International Journal of Information System & Technology
Akreditasi No. 158/E/KPT/2021 | Vol. 8, No. 6, (2025), pp. 338-344

Comparison of Naive Bayes and C4.5 Methods with Particle Swarm Optimization on Customer Loyalty Classification

Embun Fajar Wati[1], Elvi Sunita Perangin-Angin[2], Luthfi Indriyani[3]
[1,2,3]Universitas Bina Sarana Informatika, Indonesia
Email: embun.efw@bsi.ac.id[1], elvi.evt@bsi.ac.id[2], luthfi.lfy@bsi.ac.id[3]

Abstract

The Company attaches great importance to customer loyalty for the sustainability of the Company. Loyal customers will buy many times and provide great profits. In this study, the decision tree method or C4.5 and naïve bayes were used with PSO optimization for customer classification which aims to design a strategy in decision-making towards disloyal customers. Some of the stages carried out are data load into MS. Excel, data cleaning from noise, data selection as many as 238 obtained from previous research with several attributes, including, namely age, annual income, purchase amount, region, purchase frequency, and loyalty score, as well as data transformation, namely each attribute is grouped into 2 with their own criteria, data testing by modeling data through Rapidminer, Data evaluation by examining the values of accuracy, precision, recall, and AUC. Both methods have the same accuracy value of 96.67% and the same recall value of 100%. For the precision value, there is a difference of 0.6% and the precision decision tree value is higher than the naïve Bayes which is 96.16%. As for the AUC value, it is higher naïve bayes, which is 0.922 with the difference from the decision tree of 0.059. It can be concluded that the two methods in processing customer loyalty data in this study have the same accuracy, so both methods are equally good.

Keywords: *naïve bayes, decision tree, particle swarm optimization, c4.5, loyalty.*

1. Introduction

One of the business strategies that can be used by companies is the use of information technology as a medium to collect transaction data [1]. Customers have a tendency to continue subscribing to the company or otherwise unsubscribing [2]. Loyalty is used to describe a customer's willingness to continue subscribing to a Company in the long term, by buying and using its goods and services repeatedly, preferably exclusively, and willingly recommending the company's products to their friends [3]. Customer loyalty is a person's loyalty to a particular good or service.

One technique that can be used to identify customer loyalty tendencies is by data classification [4]. Classification is a learning technique that can predict the target class of a predetermined class, in addition to that classification can help find hidden data models during data analysis [5]. Customer loyalty can not only provide consistent revenue, but it creates a good reputation in the market and minimizes marketing costs to acquire new customers [6].

Researchers have conducted research on customer loyalty with several data mining methods, namely naïve bayes, decision tree, and knn methods which produce the highest percentage of naïve bayes which is 96.67% [7]. Loyal customers need to be accurately predicted in order to help in the decision-making process [8].

Data mining is a term that can be used to describe the discovery of knowledge in a database [9]. Data mining involves using specific algorithms to extract data patterns [10]. The methods that will be used in this study are naïve Bayes and decision tree or C4.5. Decision Tree is a method that partitions data recursively based on specific features to form a simple and interpretive decision tree structure [11]. C4.5 is a type of classification algorithm that allows in creating an

ISSN : 2580-7250

understandable decision tree [12]. Naive Bayes uses probability theory, which is predicting future chances or probabilities based on past experience [13].

In the case of previous studies that used both methods and the same optimization, the results of naïve Bayes were higher than the decision tree of 100% and 97% [14]. Previous research on student graduation using the c4.5 method resulted in a percentage of 97% without c4.5 journal optimization, while naïve bayes with psi optimization had a value on student graduation without comparing it to other methods 100% [15].

Both methods were chosen because they are easy to use to work independently. It is a function of data that has no other characteristics or lack of other characteristics in the same data [16]. Both of these methods will be optimized with PSO (Particle Swarm Optimization). PSO (particle swarm optimization) is used for attribute selection and helps reduce irrelevant input dimensions and improve prediction performance [17]. In this study, data will be taken from sales with several attributes or features taken from previous research [7].

2. Research Methodology

The methods or algorithms used in this study are Naïve Bayes, Decision Tree and optimization with PSO. The software used is Rapidminer. Some of the stages carried out in this study are [15]:

a) Data load
The data is transferred to the form of MS. Excel which has an extension .xls for easy processing, because one of the file forms supported by Rapidminer is ms. Excel.

b) Data cleaning
Data cleanup is carried out on noise data such as incomplete or blank, which is 0 and does not match the other data, and incomplete or duplicate data.

c) Data selection
Data was obtained from the previous study which totaled 238 and used several attributes, including, namely age, annual income, purchase amount, region, purchase frequency, and loyalty score.

d) Data transformation
The data transformation on each attribute or feature is changed to 2 types in table 1.

Table 1. Data Transformation

Age	Teenager	<=25
	Mature	>25
Annual income	Big	>=50.000
	Small	<50.000
Purchase amount	Much	>=500
	Less	<500
Purchase frequency	Often	>=15
	Seldom	<15
Loyalty score	Loyal	>=5
	Not loyal	<5

e) Data testing
Data testing is done with Rapidminer software with data mining modeling.

f) Data evaluation
Data evaluation by looking at the results of AUC, recall, and precision values from modeling in the previous stage.

3. Results and Discussion

238 customer data have been loaded or input into MS. Excel in table 2. The data used is complete, all do not need to be cleaned because they have been used in previous

research [7]. The data used for training and testing in 10 cross validations were divided into 10 data parts or subsets of the same size. Repeated 10 times in the training and testing process.

Table 2. Customer Data

Age	Annual Income	Purchase Amount	Region	Purchase Frequency	Loyalty Score
teenager	small	less	North	seldom	not loyal
mature	big	less	South	often	loyal
mature	big	much	West	often	loyal
teenager	small	less	East	seldom	not loyal
mature	small	less	North	seldom	not loyal
…	…	…	…	…	…
mature	big	less	West	often	loyal
mature	big	less	North	often	loyal
mature	big	much	South	often	loyal
mature	big	less	West	often	loyal
mature	big	less	North	often	loyal

Data mining modeling in Figure 1 with rapidminer using naïve bayes and decision tree methods and optimized with PSO, training and testing with 10-fold cross validation.

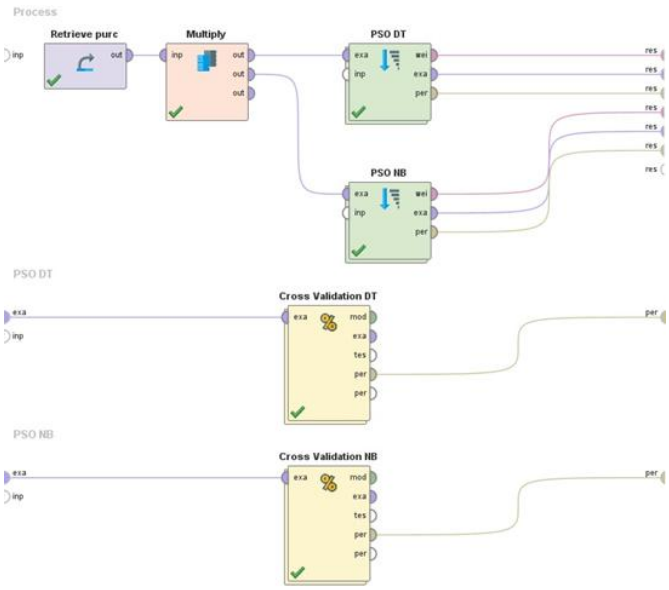

Figure 1. Processes with PSO and Cross Validation

In Figure 1, process modeling with decision tree and naïve Bayes using multiply and PSO (particle swarm optimization) with 10-fold cross validation.

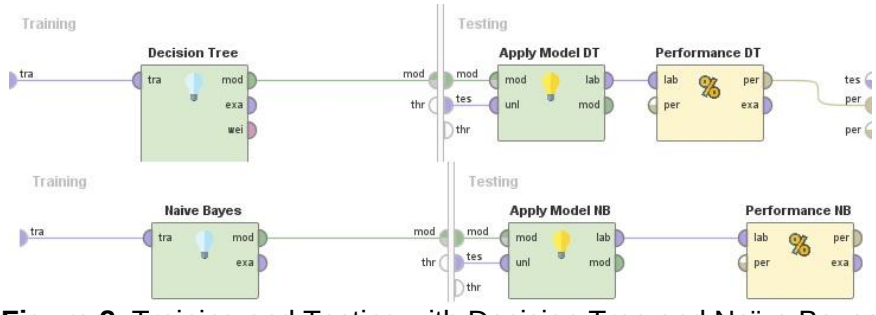

Figure 2. Training and Testing with Decision Tree and Naïve Bayes

In figure 2, training is carried out with algorithms or data mining methods, namely decision tree and naïve bayes and testing with applied models and performance. The accuracy of the decision tree can be seen in figure 3 and the accuracy of naïve bayes can be seen in figure 4.

accuracy: 96.67% +/- 4.30% (micro average: 96.64%)

	true not loyal	true loyal	class precision
pred. not loyal	43	0	100.00%
pred. loyal	8	187	95.90%
class recall	84.31%	100.00%	

Figure 3. Decision Tree Accuracy

The accuracy produced in Figure 3 is 96.67% with the decision tree method, while the accuracy produced by the Naïve Bayes method can be seen in Figure 4.

accuracy: 96.67% +/- 3.83% (micro average: 96.64%)

	true not loyal	true loyal	class precision
pred. not loyal	43	0	100.00%
pred. loyal	8	187	95.90%
class recall	84.31%	100.00%	

Figure 4. Naïve Bayes Accuracy

The accuracy produced in figure 4, namely with the naïve bayes method, is the same as 96.67%. The precision produced by the decision tree can be seen in figure 5.

precision: 96.16% +/- 4.79% (micro average: 95.90%) (positive class: loyal)

	true not loyal	true loyal	class precision
pred. not loyal	43	0	100.00%
pred. loyal	8	187	95.90%
class recall	84.31%	100.00%	

Figure 5. Precision Decision Tree

The precision produced in figure 5 is by the decision tree method of 96.16%. Meanwhile, the precision produced by naïve bayes is shown in figure 6.

precision: 96.10% +/- 4.46% (micro average: 95.90%) (positive class: loyal)

	true not loyal	true loyal	class precision
pred. not loyal	43	0	100.00%
pred. loyal	8	187	95.90%
class recall	84.31%	100.00%	

Figure 6. Precision Naïve Bayes

In figure 6, the precision value of the naïve Bayes method is 96.10%. The recall generated by the decision tree method is shown in figure 7.

recall: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: loyal)

	true not loyal	true loyal	class precision
pred. not loyal	43	0	100.00%
pred. loyal	8	187	95.90%
class recall	84.31%	100.00%	

pred. not loyal

Figure 7. Recall Decision Tree

The recall in figure 7 is generated by the decision tree method by 100%, while the recall is generated by the naïve bayes method in figure 8.

recall: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: loyal)

	true not loyal	true loyal	class precision
pred. not loyal	43	0	100.00%
pred. loyal	8	187	95.90%
class recall	84.31%	100.00%	

Figure 8. Recall Naïve Bayes

In figure 8, it can be seen that the results of modeling using the naïve Bayes method have the same recall as the decision tree, which is 100%. The resulting AUC diagram can be seen in figure 9. In figure 3 – figure 8, there are predictions of 8 loyal customers, but in real terms not loyal. The 8 customers need to be recorded and reanalyzed, in order to help in marketing strategies and in the decision-making process to retain these loyal customers.

Figure 9. AUC Decision Tree

The AUC value produced by the decision tree method in figure 9 is 0.922%. Meanwhile, the AUC value produced by the naïve Bayes method is in figure 10.

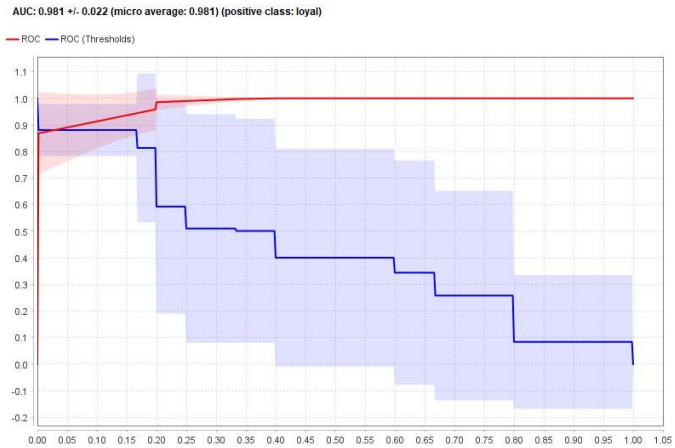

Figure 10. AUC Naïve Bayes

In figure 10, there is an AUC value resulting from the naïve Bayes method, which is 0.981%. The comparison of accuracy, precision, recall, and AUC values is shown in table 1.

Table 1. Comparison of Accuracy, Precision, Recall, and AUC Values

Method	Accuracy	Precision	Recall	AUC
Decision Tree	96.67%	96.16%	100%	0.922
Naïve Bayes	96.67%	96.10%	100%	0.981

From table 1, it is known that both methods, there is the same accuracy value of 96.67% and the same recall value of 100%. For the precision value, there is a difference of 0.6% and the precision decision tree value is higher than the naïve Bayes which is 96.16%. As for the AUC value, the naïve bayes is higher, which is 0.922 with the difference from the decision tree of 0.059.

It can be concluded that the two methods in data processing in this study have the same accuracy, so both methods are equally good. The cause of the similarities is the possibility of a model that is too simple or perhaps too complex to produce a perfect accuracy of 100%. The contribution of this study is to compare 2 methods, namely the decision tree with naïve bayes to classify customer loyalty with PSO optimization.

4. Conclusion

Both methods have the same accuracy value of 96.67% and the same recall value of 100%. For the precision value, there is a difference of 0.6% and the precision decision tree value is higher than the naïve Bayes which is 96.16%. As for the AUC value, it is higher naïve bayes, which is 0.922 with the difference from the decision tree of 0.059. It can be concluded that the two methods in processing customer loyalty data in this study have the same accuracy, so both methods are equally good. In the next research, it is hoped that research will be carried out on different cases with the same methods and optimizations, so that the best method is known.

References

[1] Basri, Windu Gata, And Risnandar, "Analisis Loyalitas Pelanggan Berbasis Model Recency, Frequency, Dan Monetary (Rfm) Dan Decision Tree Pada Pt. Solo," *Jurnal Teknologi Informasi Dan Ilmu Komputer (Jtiik)*, Vol. 7, No. 5, Pp. 943–950, 2020.

[2] Yohana Tri Widayati, Yani Prihati, And Stephanus Widjaja, "Analisis Dan Komparasi Algoritma Naïve Bayes Dan C4.5 Untuk Klasifikasi Loyalitas Pelanggan Mnc Play Kota Semarang," *Transformtika*, Vol. 18, No. 2, Pp. 161–172, 2021.

[3] Khotibul Umam, Diah Puspitasari, And Acmad Nurhadi, "Penerapan Algoritma C4.5 Untuk Prediksi Loyalitas Nasabah Pt Erdika Elit Jakarta," *Jurnal Media Informatika Budidarma*, Vol. 4, No. 1, Pp. 65–71, 2020.

[4] Musthofa Galih Pradana And Pujo Hari Saputro, "Komparasi Metode Naïve Bayes Dan C4.5 Dalam Klasifikasi Loyalitas Pelanggan Terhadap Layanan Perusahaan," *Indonesian Journal Of Business Intelligence*, Vol. 3, No. 1, Pp. 20–24, 2020.

[5] Ratih Yulia Hayuningtyas, "Analisa Klasifikasi Loyalitas Pelanggan Menggunakan Algoritma Naïve Bayes," *Jurnal Tekinkom*, Vol. 7, No. 2, Pp. 891–898, 2024.

[6] Ardiane Rossi Kurniawan Maranto, Lily Damayanti, And Irvan Rahul Ramadika, "Perbandingan Algoritma C4.5 Dan Naïve Bayes Dalam Prediksi Loyalitas Pelanggan," *Bit-Tech (Binary Digital - Technology)*, Vol. 7, No. 2, Pp. 396–405, 2024.

[7] Embun Fajar Wati, Elvi Sunita Perangin-Angin, And Luthfi Indriyani, "Customer Loyalty Classification With Comparison Of Naive Bayes, C4.5, And Knn Methods," *International Journal Of Information System & Technology*, Vol. 8, No. 3, Pp. 177–185, 2024.

[8] Ni Wayan Wardani *Et Al.*, "Prediksi Pelanggan Loyal Menggunakan Metode Naïve Bayes Berdasarkan Segmentasi Pelanggan Dengan Pemodelan Rfm," *Jurnal Manajemen Dan Teknologi Informasi*, Vol. 12, No. 2, Pp. 113–124, 2022.

[9] Muhamad Fahrurrozi And Sriyanto, "Prediksi Loyalitas Karyawan Dengan Menggunakan Fitur Optimasi Pembobotan Pso Dan Algoritma Decision Tree C4.5," *Jurnal Teknika*, Vol. 17, No. 1, Pp. 195–203, 2023.

[10] Embun Fajar Wati, Elvi Sunita Perangin-Angin, And Anggi Puspita Sari, "Prediction Of Student Graduation Using The K-Nearest Neighbors Method," *International Journal Of Information System & Technology*, Vol. 7, No. 3, Pp. 211–216, 2023.

[11] Khaerul Anam, Ade Rizki Rinaldi, And Fathurrohman, "Komparasi Algoritma Machine Learning Dalam Klasifikasi Loyalitas Nasabah Bank Berbasis Particle Swarm Optimization," *Jati (Jurnal Mahasiswa Teknik Informatika)*, Vol. 8, No. 4, Pp. 8212–8218, 2024.

[12] Embun Fajar Wati And Biktra Rudianto, "Penerapan Algoritma Knn, Naive Bayes Dan C4.5 Dalam Memprediksi Kelulusan Mahasiswa," *Jurnal Format*, Vol. 11, No. 2, Pp. 168–175, 2022.

[13] Rianti Yunita Kisworini And Muhammad Akbar Setiawan, "Peningkatan Performa Naivee Bayes Dengan Seleksi Atribut Menggunakan Chi Square Untuk Klasifikasi Loyalitas Pelanggan Grab," *Journal Of Informatics, Information System, Software Engineering And Applications*, Vol. 2, No. 2, Pp. 69–75, 2020.

[14] E. F. Wati, A. P. Sari, E. T. Alawiah, M. H. Siregar, And B. Rudianto, "Particle Swarm Optimization Comparison On Decision Tree And Naive Bayes For Pandemic Graduation Classification," In *2nd International Conference On Advanced Information Scientific Development (Icaisd)*, 2021, Pp. 1–11.

[15] Embun Fajar Wati, Elvi Sunita Perangin-Angin, And Anggi Puspita Sari, "Improved Naive Bayes Algorithm With Particle Swarm Optimization To Predict Student Graduation," *International Journal Of Information System & Technology*, Vol. 7, No. 6, Pp. 386–391, 2024.

[16] E. F. Wati And B. Sudrajat, "Application Of Naive Bayes Method For Diagnosis Of Pregnancy Disease," *International Journal Of Information System & Technology*, Vol. 6, No. 1, Pp. 93–100, 2022.

[17] Moch. Rizki Kurniawan, Puspita Nurul Sabrina, And Ridwan Ilyas, "Prediksi Customer Churn Pada Perusahaan Telekomunikasi Menggunakan Algoritma C4.5 Berbasis Particle Swarm Optimization," *Jati (Jurnal Mahasiswa Teknik Informatika)*, Vol. 7, No. 5, Pp. 3369–3375, 2023.