International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

Analysis Determination of Data Mining Classification Method for Prediction Level Renewal of Life Insurance Policy

Dyan Aulia Purwanto[1], Riza Adrianti Supono[2]
[1]Master Student, Faculty of Technology and Engineering, Business Information System, University of Gunadarma, Indonesia
[2]Faculty of Technology and Engineering, University of Gunadarma, Indonesia
Email address: [1]hutama08@gmail.com, [2]adrianti@staff.gunadarma.ac.id

Abstract*—This study identified the pattern of policy holders in carrying out policy renewal from 2015 to 2018 sourced from the MITRA ASRI application database. Various data mining algorithms from the classification technique are tested to identify patterns in policy renewal using WEKA Tools, which consist of 11 classification techniques those are algorithm of Naïve Bayes, Neural Network (Multilayer Perceptron), SMO (Support Vector Machine), K-Nearest Neighbor, Decorate, Jrip, PART, ADTree, the C4.5 algorithm, REPTree and Simple CART. Then get the best results. From the results of classification, the C4.5 algorithm is the best algorithm in the accuracy of the renewal class and the AUC value with the greatest average accuracy of the classification technique that is tested for each attribute that influences it. The resulting pattern can be a rule that will be implemented as a new module in the MITRA ASRI application that serves to provide suggestions for filling out renewal policies that are placed automatically. The order of attributes that predominantly appears in each policy renewal are attributes of age, number of family members, type of claim, frequency sum claim, and choice of package.*

***Keywords*—** *Data Mining, Renewal Policy, Life Insurance, Claims, Classification Methods, WEKA.*

I. INTRODUCTION

The use of information technology has proven to facilitate human performance in carrying out a job. This is what causes information technology to be applied in a variety of fields, including the insurance business. Customer service is key in the world of insurance business to continue to grow and innovate in providing a service for products sold, so companies must be able to form and implement new breakthroughs in order to compete by using Information Technology Facilities to be able to have a positive influence on customer service and performance company. Growth can enable organizations to obtain from the scale of profits, to improve their position among industry competitors, and provide more opportunities for professional development and progress to employees (Mello, 2010).

One of the life insurance companies has three operational network divisions including individual, group and sharia life insurance divisions. The company hopes that the three divisions can add services to the community to obtain indirect protection. Competition in the insurance business world requires breakthroughs and strategies to ensure business continuity. One of the company's main assets is customer data and a history of policies available in large quantities. The availability of these data demands the existence of technology that can utilize the data to strengthen business strategies. Prediction of customer interest is very important for insurance companies, because it can determine the interest of customers to renew their insurance policies in contributing to the company for the business continuity of the insurance company.

This results in a need for technology that can use it in exploring new knowledge, which can help in implementing insurance business strategies by utilizing a very large amount of data, companies can certainly find a variety of information. One of the information that can be generated is in the form of information about the renewal or renewal of the insurance policy in submitting customer claims to the type of insurance chosen. The information produced is very important for an insurance company, where with information on the level of customer claims, insurance companies can make decisions in implementing the right strategy to offer other insurance products to prospective customers based on the level of customer claims and can increase prospective new customers in an area.

Data mining technology can be utilized in life insurance customer data and transaction data that occurs in insurance premiums and claims data. From these data, potential customers will be classified in renewing the policy or terminating the policy in filing insurance claims held by the customer. Classification can be interpreted as a process of finding a model or function that describes and distinguishes a class of data objects, with the aim of using a model produced by patterns in decision making of customers in renewing insurance policies.

Data Mining is a very useful technology to help companies find very important information from their data warehouse. Data Mining predicts trends and characteristics of business behavior that are very useful to support important decision making. Automated analysis carried out by data mining outweighs traditional decision support systems that are already widely used. Data Mining explores the database to find hidden patterns, look for predictive information that might be forgotten by business people because it lies outside of their expectations (Santosa, 2007).

This study uses a classification method in the Data Mining approach to obtain algorithms that are suitable in predicting

Dyan Aulia Purwanto and Riza Adrianti Supono, "Analysis determination of data mining classification method for prediction level renewal of life insurance policy," *International Research Journal of Advanced Engineering and Science*, Volume 4, Issue 1, pp. 299-302, 2019.

life insurance customers' decisions by comparing eleven classification techniques used.

II. Previous Research

This research was carried out not apart from the result of previous studies that have been conducted as comparison and study materials. The result of the research that made the comparison is inseparable from the topic of the research, which is about the method of Classification Research.

Based on a journal entitled comparative analysis of the performance classification methods in data mining by (Ari Wibowo, 2015), the result of this study revealed accuracy of the Bagging CART method is better with the classification tree produced is a very complex, because this tree is formed by all predictor variables.

According to (William Frado Pattipeilohy, Arief Wibowo, and Dyah Retno Utari, 2017) in his journal entitled information system modeling and prototyping for predictions of renewal car insurance policy using c4.5 algorithm, in order to predict how likely the customer will renew the insurance policy.

According to (Md. Rafiqul Islam and Md. Ahsan Habib, 2015) in his journal entitled a data mining approach to predict prospective business sectors for lending in retail banking using decision tree, that by applying data mining task procedures for analysis of prospective business sectors in retail banking using the Decision Tree method.

According to (Mhd. Rido Hidayatsyah, 2013) in his journal entitled application of decision tree method in giving loans to debtors with c4.5 algorithm, that the C4.5 algorithm method works well and can be applied in testing the risk of prospective customers.

According to (Ni G. A. P. Harry Saptarini, 2016) in his journal entitled determination of employee based talents using data mining concepts, that the accuracy of the fuzzy C4.5 algorithm with 3 linguistic terms has the highest level of accuracy.

According to (Rizal Amegia Saputra, 2014) in his journal entitled comparison of data mining classification algorithms to predict Tuberculosis (TB): case study of karawang sukabumi health center, that the highest accuracy of Naïve Bayes

III. Research Methods

The research uses a quantitative approach in the data analysis phase, using Classification techniques to obtain a pattern of filling policy renewals from the MITRA ASRI database.

In this methodology, the cycle of data mining process is divided into 6 stages using the CRISP-DM model (Cross Standard Industry for Data Mining) (Rokach, 2010) including business/research understanding phase, data understanding phase, data preparation, modelling phase, evaluation phase, deployment phase, where the dependence between each stage is illustrated by an arrow. Techniques for analyzing data using Quantitative Data, Analysis carried out through policy data uses testing on 11 classification techniques.

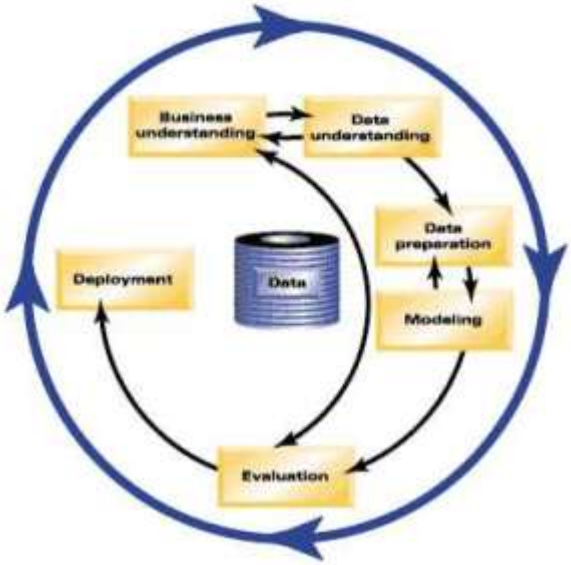

Fig. 1. Cross Industry Standard Process for Data Mining Method (CRISP-DM.org).

A. Proposed System Analysis

Based on the results of analysis, It is expected that after going through each stage a knowledge will be generated that can be utilized for the benefit of the company. The identified pattern will be used as knowledge for the MITRA ASRI application so that it has the ability to be able to provide policy renewal proposals in the Actuarial Department section in preparing a draft change in policy renewal in the insurance company AJB Bumiputera 1912.

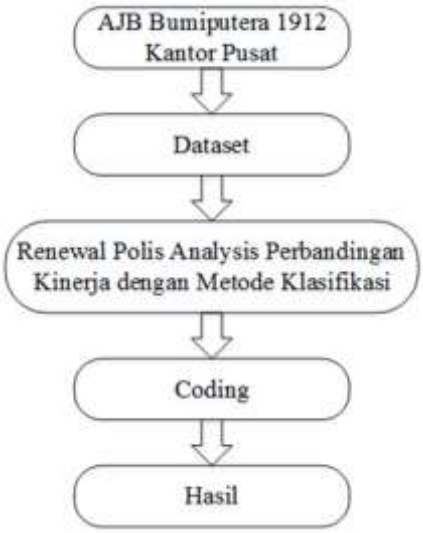

Fig. 2. Model Building.

From the formed model shows that the data is combined into a dataset which in the end is carried out a comparative performance analysis process by using eleven classification techniques to get the best policy renewal results and then implemented into an application program to give a decision to the insurance agent or marketing that is.

Dyan Aulia Purwanto and Riza Adrianti Supono, "Analysis determination of data mining classification method for prediction level renewal of life insurance policy," *International Research Journal of Advanced Engineering and Science*, Volume 4, Issue 1, pp. 299-302, 2019.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

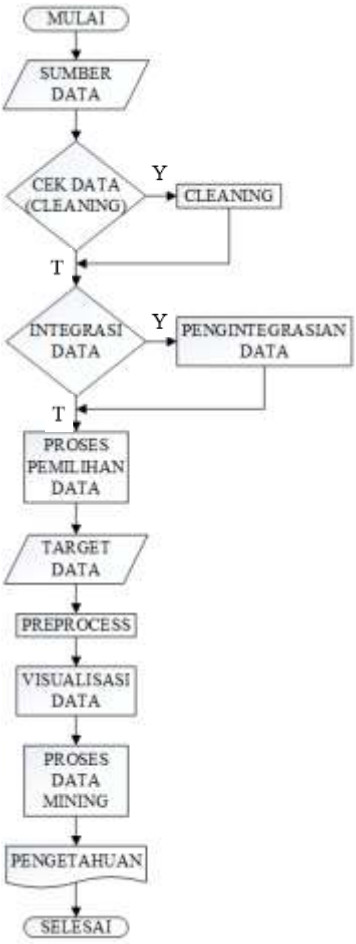

Fig. 3. System Process Flow.

IV. Research Results And Discussion

Train sets and Test sets to be used are claims policy data. For claims policy data is taken from the data of policy holders who have activated their policies and data of expired policy holders where the active period has passed than 1 year. Of all the total data in the actuarial report table which amounted to 445,429 policy data consisting of 5,395 policy data that filed claims and 440,034 policy data that did not make claims. Whereas in this study, only 2,698 policies will be used to make claims or those that do not make claims. Based on data extraction carried out on the MITRA ASRI database in accordance with the existing policy renewal, the number of training sets and testing sets can be seen in the following Table.

TABLE I. Table Number of Record Data Used.

Pembaharuan Polis / *Renewal*	**Polis Klaim (*Train Set*)**	**Polis Klaim (*Test Set*)**
T	1.296	381
Y	862	159

Each policy that does renewal or does not necessarily have different criteria, then for this study the data used to predict policy renewal / renewal is to look at supporting attributes in determining the appropriate method in applying policy renewal predictions. Attributes needed to obtain a dataset whose results will be used in classification techniques, then the policyholders' report attributes can be described in the following table.

TABLE II. Table Attributes of the Policyholder Report.

Atribut	**Tipe Data**	**Range Data / Keterangan**
Usia	Numeric	Usia pemegang polis saat membuka polis baru / Dalam satuan tahun
Jumlah anggota keluarga	Numeric	Jumlah anggota keluarga yang Ter-*cover* pada polis / 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Sum insured	Numeric	Total uang pertanggungan berdasarkan jumlah anggota keluarga / Dalam satuan rupiah
Klaim frekuensi	Numeric	Berdasarkan intensitas pengajuan klaim pada satu polis / 1, 2, 3
Pilihan paket	Teks	Jenis pilihan paket yang dipilih oleh pemegang polis / A1, A2, A3, A4, A5, B1, B2, B3, B4, B5
Penyebab meninggal	Teks	Penyebab meninggal yang diajukan saat klaim / Sakit, kecelakaan, lainnya, NULL
Renewal	Teks	Status dari pembaharuan polis yang telah *expired* / Y, T

A. Comparative Results Analysis

Based on the results of testing the Actuarial Report dataset which contains information about policy holders who carry out policy updates / updates on their policies. By using 80% percentage of training data and 20% testing data, the best prediction results are obtained from the comparative calculation of AUC value for the Naïve Bayes algorithm method, Multilayer Perceptron / Neural Network, SMO (Support Vector Machine), K-Nearest Neighbor, Decorate, Jrip, PART, ADTree, C4.5, REPTree and Simple CART Algorithm used and can be seen in the following table

TABLE III. Table Comparative Results of Accuracy Value Confusion Matrix and AUC on Data Testing Renewal.

Algoritma	**Perbandingan Nilai *Accuracy***	**Perbandingan Nilai *AUC***
Naïve Bayes	79,05%	0,7615
Multilayer Perceptron / Neural Network	92,1%	0,8905
SMO (*Support Vector Machine)*	82,7%	0,768
K-Nearest Neighbor	91,3%	0,8865
Decorate	93,45%	0,922
JRip	92,85%	0,9065
PART	93,2%	0,927
ADTree	89,9%	0,8735
Algoritma C4.5	**93,45%**	**0,9375**
REPTree	92,7%	0,9265
SimpleCart	92,3%	0,88

The table above explains that the Accuracy and AUC comparison of each method and the average value of the calculation results are taken from training data and testing data. It is seen that the highest C4.5 Algorithm method with 93.45% accuracy value and AUC 0.9375. It is known from the eleven algorithm methods that are compared, there are only five algorithm methods which are included in the classification very well, because it has an AUC value between 0.90 - 1.00 consisting of the C4.5, PART, REPTree, Decorate

Dyan Aulia Purwanto and Riza Adrianti Supono, "Analysis determination of data mining classification method for prediction level renewal of life insurance policy," *International Research Journal of Advanced Engineering and Science*, Volume 4, Issue 1, pp. 299-302, 2019.

and JRip algorithms . While the method included in the good classification is the one that has an AUC value between 0.80 - 0.90 which consists of the Multilayer Perceptron / Neural Network, ADTree, K-Nearest Neighbor and SimpleCart methods. And the last method included in the classification is quite good which has an AUC value between 0.70 - 0.80 which consists of the SMO (Support Vector Machine) and Naïve Bayes methods.

Noting the accuracy value obtained from the classification techniques carried out, according to the classification of accuracy values generated in the application of classification techniques is to use the C4.5 algorithm method in Analyzing the Determination of Data Mining Classification Methods for Predicting the Renewal of Life Insurance Policy in the excellent classification category.

V. CONCLUSIONS

The results of the research were carried out by testing twice on the six attributes used in the classification process with renewal attributes as predictor determinants and predictive results and making comparisons with eleven algorithm methods, concluding that based on the acquisition of customer data insurance policies from 2015 until 2018 is correct. After testing a dataset, the algorithm that is suitable for life insurance companies is the highest C4.5 algorithm with an accuracy of 93.45% and AUC 0.9375, so the accuracy rate is almost close to 1.00 or very valid compared to other algorithm methods. So thus the results of the patterns in decision making that can be applied is to use the C4.5 algorithm classification technique by producing the right decision tree.

Suggestions

Based on the above conclusions and the result of the research conducted, there are several things that need to be considered by It is necessary to develop a web service that can bridge WEKA's JAVA-based data mining tools with PHP-based MITRA ASRI, so that both can communicate with each other and can be used to facilitate the process of inputting data from MITRA ASRI to WEKA and output predictions from WEKA to MITRA ASRI with automatically.

REFERENCES

[1] Ali, Z. (2008). *Hukum Asuransi Syariah*. Jakarta: Sinar Grafika.
[2] Amrin, A. (2006). *Asuransi Syariah (keberadaan dan kelebihan di tengah asuransi konvensional)*. Jakarta: PT Elex Media Komputindo.
[3] Amrin, A. (2011). *Meraih Berkah Melalui Asuransi Syariah Ditinjau Dari Perbandingan Dengan Asuransi Konvensional*. Jakarta: PT. Elex Komputindo.
[4] Anupam Shukla, R. T., Rahul Kala. (2010). *Real Life Applications of Soft Computing*. CRC Press, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton: Taylor and Francis Group, LLC.
[5] Depdikbud. (1996). *Kamus Besar Bahasa Indonesia*. Jakarta: Balai Pustaka.
[6] Efraim Turban, J. E. A., Ting-Peng Liang. (2005). *Decision Support Systems and Intelligent Systems* (7th ed.). New Jersey, U.S.A.: Prentice-Hall, Inc.
[7] Haykin, S. (2009). *Neural Networks and Learning Machines* (3rd ed.). Upper Saddle River, NewJersey 07458: Pearson Education, Inc.
[8] Hermawati, F. A. (2013). *Data Minig* (I ed.). Yogyakarta: Andi Offset.
[9] Undang-Undang Republik Indonesia Nomor 2 Tahun 1992 Tentang Usaha Perasuransian, (1992).
[10] Iqbal, M. (2005). *Asuransi Umum Syariah Dalam Praktik*. Jakarta: Gema Insani Press.
[11] Kamber, J. H. M. (2006). *Data Mining: Concept and Technique Second Edition*. Oxford, UK: Diane Cerra, Elsevier Science.
[12] Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. Hoboken, New Jersey: John Wiley & Sons, Inc.
[13] Leo Breiman, J. F., Charles J. Stone, R.A. Olshen. (1984). *Classification and Regression Trees* (1st ed.). New York, NY: Chapman And Hall: Chapman and Hall/CRC.
[14] Luthfi, K. E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset.
[15] Mello, J. A. (2010). *Strategic Human Resource Management* (3rd ed.). United States of America: South-Western College Pub.
[16] Nello Cristianini, J. S.-T. (2000). *An Introduction to Support Vector Machines*. Cambridge University Press: The Press Syndicate of the University of Cambridge.
[17] Neuman, W. L. (2000). *Social Research Methods, Qualitative and Quantitative Approach* (4th ed.). USA: Allyn & Bacon.
[18] Pasaribu, C. (2004). *Hukum Perjanjian Dalam Islam*. Jakarta: Sinar Grafika.
[19] Pehlivanli, Y. D. D. A. C. (2011). The Comparison of Data Mining Tool.
[20] Pramudiono, I. (2003). Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data. *IlmuKomputer.com*.
[21] Puspitaningrum, D. (2006). *Pengantar Jaringan Syaraf Tiruan*. Yogyakarta: Andi.
[22] Roger J., L., M.D., Ph.D. (2000). An Introduction to Classification and Regression Tree (CART) Analysis. *Presented at the 2000 Anual Meeting of Society For Academy Emergency Medicine in San Fransisco, California*.
[23] Rokach, O. M. L. (2010). *Data Mining and Knowledge Discovery Handbook* (2 nd ed.). Israel: Springer.
[24] Saleh, A. (2015). Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. 207.
[25] Santosa, B. (2007). *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
[26] Sula, M. S. (2004). *Asuransi Syariah: Life and General: Konsep dan Sistem Operasional*. Jakarta: Gema Insani.
[27] Tan, P.-N. (2006). *Introduction to Data Mining*. Boston, MA, USA: Addison-Wesley Longman PublishingCo.Inc.
[28] Thabtah, F. A. (2007). A Review of Associative Classification Mining. *The Knowledge Enineering Review, 22:1*, 37-65.
[29] Umar, H. (2008). *Metodologi Penelitian Untuk Skripsi dan Tesis Bisnis*. Jakarta: PT. Raja Grafindo Persada.
[30] Vulandari, R. T. (2017). *Data Mining: Teori dan Aplikasi Rapidminer*. Yogyakarta: Gava Media.
[31] Wibowo, A. (2011). Prediksi Nasabah Potensial Menggunakan Metode Klasifikasi Pohon Biner. 7.

Dyan Aulia Purwanto and Riza Adrianti Supono, "Analysis determination of data mining classification method for prediction level renewal of life insurance policy," *International Research Journal of Advanced Engineering and Science*, Volume 4, Issue 1, pp. 299-302, 2019.