantitative measure pada aplikasi pendukung keputusan. Yogyakarta: Universitas Islam Indonesia; 2008.