JURNAL RISET INFORMATIKA **P-ISSN: 2656-1743 |E-ISSN: 2656-1735**
Vol. 5, No. 4. September 2023 DOI: https://doi.org/10.34288/jri.v5i4.246
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Naive Bayes and Decision Tree Algorithms for BRI Life Sharia Insurance Product Classification

Rika Astuti

Teknologi Informasi, Universitas Siber Indonesia, Jakarta Selatan
Jakarta, Indonesia
rika.astuti93@gmail.com

Abstract
With the variety of sharia insurance products currently available, competition in the insurance world is increasing. The author differentiates between the Naive Bayes and Decision Tree classification methods. Using the Naive Bayes method and decision trees, the author can classify sharia products by looking at the highest accuracy value. The author uses a collection of data obtained by retrieving the dataset on the core Sharia business system in the form of a web application called SKALA (Administrator) used by user support staff. Data is classified using the naive Bayes method, decision trees. The research results using decision trees have an accuracy value of 91.95%. The prediction data with the highest precession value is found in the 1159 Insurance product with a precession of 92.48%. The lowest value is found in Life Insurance products, with a class precision value of 82.99%. Based on research results using the Naïve Bayes Classifier and Decision Tree methods, the classification of Sharia Insurance Products is based on the highest accuracy value. From the percentage results, classifying BRI Life individual sharia insurance products using the Naïve Bayes classifier has an accuracy value of 81.66% and a Decision Tree of 91.95%.

Keywords: Decision Tree, NaiveBayes Classifier, Classification

Abstrak
Dengan beragamnya produk asuransi syariah yang ada saat ini, persaingan dalam dunia asuransi semakin meningkat. Penulis membedakan metode klasifikasi Naive Bayes dan Decision Tree. Dengan menggunakan metode Naive Bayes dan pohon keputusan, penulis dapat mengklasifikasikan produk syariah dengan melihat nilai akurasi tertinggi. Penulis menggunakan kumpulan data yang diperoleh dengan mengambil dataset pada inti sistem bisnis syariah berupa aplikasi web bernama SKALA (Administrator) yang digunakan oleh staf pendukung pengguna. Data diklasifikasikan menggunakan metode naif Bayes, pohon keputusan. Hasil penelitian menggunakan pohon keputusan mempunyai nilai akurasi sebesar 91,95%. Dimana data prediksi dengan nilai presesi tertinggi terdapat pada produk Asuransi 1159 dengan presesi sebesar 92,48%. Nilai terendah terdapat pada produk Asuransi Jiwa dengan nilai presisi kelas sebesar 82,99%. Berdasarkan hasil penelitian dengan menggunakan metode Naïve Bayes Classifier dan Decision Tree, dengan pengklasifikasian Produk Asuransi Syariah berdasarkan nilai akurasi tertinggi. Dari hasil persentase terlihat bahwa pengklasifikasian produk asuransi syariah perorangan BRI Life menggunakan Naïve Bayes classifier mempunyai nilai akurasi sebesar 81,66% dan Decision Tree sebesar 91,95%.

Kata Kunci: Decision Tree, Naive Bayes, Klasifikasi

INTRODUCTION

In today's digital era, data mining has become a strategic step used in various industries. Data classification and reporting are two examples of data mining applications (Andarista & Jananto, 2022). Current developments in information technology allow the collection of vast amounts of data (Mukaromah, 2020). Apart from that, many classification algorithms are widely used. The classification model process determines the relationships of a class, grouping classes and determining (Rachmawati, 2007). Relationships are based on the categories that have been created (Herdiansah et al., 2022). One way to classify is through information extraction model analysis, which produces information class classifications. Because service class influences service quality, classification is essential in determining service class (Hermanto, Mustopa, & Kuntoro, 2020). The results obtained from a classification using the naive Bayes method and decision trees on Sharia insurance products can be implemented to determine accuracy values (Ubaedi & Djaksana,

2022). In today's digital era, easy data access and storage is critical. Planning, decision-making, monitoring, and evaluation can be done by processing data into information (Fatmawati, Kramanandita, & Miza, 2022). A specific method is required to extract the data that will be processed to obtain helpful information for the user. Data mining is a method for automatically extracting knowledge from large amounts of data by finding interesting patterns (Elfaladonna & Rahmadani, 2019). This research determines the Accuracy of Sharia insurance products using the Naive Bayes classifier and decision tree methods. The Sharia business web application is a Kcore system that can monitor individual Sharia insurance participant data. This method is used as a classification to determine the accuracy value with the highest percentage (Tangkelayuk & Mailoa, 2022). The author collects insurance product data to classify Sharia insurance products. The methods used are Naïve Bayes and decision trees. The Naïve Bayes method is a classification method that can be used to predict the probability of membership in a class. The Naïve Bayes classifier method has advantages over other methods (Sari, Firdausi, & Azhar, 2020). Large-scale data (Yudistira, 2021). Algorithm comparisons were conducted to determine the classification with the highest accuracy value based on Sharia insurance products (Akbar, Saputra, Maulaya, Hidayat, & Rahmaddeni, 2022). Decision Tree is an additional algorithm that can be used to carry out product classification analysis, and the process is to convert data from a table into a tree and then convert the tree into rules (Normawati & Prayogi, 2021). The advantage of the Naive Bayes algorithm is that it is easy and fast to predict classes with a test data set (Fitri, Yuliani, Rosyida, & Gata, 2020). The Naive Bayes classification method is quite simple in probability classification. Adding frequencies and combining values is almost the same as the naive Bayes method according to the attributes (Tandian, Laia, & Saputra, 2019). For several classes, decision trees are a simple example of a classification technique. After the data is divided, the algorithm will choose the attributes that best separate the data based on specific criteria (Setiyani, Wahidin, Awaludin, & Purwani, 2020). Decision trees to solve problems regarding data classification (Penira, Samsudin, & Ikhwan, 2022). This algorithm is handy when the data set is large. To use the decision tree algorithm, it must provide a dataset where each data point has a bunch of features and a corresponding result or decision. The algorithm then analyzes the data and creates a tree structure based on discovered patterns and relationships. Once a decision tree is created, you can use it to make predictions or classify new data points (Sartika & Sensuse, 2017). The research results show that the Naive Bayes Classifier can classify sharia insurance products into several classes. The results were obtained using the Naive Bayes Classifie method with an accuracy value of 97%, a Precision value of 99.8%, a Recall of 97%, and an Error Rate of 3%. The next stage is to determine how much data there is for each attribute. Then, the respective accuracy values are 80.137%, and the precision values are 78.0822%. And the combined accuracy percentage value was obtained at 83.5616%, namely 90%. The results of the confusion matrix evaluation show that the student and employee classes have the best consumer status. Previous research used the Naive Bayes classifier and Decision Tree methods. The data that will be classified is the type of payment and sales transactions.

RESEARCH METHOD

The research flow steps are as follows:

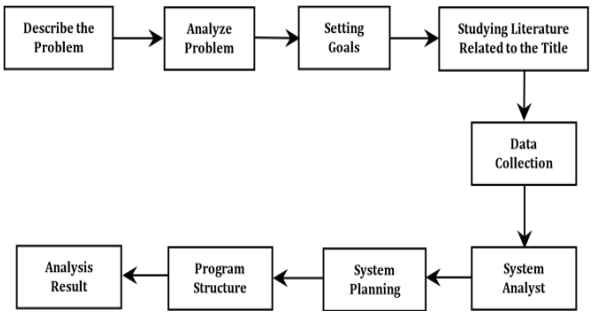

Figure 1. Research flow

Based on the process flow in Figure 1, the stages are as follows.

1. Problem specification.

Analysis using the Naive Bayes and Decision Tree method to determine the correct accuracy value.

2. Problem Analysis

The problem of very diverse insurance products has led to the development of competition in the insurance world. This creates issues with customers' interest in purchasing individual insurance products.

3. Setting Goals

This research aims to determine the Decision Tree's accuracy value and Naive Bayes' accuracy value in classifying individual Sharia insurance products.

4. literature review

The author conducted literature by searching for several sources obtained from journals and books.

5. data collection techniques

The data collection technique is taken through historical data on individual Sharia insurance closures from 01 April 2022 to 30 June 2022. This data contains 34 attributes and 1,565 records.

6. System Analysis

system analysis of problem-solving by classifying using Decision Trees and the Naive Bayes method.

7. System Design

System Design describes the research flow that will be created.

8. Program Structure

a structured program from data retrieval techniques, data collection, and data processing to data implementation.

9. Analysis Results

The research results show that insurance product classification using Decision Tree and Naive Bayes can be implemented to determine the accuracy value of individual Sharia insurance product classification.

Naive Bayes Classifier

The Naive Bayes method is a method for data mining classification in decision-making. The flowchart structure includes internal nodes representing tests or attributes at each stage. Each branch represents a result for an attribute, while the path from the leaves to the root represents the rules for classification. Decision tree data is usually displayed in table form with attributes. Attributes represent parameters called criteria. Displaying data in tabular form, processed into several attributes, results from Accuracy.

Naive Bayes formula is as follows:

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}$$

Where:
If it is known that event B has occurred,
P(A|B) is the probability of event A occurring, and
P(B|A) is the probability of event B occurring.
P(A) is the probability of event A occurring, and
P(B) is the probability that event B will happen.

Table 1. Naive Bayes Classifier Proses

NO	Product	Verification status	JUP	Contribution	Payment
1	BRILIFE INSURANCE	Verify	720000000	6000000	Monthly
2	BRILIFE INSURANCE	Verify	60000000	500000	Monthly
3	BRILIFE INSURANCE	Verify	48000000	200000	Monthly
4	BRILIFE INSURANCE	Verify	120000000	1000000	Monthly
5	BRILIFE INSURANCE	Verify	360000000	1500000	Monthly
6	BRILIFE INSURANCE	Verify	120000000	1000000	Monthly
7	BRILIFE INSURANCE	Verify	240000000	1000000	Monthly
8	BRILIFE INSURANCE	Verify	60000000	500000	Monthly
9	BRILIFE INSURANCE	Verify	120000000	1000000	Monthly
10	BRILIFE INSURANCE	Verify	60000000	500000	Monthly
11	BRILIFE INSURANCE	Verify	480000000	600000	annual
12	BRILIFE INSURANCE	Verify	120000000	3000000	annual
13	BRILIFE INSURANCE	Verify	40000000	1000000	annual
14	BRILIFE INSURANCE	Verify	40000000	1000000	annual
15	BRILIFE INSURANCE	Verify	40000000	1000000	annual

The results of the Daily Report data testing where the verified status is active will then be spread across all products. Sharia products include BRI Life Link, Student Syariah, Profitera, and Purnadana Syariah. If the police status of Surrender is the product of the Sharia Life Link Protection, all results of the classification model are according to the modeling process. As seen in the figure 2.

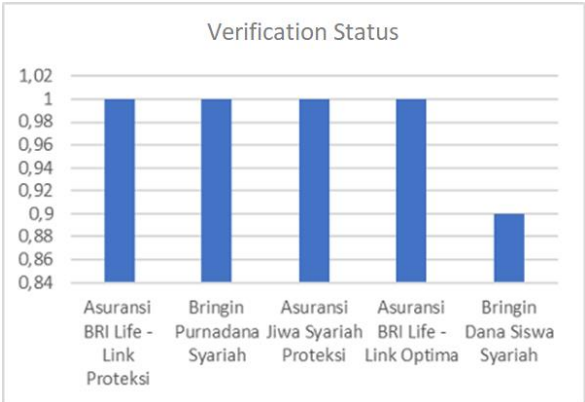

Figure 2. Naive Bayes Method Classification Graph

Then, from the data available, most insurance customers in all Sharia products have verified status. For the customer status that has not been verified, there is BRI Life Link Protection. Sharia products include BRI Life Link, Student Syariah, Profitera, and Purnadana Syariah.

Table 2. Naïve Bayes Classifier Accuracy Percentage

Accuracy 81,66%						
PRODUCT	TRUE BRI LIFE INSURANCE	TRUE LIFE INSURANCE	TRUE WANT DONATIONS	TRUE RETIRED	TRUE BRILLIANT INSURANCE	CLASS PRECESSION
INSURANCE PRODUCT	1131	133	45	4	0	86,14%
LIFE INSURANCE PRODUCTS	43	25	2	0	0	35,71%
PRODUCTS WANT DONATIONS	3	1	2	0	0	33,33%
RETIRED PRODUCTS	15	34	0	112	0	69,14%
BRILLIANT INSURANCE PRODUCT	2	1	3	0	8	57,14%
CLASS RECALL	94,64%	12,89%	3,85%	95,55%	100.00%	

From Accuracy, with the insurance analysis, customers' interest in BRI Life insurance product services is 81.66%.

Table 3. Naive Bayes Classifier Accuracy Error Percentage

Classification Error 18,34%						
PRODUCT	TRUE BRI LIFE INSURANCE	TRUE LIFE INSURANCE	TRUE WANT DONATIONS	TRUE RETIRED	TRUE BRILLIANT INSURANCE	CLASS PRECESSION
BRILLIANT INSURANCE PRODUCT	1131	133	45	4	0	86,14%
LIFE INSURANCE PRODUCTS	43	25	2	0	0	35,71%
PRODUCTS WANT DONATIONS	3	1	2	0	0	33,33%
RETIRED PRODUCTS	15	34	0	112	0	69,14%
BRILLIANT INSURANCE PRODUCT	2	1	3	0	8	57,14%
`Class Recall	94,64%	12,89%	3,85%	95,55%	100.00%	

From the Error Classification, with an analysis of insurance customers less interested in BRI Life insurance product services, it is 18.34%.

Table 4. Table 5. Percentage Classification Error of the Decision Tree

Classification Error: 8,05%						
PRODUCT	TRUE BRI LIFE INSURANCE	TRUE LIFE INSURANCE	TRUE WANT DONATIONS	TRUE RETIRED	TRUE BRILLIANT INSURANCE	CLASS PRECESSION
BRILLIANT INSURANCE PRODUCT	1160	70	25	0	0	92,48%
LIFE INSURANCE PRODUCTS	22	122	2	1	0	82,99%
PRODUCTS WANT DONATIONS	4	1	25	0	0	83,33%
RETIRED PRODUCTS	0	0	0	115	0	100,00%
BRILLIANT INSURANCE PRODUCT	0	1	0	0	8	88,89%
CLASS RECALL	97,82%	62,89%	48,00%	99,14%	100.00%	

From the Error Classification, with analysis of insurance customers less interested in BRI Life insurance product services, it is 8.05%.

The highest test results for insurance products are JUP ≤ 1 billion, namely for BRI Life Link Protection Sharia products, and a contribution value of ≤ 20 million for buying BRI Life Link Protection Sharia products. As seen in Figure 3.

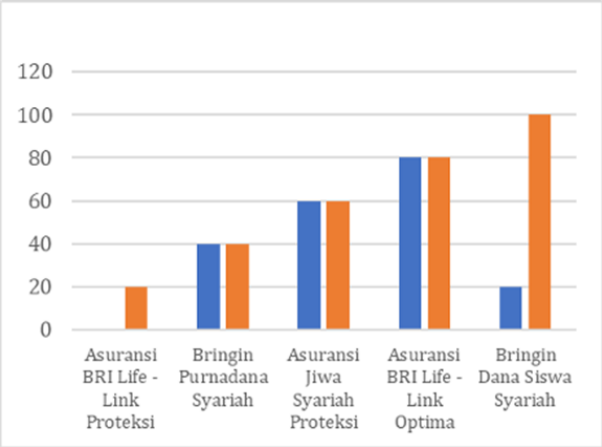

Figure 3. Decision Tree Method Classification Graph

For data that is ≤ 20 million spread all products and > 20 million ≤ 80 Million There are products Profitera, BRI Life Link Optima Shariah, and Shariahi Students and > 80 Million there are products profiter and BRI life link optima shariah. As seen in Figure 8.

Table 5. Percentage Accuracy of the Decision Tree

Accuracy 91,95 %						
PRODUCT	TRUE BRI LIFE INSURANCE	TRUE LIFE INSURANCE	TRUE WANT DONATIONS	TRUE RETIRED	TRUE BRILLIANT INSURANCE	CLASS PRECESSION
BRILLIANT INSURANCE PRODUCT	1160	70	25	0	0	92,48%
LIFE INSURANCE PRODUCTS	22	122	2	1	0	82,99%
PRODUCTS WANT DONATIONS	4	1	25	0	0	83,33%
RETIRED PRODUCTS	0	0	0	115	0	100,00%
BRILLIANT INSURANCE PRODUCT	0	1	0	0	8	88,89%
CLASS RECALL	97,82%	62,89%	48,00%	99,14%	100.00%	

From Accuracy, with the insurance analysis, customers' interest in BRI Life insurance product services is 91.95% accurate.

CONCLUSIONS AND SUGGESTIONS

Conclusion

BRI Life Sharia individual insurance product classification has an accuracy value of 91.95% for the Decision Tree. BRI Life Sharia's unique insurance product classification has an accuracy value of 81.66% for Naive Bayes. Decision Trees and Naive Bayes can be implemented to determine the value accuracy of the classification of BRI Life Sharia individual insurance products.

Suggestion

BRI Life can implement product classification on a macro basis to assess insurance customer interest. Further research can be conducted with a more extensive data set to produce higher value accuracy.

REFERENCES

Akbar, F., Saputra, H. W., Maulaya, A. K., Hidayat, M. F., & Rahmaddeni, R. (2022). Implementasi Algoritma Decision Tree C4.5 dan Support Vector Regression untuk Prediksi Penyakit Stroke. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *2*(2), 61–67. https://doi.org/10.57152/MALCOM.V2I2.426

Andarista, R. R., & Jananto, A. (2022). Penerapan Data Mining Algoritma C4.5 Untuk Klasifikasi Hasil Pengujian Kendaraan Bermotor. *Jurnal Tekno Kompak*, *16*(2), 29–43. https://doi.org/10.33365/JTK.V16I2.1525

Elfaladonna, F., & Rahmadani, A. (2019). Analisa Metode Classification-Decission Tree dan Algoritma C. 45 untuk Memprediksi Penyakit Diabetes dengan Menggunakan Aplikasi Rapid Miner. *SINTECH (Science and Information Technology) Journal*, *2*(1), 10–17. https://doi.org/10.31598/SINTECHJOURNAL.V2I1.293

Fatmawati, T., Kramanandita, R., & Miza, R. (2022). Rancangan Implementasi Enterprise Resource Planning (ERP) pada Sistem Pengelolaan Sales Order PT Jaya Mandiri Indotech. *Jurnal Teknologi Dan Manajemen*,

20(1), 33–44. https://doi.org/10.52330/JTM.V20I1.49

Fitri, E., Yuliani, Y., Rosyida, S., & Gata, W. (2020). Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine. *Jurnal Transformatika*, *18*(1), 71–80. https://doi.org/10.26623/TRANSFORMATIKA.V18I1.2317

Herdiansah, A., Indra Borman, R., Nurnaningsih, D., Aristo, A., Sinlae, J., Ridlo, R., & Hakim, A. (2022). Klasifikasi Citra Daun Herbal Dengan Menggunakan Backpropagation Neural Networks Berdasarkan Ekstraksi Ciri Bentuk. *JURIKOM (Jurnal Riset Komputer)*, *9*(2), 388–395. https://doi.org/10.30865/JURIKOM.V9I2.4066

Hermanto, H., Mustopa, A., & Kuntoro, A. Y. (2020). Algoritma Klasifikasi Naive Bayes Dan Support Vector Machine Dalam Layanan Komplain Mahasiswa. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, *5*(2), 211-220. https://doi.org/10.33480/jitk.v5i2.1181

Mukaromah, E. (2020). Pemanfaatan Teknologi Informasi dan Komunikasi dalam Meningkatkan Gairah Belajar Siswa. *Indonesian Journal of Education Management & Administration Review*, *4*(1), 175–182. https://doi.org/10.4321/IJEMAR.V4I1.4381

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, *5*(2), 697–711. https://doi.org/10.30645/J-SAKTI.V5I2.369

Penira, A., Samsudin, S., & Ikhwan, A. (2022). Data Mining Memprediksi Kecenderungan Calon Nasabah dalam Memilih Produk Asuransi PT AJS Bumiputera Medan. *Infoman's : Jurnal Ilmu-Ilmu Manajemen Dan Informatika*, *16*(1), 61–70. Retrieved from http://ejournal.stmik-sumedang.ac.id/index.php/infomans/article/view/403

Rachmawati, I. N. (2007). Pengumpulan Data Dalam Penelitian Kualitatif: Wawancara. *Jurnal Keperawatan Indonesia*, *11*(1), 35–40. https://doi.org/10.7454/JKI.V11I1.184

Sari, V. R., Firdausi, F., & Azhar, Y. (2020). Perbandingan Prediksi Kualitas Kopi Arabika dengan Menggunakan Algoritma SGD, Random Forest dan Naive Bayes. *Edumatic: Jurnal Pendidikan Informatika*, *4*(2), 1–9. https://doi.org/10.29408/edumatic.v4i2.2202

Sartika, D., & Sensuse, D. I. (2017). Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *3*(2), 151–161. https://doi.org/10.35957/JATISI.V3I2.78

Setiyani, L., Wahidin, M., Awaludin, D., & Purwani, S. (2020). Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review. *Faktor Exacta*, *13*(1), 35–43. https://doi.org/10.30998/FAKTOREXACTA.V13I1.5548

Tandian, C., Laia, Y., & Saputra, A. (2019). Penerapan Data Mining Dalam Memprediksi Pemenang Klub Sepak Bola Pada Ajang Liga Champion Dengan Algoritma C.45. https://doi.org/10.34012/jusikom.v2i2.397

Tangkelayuk, A., & Mailoa, E. (2022). The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *9*(2), 1109–1119. https://doi.org/10.35957/JATISI.V9I2.2048

Ubaedi, I., & Djaksana, Y. M. (2022). Optimasi Algoritma C4. 5 Menggunakan Metode Forward Selection Dan Stratified Sampling Untuk Prediksi Kelayakan Kredit. *JSiI (Jurnal Sistem Informasi)*, *9*(1), 17–26. https://doi.org/10.30656/JSII.V9I1.3505

Yudistira, N. (2021). Peran Big Data dan Deep Learning untuk Menyelesaikan Permasalahan Secara Komprehensif. *EXPERT: Jurnal Manajemen Sistem Informasi Dan Teknologi*, *11*(2), 78–89. https://doi.org/10.36448/expert.v11i2.2063