International Journal of Information System & Technology
Akreditasi No. 158/E/KPT/2021 | Vol. 9, No. 1, (2025), pp. 16-20

Customer Loyalty Classification With Random Forest Algorithm

Anggi Puspita Sari[1], Astrid Noviriandini[2], Sifa Fauziah[3]
[1,2,3]Universitas Bina Sarana Informatika, Indonesia
Email: anggi.apr@bsi.ac.id[1], astrid.asv@bsi.ac.id[2], sifa.saz@bsi.ac.id[3]

Abstract

Customer loyalty is very important for the survival of the company. Because with customers who have customer loyalty, they will make purchases regularly. Customer loyalty needs to be maintained to increase profits. The method is to classify loyal customers with non-loyal ones, in order to retain loyal customers and set strategies for non-loyal customers. The method used is classification with random forest with cleaning stages that can clean data from noise or empty data or data that does not match, selection that can select some data to be processed for classification, transformation that can change data into two or three formats, classification with random forest with split validation using testing data and training data and with rapidminer software. Evaluation by checking the results of the classification with random forest in the form of accuracy, precision, recall, and AUC. The results of the classification show from the accuracy table that the prediction of loyal and true loyal customers is 129 more than the prediction of not loyal and true not loyal customers which is 32. The accuracy result is 96.41% which shows that the data is really accurate with very high results. The recall result is 98.47%, while the precision result is 96.99%.

Keywords: *random forest, customer loyalty, classification, prediction, non loyal*

1. Introduction

The development of technological progress and tight market competition means that today's businesses are required to shift their company's attention from products to customers. [1]. Customer retention is a measurement focus in an industry with increasing competition. [2]. Customer satisfaction has a significant influence on customer loyalty and customer satisfaction is also a measure of customer loyalty to a company [3]. Research into customer loyalty is urgent.

The main components of the fourth industrial revolution include artificial intelligence (AI), the Internet of Things, and big data [4]. AI (artificial intelligence) based predictive methods are currently being used to increase customer loyalty [5]. Customer data contains valuable insights into customer behavior and preferences that can lead businesses to effective personalized marketing and service optimization [6]. Optimal business benefits can be achieved by capturing customer data and producing important information [7].

The purpose of this research is to apply machine learning classification techniques to predict customer loyalty in companies so that companies can use the results to create possible solutions for customer relationship management. [8]. This study also aims to investigate the main factors that influence customer loyalty [9]. The resulting model can predict customer loyalty by looking at existing feature criteria and become a decision support system for management [10]. These findings confirm that machine learning approaches can be used to understand patterns in greater depth [11].

Random Forest predicts the response based on the most frequently occurring class as a result of predictions from k classification trees [12]. The selection of random forest for customer loyalty prediction classification because in several studies it produces a high percentage. The research on telecommunications customer loyalty produces an accuracy value of 81% and an ROC AUC value of 0.89 [13]. Customer loyalty research at insurance companies produces accuracy values above 90% and AUC values above 0.90 [14].

ISSN : 2580-7250

2. Research Methodology

The research methodology used is taken from previous research, namely [15]:

(a) Cleaning
Cleaning is performed on data that is mixed with noise, such as incomplete or empty information, namely 0, as well as data that is inconsistent with other information, and data that has deficiencies or is duplicated.

(b) Selection
Information was obtained from previous research involving 238 data and utilizing several characteristics, such as age, annual income, number of transactions, location, shopping frequency, and loyalty level.

(c) Transformation
Transformation that can change data into two formats. Data changes in each attribute or feature section are divided into two categories as shown in table 1.

Table 1. 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

(d) Classification
Classification with random forest with split validation using testing data and training data and with rapidminer software

(e) Evaluation
Evaluation by checking the results of the classification with random forest in the form of accuracy, precision, recall, and AUC.

3. Results and Discussion

The existing data is comprehensive, there is no need to clean it completely because it has been used in previous research. Data classification using split validation.

Table 2. Customer Loyalty 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

In table 2, the attributes or features of customer loyalty data consist of age, annual income, purchase amount, region, purchase frequency, loyalty score. Customer loyalty data classification modeling with random forest and using rapid miner is shown in Figure 1.

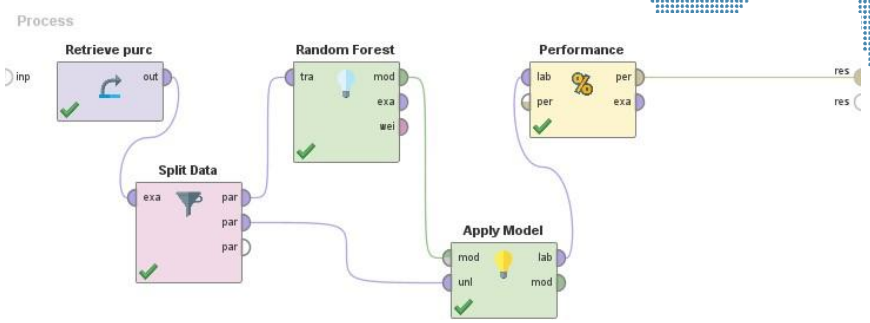

Figure 1. Modeling with random forest and split validation

In Figure 1, the classification is carried out with solid data with a ratio of 0.3:0.7, which means that the training data is 30% and the testing data is 70%.

accuracy: 96.41%

	true not loyal	true loyal	class precision
pred. not loyal	32	2	94.12%
pred. loyal	4	129	96.99%
class recall	88.89%	98.47%	

Figure 2. Accuracy results

In Figure 2, the accuracy results show a value of 96.41% with 32 predictions of not loyal and true not loyal and 129 predictions of loyal and true loyal. For the remaining data, namely 4 predictions of loyal and true not loyal and 2 predictions of not loyal and true loyal, they do not affect decision making and strategy by management because the amount of data is small.

precision: 96.99% (positive class: loyal)

	true not loyal	true loyal	class precision
pred. not loyal	32	2	94.12%
pred. loyal	4	129	96.99%
class recall	88.89%	98.47%	

Figure 3. Precision Results

In Figure 3, there is a precision result of 96.99% which shows that the loyal prediction ratio is positive compared to the overall positive predicted results.

recall: 98.47% (positive class: loyal)

	true not loyal	true loyal	class precision
pred. not loyal	32	2	94.12%
pred. loyal	4	129	96.99%
class recall	88.89%	98.47%	

Figure 4. Recall Results

In Figure 4, there is a recall result of 98.47% which shows that the ratio of positive loyal predictions compared to the total data that is truly positive. In Figure 5, there is an AUC result on the ROC curve of 0.996 which shows good performance of the random forest classifier in distinguishing positive and negative classes. From the overall results, it was found that the percentage of accuracy, precision, recall, and AUC had very good values, which proves that the random forest algorithm can classify customer loyalty data very well.

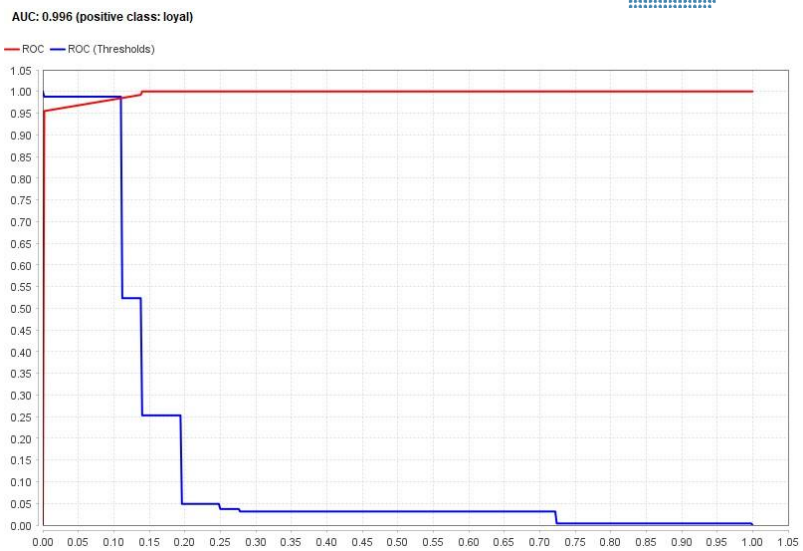

Figure 5. AUC Results

4. Conclusion

The accuracy results show a value of 96.41% with 32 predictions of not loyal and true not loyal and 129 predictions of loyal and true loyal. The precision result of 96.99% which shows that the loyal prediction ratio is positive compared to the overall positive predicted results. The recall result of 98.47% which shows that the ratio of positive loyal predictions compared to the total data that is truly positive. The AUC result on the ROC curve of 0.996 which shows good performance of the random forest classifier in distinguishing positive and negative classes. From the overall results, it was found that the percentage of accuracy, precision, recall, and AUC had very good values, which proves that the random forest algorithm can classify customer loyalty data very well. In further research, it is hoped that the same data can be processed with the same algorithm but with different validation and addition methods.

References

[1] M Ainur Syawaludin, Rifki Hidayat, And Nurmalitasari, "Prediksi Churn Pelanggan Multinational Bank Menggunakan Algoritma Machine Learning," *Simpatik: Jurnal Sistem Informasi Dan Informatika*, Vol. 4, No. 2, Pp. 89–97, 2024.

[2] Y. Suhanda, L. Nurlaela, I. Kurniati, A. Dharmalau, And I. Rosita, "Predictive Analysis Of Customer Retention Using The Random Forest Algorithm," *Tiers Information Technology Journal*, Vol. 3, No. 1, Pp. 35–47, Jun. 2022, Doi: 10.38043/Tiers.V3i1.3616.

[3] Fannisa Salsabila Pratiwi, Mula Agung Barata, And Aprillia Dwi Ardianti, "Implementasi Metode Smote Dan Random Over-Sampling Pada Algoritma Machine Learning Untuk Prediksi Customer Churn Di Sektor Perbankan," *Jurnal Sistem Informasi Dan Informatika (Simika)*, Vol. 8, No. 1, Pp. 87–98, 2025.

[4] Y. Lee, K. Na, J. Rhim, And E. Kim, "Primary Determinants And Strategic Implications For Customer Loyalty In Pet-Related Vertical E-Commerce: A Machine Learning Approach," *Systems*, Vol. 13, No. 3, P. 175, Mar. 2025, Doi: 10.3390/Systems13030175.

[5] M. T. Atay And M. Turanli, "Analysis Of Customer Churn Prediction Using Logistic Regression, -Nearest Neighbors, Decision Tree And Random Forest

Algorithms," *Adv Appl Stat*, Vol. 92, No. 2, Pp. 147–169, Dec. 2024, Doi: 10.17654/0972361725008.

[6] Zhuoran Li, "Customer Segmentation And Churn Prediction Based On K-Means And Random Forest: A Case Study Of E-Commerce Data," *Eurasia Journal Of Science And Technology*, Vol. 7, No. 2, 2025, Doi: 10.61784/Ejst3071.

[7] Maya Cendana And Silvester Dian Handy Permana, "Analisis P Erbandingan A Lgoritma Naive Bayes, J48, Dan Random Forest Tree Dalam P Eningkatan Loyalitas P Elanggan Umkm Dengan Voucher B Elanja," *Jurnal Integrasi*, Vol. 11, No. 2, Pp. 140–145, 2019.

[8] I. Z. P. Hamdan And M. Othman, "Predicting Customer Loyalty Using Machine Learning For Hotel Industry," *Journal Of Soft Computing And Data Mining*, Vol. 3, No. 2, Aug. 2022, Doi: 10.30880/Jscdm.2022.03.02.004.

[9] Yaoyi Huang, Jiaxun Chen, Bobin Zhang, And Feiyu Jin, "Study On The Key Factors Influencing Consumer Loyalty In Fresh Food E-Commerce Platforms Based On Svm-Rf Model," *Highlights In Business, Economics And Management*, Vol. 49, Pp. 120–125, 2025.

[10] S. Amalia, I. Deborah, And I. N. Yulita, "Comparative Analysis Of Classification Algorithm: Random Forest, Spaarc, And Mlp For Airlines Customer Satisfaction," *Sinergi*, Vol. 26, No. 2, P. 213, Jun. 2022, Doi: 10.22441/Sinergi.2022.2.010.

[11] A. Maehendrayuga, A. Setyanto, And Kusnawi, "Analisa Prediksi Turnover Karyawan Menggunakan Machine Learning," *Bit-Tech*, Vol. 7, No. 2, Pp. 648–659, Dec. 2024, Doi: 10.32877/Bt.V7i2.1999.

[12] Ayunda Pratiwi, Khairil Anwar Notodiputro, And Hari Wijayanto, "Pemodelan Loyalitas Konsumen Susu Pertumbuhan Dalam Mengikuti Program Rewards Menggunakan Metode Random Forest Dan Neural Network," *Xplore*, Vol. 2, No. 2, Pp. 41–48, 2018.

[13] Lisa Nusrotul Wakhidah, Akhmad Khanif Zyen, And Buang Budi Wahono, "Evaluation Of Telecommunication Customer Churn Classification With Smote Using Random Forest And Xgboost Algorithms," *Journal Of Applied Informatics And Computing (Jaic)*, Vol. 9, No. 1, Pp. 89–95, 2025.

[14] R. Pratama, M. I. Herdiansyah, D. Syamsuar, And A. Syazili, "Prediksi Customer Retention Perusahaan Asuransi Menggunakan Machine Learning," *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, Vol. 12, No. 1, Pp. 96–104, Mar. 2023, Doi: 10.32736/Sisfokom.V12i1.1507.

[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] 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.