

ANALISIS DAN KOMPARASI ALGORITMA NAÏVE BAYES DAN C4.5 UNTUK KLASIFIKASI LOYALITAS PELANGGAN MNC PLAY KOTA SEMARANG

Yohana Tri Widayati¹, Yani Prihati², Stephanus Widjaja³

¹Universitas AKI Semarang

Jl. Imam Bonjol 15-17 Semarang, (024) 3552555, email: yohana.tri@unaki.ac.id

²Universitas AKI Semarang

Jl. Imam Bonjol 15-17 Semarang, (024) 3552555, email: yani.prihati@unaki.ac.id

³STIMIK AKI Pati

Jl. Kamandowo 13 Pati, (0295) 382087, email: stephanuswidjaja@gmail.com

ARTICLE INFO

Article history:

Received 08 August 2020

Received in revised form 05 January 2021

Accepted 27 January 2021

Available online 31 January 2021

ABSTRACT

Customer loyalty is an important factor to the survival of a company. The problem facting today is related to the descreasing number of customer loyalty that happened to PT MNC Play. If this is allowed to continue it is not impossible this will endanger the continuity of the company's business.

Keep in mind the factors that cause customers to have loyal and disloyal status, data mining classification techniques can be used to classify loyal and disloyal customers. Many data mining classification algorithms can be used, so it needs to be comparative to know the accuracy of each algorithm, the algorithm used is C4.5 and Naïve Bayes. The data used were 28,899MNC Play customer of Semarang City.

The results of the classification process were evaluated using cross validation, confusion matrix, ROC Curve to find the more accurate data mining classification algorithm to determine loyal and disloyal customers.

Keywords: Classification, Customers Loyalty, Naïve Bayes, C4.5

1. Introduction

Loyalitas pelanggan adalah kesetiaan seseorang terhadap suatu barang atau jasa tertentu. Tingkat loyalitas pelanggan yang mereka miliki merupakan aset perusahaan yang berharga nilainya. [1]. Loyalitas pelanggan akan menjadi kunci sukses, tidak hanya dalam jangka pendek tetapi keunggulan bersaing secara berkelanjutan.

MNC Play adalah salah satu perusahaan penyedia jasa internet provider di kota Semarang. Masalah loyalitas pelanggan yang terjadi di MNC Play saat ini adalah banyaknya pelanggan yang sudah tidak menggunakan layanan internet dari MNC Play dan memilih tidak berlangganan lagi.

Received August 08, 2020; Revised January 05, 2021; Accepted January 27, 2021

Klasifikasi loyalitas pelanggan digunakan untuk mempermudah perusahaan mengetahui pelanggan loyal dan tidak loyal dengan melakukan analisa secara periodic setiap bulan. Oleh karena itu diperlukan metode yang cocok untuk mendapatkan hasil yang akurat untuk mengetahui tingkat klasifikasi pelanggan mana yang perlu menjadi perhatian.

Belum diketahuinya algoritma yang paling akurat untuk melakukan klasifikasi loyalitas pelanggan MNC Play kota Semarang, sehingga perlu dilakukan analisis dan komparasi untuk mengetahui algoritma apa yang paling akurat pada klasifikasi loyalitas pelanggan MNC Play.

Dalam jurnal ini akan menerapkan teknik algoritma data mining. Algoritma yang akan dianalisis adalah *Naïve Bayes* dan *C4.5*. [2, 3]

2. Research Method

Penelitian ini menggunakan metode penelitian eksperimen. Penelitian eksperimen melibatkan penyelidikan perlakuan pada atribut parameter atau variabel tergantung dari penelitiannya dan menggunakan tes yang dikendalikan oleh peneliti. Adapun tahapan yang dilakukan sebagai berikut: [4].

1. Pengumpulan data (*Data Gathering*)

Pada tahap ini ditentukan data yang akan diproses yaitu dataset MNC Play, kemudian mengintegrasikan semua data kedalam dataset, termasuk membagi variabel yang diperlukan dalam proses. Tahap ini dilakukan sebagai langkah awal dari suatu penelitian.

Untuk memperoleh data yang benar-benar akurat, maka penentuan jenis dan sumber data sangatlah penting. Sumber data pada penelitian ini adalah dataset MNC Play Kota Semarang yang terdiri dari 28.898 record dan 45 atribut, dataset tersebut diambil dari bulan Desember 2015 sampai dengan bulan Oktober 2018. Dataset ini memiliki klasifikasi loyalitas pelanggan tersebut sesuai dengan standar perusahaan MNC Play yang memiliki 2 label/class:

- 1) *Permanent Activation* (pelanggan yang loyal)
- 2) *Termination* (pelanggan yang tidak loyal)

2. Pengolahan awal data (*Data Pre-processing*)

Tahap pengolahan awal data dilakukan untuk mempersiapkan data yang benar-benar valid sebelum diproses pada tahap berikutnya:

- 1) *Data integrasi*
Data pelanggan yang terdiri dari permanent dan termination yang diperoleh diintegrasikan sebagai satu kesatuan data. Dalam hal ini dataset MNC Play kota Semarang dapat dikatakan data yang tersimpan didalamnya berisi data yang konsisten dan tidak bersifat noise.
- 2) *Seleksi fitur (atribut)*
Seleksi fitur digunakan sebagai input untuk proses klasifikasi. Seleksi fitur dilakukan dengan mengambil sebagian variabel pada seluruh atribut yang ada. Sedangkan pada dataset MNC Play kota Semarang dengan populasi data yang memiliki banyak record maka tidak diperlukan sampling data karena pada dataset tersebut dapat digunakan keseluruhan tanpa mengambil sample.
- 3) *Data cleansing*
Pada tahap ini, dilakukan penghapusan data yang sama atau bahkan kata yang tidak lengkap, sedangkan pada dataset MNC Play data merupakan bagian yang tidak terpisahkan dari atribut satu dengan yang lain, dengan kata lain tidak ada data yang dihapus pada dataset.
- 4) *Transformasi data*

Transformasi data dilakukan dimana data yang ada bukan merupakan data yang bertipe sama dan merubah tipe data dari string ke bentuk numerik. Pada tahap transformasi data ini tidak diperlukan scaling atau lebih tepatnya mengkompres dataset agar menjadi data yang sederhana. Transformasi data pada dataset MNC Play merubah beberapa tipe atribut data agar dikenali oleh RapidMiner, seperti tabel 1 di bawah ini:

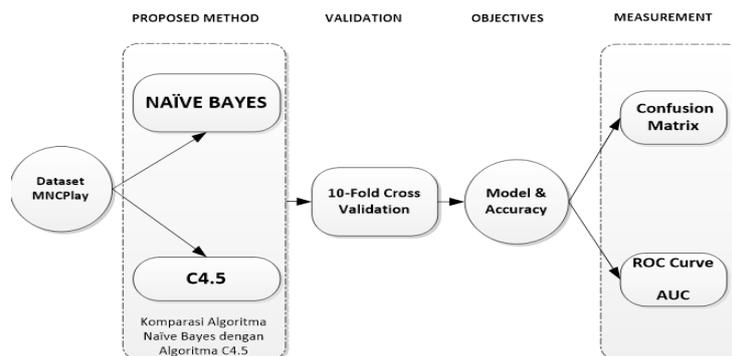
Tabel 1: Tipe Atribut Data

CUSTOMER_STATUS	Binominal	Label/Class
CUSTOMER_CODE	Polynominal	Attribute
SUSPEND_DATE	Date_time	Attribute
TERMINATION_DATE	Date_time	Attribute
NO_ACCOUNT	Integer	Attribute
NAMA_PELANGGAN	Polynominal	Attribute
CUSTOMER_CATEGORY	Polynominal	Attribute
SALES_CODE	Polynominal	Attribute
Sales	Polynominal	Attribute
TIsales	Polynominal	Attribute
sm	Polynominal	Attribute
FORMSIGNED	Date_time	Attribute
LINKACTIVE	Date_time	Attribute

3. Metode yang diusulkan (*Proposed Method*)

Tahap ini akan membahas metode yang akan digunakan untuk penelitian ini. Berikut ini tahap yang akan dilakukan dalam penelitian. Proses pertama adalah melakukan klasifikasi menggunakan algoritma Naïve Bayes, hasil proses klasifikasi di evaluasi dengan menggunakan *Confusion Matrix* dan *ROC Curve* untuk mengukur performa atau tingkat akurasi.

Pada proses berikutnya melakukan klasifikasi menggunakan algoritma C4.5. Hasil performa akurasi kedua algoritma tersebut dibandingkan menggunakan *Confusion Matrix* dan *ROC Curve*. Pada tahap ini akan dijelaskan metode yang akan digunakan untuk penelitian. Berikut ini adalah tahap yang akan dilakukan:



Gambar 1: Metode yang diusulkan

Pada tahap pre-processing data dilakukan peng-integrasian data, seleksi fitur, proses cleansing dan proses transformasi data. Proses selanjutnya adalah melakukan komparasi klasifikasi data mining berdasarkan data yang disediakan untuk mengetahui tingkat

akurasi yang paling baik. Setelah dilakukan proses komparasi, hasil proses klasifikasi dievaluasi dengan menggunakan *Confusion Matrix* dan *ROC Curve* untuk mengukur performa atau tingkat akurasi masing-masing klasifikasi data mining.

4. Eksperimen dan pengajuan model (*Model Testing and Experimen*)

Pada tahapan ini menjelaskan tentang teknik pengujian yang digunakan. Tahap modeling untuk mengklasifikasikan tingkat loyalitas pelanggan MNC Play dengan menggunakan komparasi dua algoritma *Naïve Bayes* dan *C4.5*. Dalam melakukan penelitian ini diperlukan eksperimen dan proses pengujian model yang diusulkan. Proses eksperimen dan pengujian model menggunakan bagian dari dataset yang ada. Semua dataset kemudian diuji dengan metode yang diusulkan pada tools *RapidMiner*. Pengujian model berdasarkan perhitungan metode *X-Validation* dan hasil pengujian model berupa nilai akurasi dan nilai *AUC*.

5. Evaluasi dan validasi hasil (*Result Evaluation*)

Pada tahap ini akan membahas tentang hasil evaluasi dari eksperimen yang telah dilakukan. Model yang terbentuk akan diuji dengan menggunakan *Confusion Matrix* untuk mengetahui tingkat akurasi. *Confusion Matrix* akan menggambarkan hasil akurasi mulai dari prediksi positif yang benar, prediksi positif yang salah, prediksi negatif yang benar dan prediksi negatif yang salah. Akurasi akan dihitung dari seluruh prediksi yang benar (baik prediksi positif dan negatif). Semakin tinggi nilai akurasi, semakin baik pula model yang dihasilkan.

Pengujian juga diukur dengan menggunakan *AUC*, semakin tinggi nilai *AUC*, maka semakin baik pula model klasifikasi yang terbentuk. Dari hasil yang diperoleh kemudian dibandingkan untuk mengetahui ada perubahan atau tidak. Diharapkan dalam penelitian ini klasifikasi antara kedua algoritma *Naïve Bayes* dan *C4.5* dapat bekerja dengan baik.

3. Results and Analysis

Pengujian keakuratan loyalitas pelanggan dengan menggunakan dua algoritma klasifikasi data mining yaitu algoritma *Naïve Bayes* dan algoritma *C4.5*.

Hasil pengujian menggunakan dataset MNC Play dengan dua label yaitu *permament activation* dan *termination* adalah sebagai berikut:

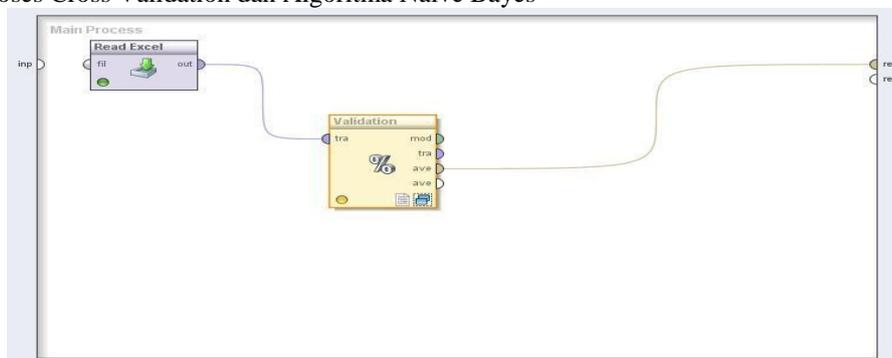
3.1. Algoritma *Naïve Bayes*

Algoritma *Naïve Bayes* merupakan pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class [5, 6].

Naïve Bayes adalah metode yang baik karena mudah dibuat, tidak membutuhkan skema estimasi parameter perulangan yang rumit.

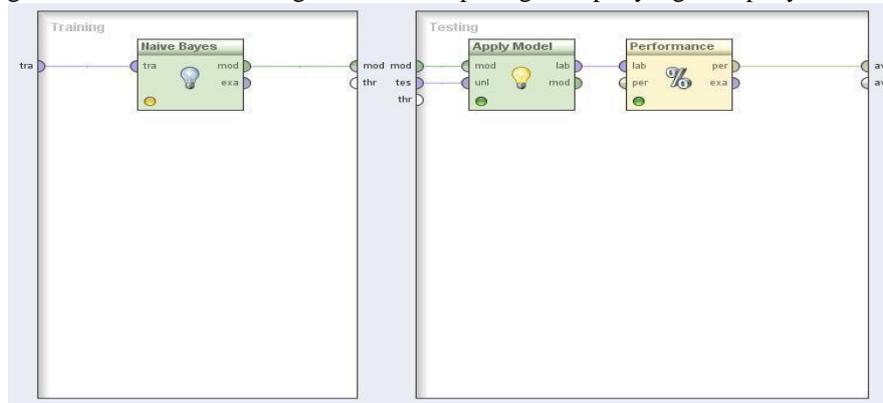
Hasil pengujian menggunakan dataset MNC Play dengan dua label yaitu *permament activation* dan *termination*. Berikut ini adalah hasil perhitungan mencari nilai akurasi dengan menggunakan *RapidMiner*, yaitu sebagai berikut:

1. Proses Cross Validation dan Algoritma *Naïve Bayes*



Gambar 2: Proses X-validation *Naïve Bayes*

Proses cross validation (X-Validation), validasi silang (Cross Validation) adalah nama operator yang melakukan validasi silang dan contoh seperangkat input yang mempunyai tabel.



Gambar 3: Proses Algoritma Naïve Bayes

Setiap performance kinerja yang dikirimkan pada port di sebelah kanan akan dirata-rata bersama-sama dan dikirim ke port output dari operator Cross Validation. Untuk memperhitungkan efek preprocessing seperti noemalisasi atau pemilihan fitur yang dapat memiliki dampak signifikan.

Hasil pengujian akurasi mengikuti kaidah aturan dari confusion matrix, data yang melalui tahap pengolahan data. Untuk klasifikasi algoritma Naïve Bayes dari dataset MNC Play dapat dilihat di tabel validasi algoritma Naïve Bayes sebagai berikut:

Tabel 2: Confusion Matrix Algoritma Klasifikasi Naïve Bayes

CLASSIFICATION	PREDICTED CLASS	
	Class = YES	Class = No
Class = YES	22337	6526
Class = No	0	0

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \\
 &= \frac{22337 + 0}{22337 + 6526 + 0 + 0} \\
 &= 0,77389 \times 100\% \\
 &= 77,39\%
 \end{aligned}$$

Hasil tersebut sesuai dengan hasil dari RapidMiner seperti dapat dilihat pada Gambar 4.

2. Hasil Akurasi Algoritma Naïve Bayes

Criterion Selector

accuracy
precision
recall
AUC (optimistic)
AUC
AUC (pessimistic)

Multiclass Classification Performance Annotations

Table View Plot View

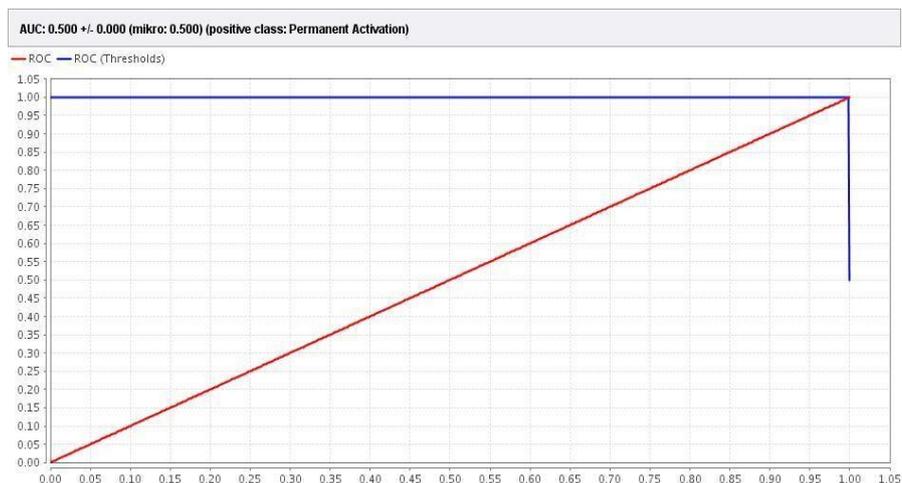
accuracy: 77.39% +/- 0.01% (mikro: 77.39%)

	true Termination	true Permanent Activation	class precision
pred. Termination	22337	6526	77.39%
pred. Permanent Activation	0	0	0.00%
class recall	100.00%	0.00%	

Gambar 4: Hasil Akurasi Algoritma *Naïve Bayes*

Evaluasi dan validasi pada pengujian ini mengikuti aturan kurva ROC yang disebut AUC (Area Under the ROC Curve) seperti yang terlihat pada gambar 5.

3. Hasil AUC (*Area Under the ROC Curve*)



Gambar 5: ROC Curve Algoritma *Naïve Bayes*

Kurve Receiver Operating Characteristic (ROC) digunakan untuk mengekspresikan data Confusion Matrix. Untuk membandingkan nilai kerja kurva biru dan kurva merah dalam bentuk angka maka dapat dilakukan dengan membandingkan panjang garis di bawah kurva atau Area Under Curve (AUC), dari gambar dapat disimpulkan panjang area kurve berwarna biru lebih panjang dibandingkan panjang area kurve berwarna merah. Pada gambar 5 threshold berhenti pada titik 0,50 yang menandakan klasifikasi tersebut gagal karena garis biru menurun. Hal ini menunjukkan bahwa model algoritma *Naïve Bayes* dikategorikan gagal.

3.2. Algoritma C4.5.

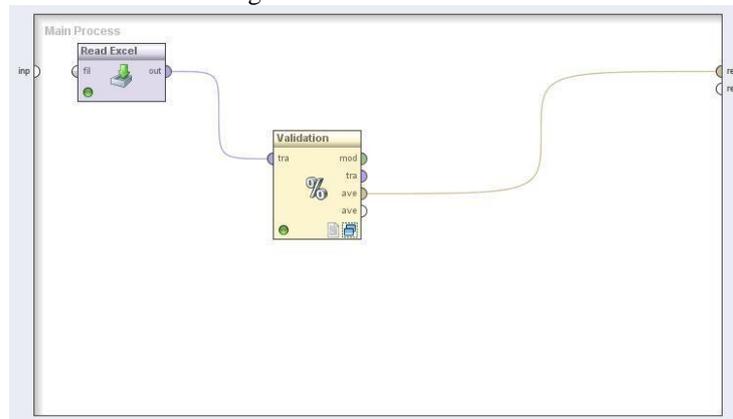
Diantara beberapa algoritma klasifikasi, algoritma C4.5 merupakan algoritma terbaik dari 10 algoritma data mining yang biasa digunakan dalam penelitian. Algoritma C4.5 mengubah fakta yang sangat besar menjadi pohon keputusan yang mpresenasikan aturan.

Pengujian menggunakan dataset MNC Play dengan dua label yaitu permanent activation

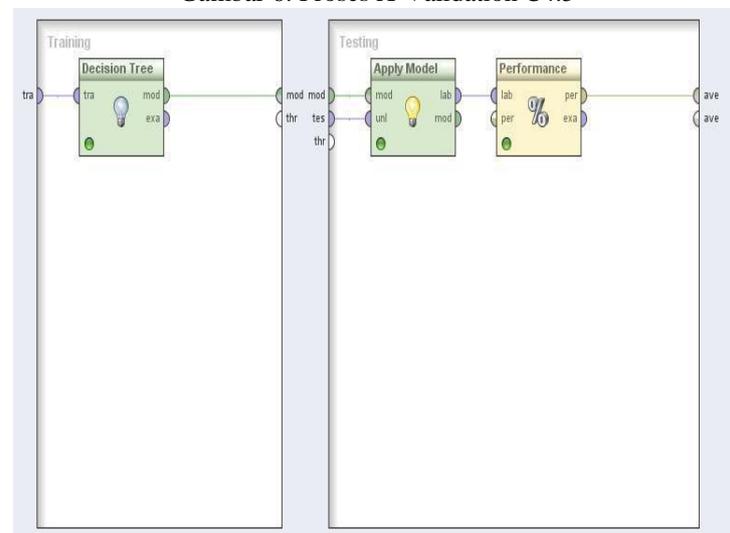
dan termination. [7]. [8].

Berikut ini adalah hasil perhitungan mencari nilai akurasi dengan menggunakan RapidMiner, yaitu gambar yang menunjukkan proses algoritma dari algoritma C4.5, sebagai berikut:

1. Proses Cross Validation dan Algoritma C4.5



Gambar 6: Proses X-Validation C4.5



Gambar 7: Proses Algoritma C4.5

Dalam validation terdapat dua kolom, training dan testing. Didalam kolom training terdapat algoritma klasifikasi yang diterapkan yaitu C4.5, sedangkan didalam kolom testing terdapat Apply Model untuk menjalankan algoritma /model C4.5 dan performance untuk mengukur peforma dari model C4.5.

Untuk mencari keakuratan klasifikasi pada dataset MNC Play maka pada proses validasi. Hasil klasifikasi menggunakan dataset MNC Play dengan algoritma C4.5 seperti pada gambar 6. Hasil pengujian akurasi mengikuti kaidah aturan dari confusion matrix, data yang melalui tahap pengolahan data.

Untuk mencari jumlah kasus keseluruhan maka digunakan teory entropy untuk menghitung jumlah kasus seluruhnya: [9].

$$\text{Entropy (S)} = \sum_{i=0}^n -p_i * \log_2 p_i$$

$$\text{Entropy [22372,6562} = \frac{22372}{28898} \times \log_2 \frac{22372}{28898}]$$

$$(0,774 \times -0,774) - (0,225 \times -0,225) (0,774 \times$$

$$-0,111) - (0,225 \times -0,647)$$

$$0,663 + 0,422$$

$$= 1,085$$

Dari hasil tersebut nilai 1,085 adalah hasil keseluruhan entropy dari dataset MNC Play. Untuk klasifikasi algoritma C4.5 dari dataset MNC Play dapat dilihat di tabel validasi algoritma C4.5.

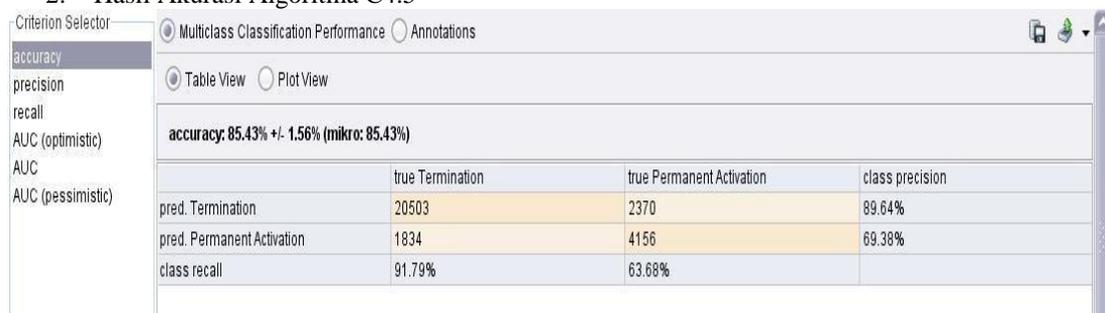
Tabel 3: Confusion Matrix Algoritma Klasifikasi C4.5

CLASSIFICATION	PREDICTED CLASS	
	Class = YES	Class = No
Class = YES	20503	2370
Class = No	1834	4156

Berikut persamaan rumus untuk menghitung tingkat akurasi pada matriks validasi algoritma C4.5.

$$\begin{aligned} \text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \\ &= \frac{20503 + 4156}{20503 + 4156 + 1834 + 2370} \\ &= \frac{24659}{28863} \times 100\% \\ &= 85,43\% \end{aligned}$$

2. Hasil Akurasi Algoritma C4.5

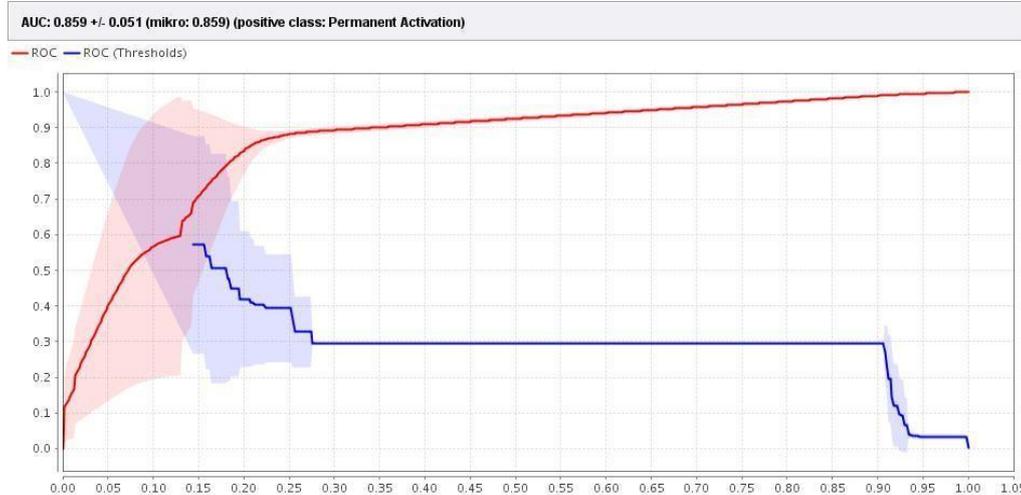


	true Termination	true Permanent Activation	class precision
pred. Termination	20503	2370	89.64%
pred. Permanent Activation	1834	4156	69.38%
class recall	91.79%	63.68%	

Gambar 8: Hasil Akurasi Algoritma C4.5

Standar deviasi memberi gambaran tentang seberapa kuat model semakin kecil standar deviasi, semakin rendah tingkat ketergantungan kinerja model pada kumpulan data uji. Evaluasi dan validasi pada penelitian ini mengikuti aturan *AUC*, dengan perhitungan nilai *AUC*:

3. Hasil *AUC* (Area Under the ROC Curve) *C4.5*



Gambar 9 : *ROC Curve C4.5*

Untuk membaca kurva tersebut pada kinerja algoritma klasifikasi adalah: *Bad*, jika kurva yang dihasilkan mendekati garis baseline atau garis yang melintang dari titik 0,0. *Good*, jika kurva mendekati titik 0,1. Pada gambar 4.9 dapat dilihat 2 kurva, yaitu kurva dengan warna biru dan kurva dengan warna merah.

Dapat disimpulkan kinerja kurva berwarna biru lebih bagus dibandingkan kinerja kurva berwarna merah. Dengan cara tersebut maka dapat dengan mudah membandingkan bagus atau tidaknya klasifikasi algoritma *C4.5*. Dari gambar dapat disimpulkan panjang area kurva berwarna biru lebih besar dibanding.

- SUSPEND_DATE > 1438084800000 DAN NO_ACCOUNT >1000023999
MAKA PERMANENT ACTIVATION
- 11) JIKA linkactive ≤ 1503230400000 DAN linkactive ≤ 1490788800000 DAN CUSTOMERCOD =SBR DAN NO_ACCOUNT ≤ 1000079439500 DAN SUSPEND_DATE > 1438084800000 DAN NO_ACCOUNT ≤ 1000023999 DAN NO_ACCOUNT > 100019895500 MAKA TERMINATION
 - 12) JIKA linkactive ≤ 1503230400000 DAN linkactive ≤ 1490788800000 DAN CUSTOMERCOD =SBR DAN NO_ACCOUNT ≤ 1000079439500 DAN SUSPEND_DATE > 1438084800000 DAN NO_ACCOUNT ≤ 1000023999 DAN NO_ACCOUNT ≤ 100019895500 MAKA PERMANENT ACTIVATION
 - 13) JIKA linkactive ≤ 1503230400000 DAN linkactive ≤ 1490788800000 DAN CUSTOMERCOD =SBR DAN NO_ACCOUNT ≤ 1000079439500 DAN SUSPEND_DATE ≤ 1438084800000 DAN OPENWO ≤ 1431345600000 MAKA PERMANENT ACTIVATION
 - 14) JIKA linkactive ≤ 1503230400000 DAN linkactive ≤ 1490788800000 DAN CUSTOMERCOD =SBR DAN NO_ACCOUNT ≤ 1000079439500 DAN SUSPEND_DATE ≤ 1438084800000 DAN OPENWO > 1431345600000 MAKA TERMINATION
 - 15) JIKA linkactive ≤ 1503230400000 DAN linkactive ≤ 1490788800000 DAN CUSTOMERCOD =SBR DAN NO_ACCOUNT > 1000079439500 MAKA TERMINATION
 - 16) JIKA linkactive ≤ 1503230400000 DAN linkactive ≤ 1490788800000 DAN CUSTOMERCOD =SPP MAKA TERMINATION
 - 17) JIKA linkactive ≤ 1503230400000 DAN linkactive ≤ 1490788800000 DAN CUSTOMERCOD =PA MAKA TERMINATION

Pada penelitian yang menguji keakuratan klasifikasi loyalitas pelanggan dengan menggunakan algoritma *Naïve Bayes* dan algoritma *C4.5* untuk mencari algoritma klasifikasi terbaik dengan dataset dari MNC Play.

Pada eksperimen tahap awal *Naïve Bayes* adalah metode yang baik karena mudah dibuat, tidak membutuhkan skema estimasi parameter perulangan yang rumit, ini berarti bisa diaplikasikan untuk dataset berukuran besar (Wu & Kumar, 2009). Mudah diintegrasikan sehingga pengguna yang tidak punya keahlian dalam bidang teknologi klasifikasi pun bisa mengerti.

Naive Bayes merupakan algoritma yang dapat meminimalkan tingkat kesalahan dibandingkan dengan semua pengklasifikasi lainnya. Setelah diuji algoritma *Naïve Bayes* memperoleh hasil akurasi sebesar 77,39% dengan dengan nilai *AUC* 0,50 dan termasuk kategori *gagal*. *AUC* 0,50 yang masuk dalam kategori gagal dengan arti pada algoritma *Naïve Bayes* kali ini tidak dapat mengelompokkan secara akurat dari dataset yang ada.

Pada eksperimen tahap kedua menggunakan algoritma *C4.5*, Teori entropi diadopsi untuk memilih pemecahan atribut yang tepat untuk algoritma *C4.5*, dengan menyatakan jumlah rata-rata informasi yang dibutuhkan untuk mengklasifikasikan sampel dengan memperoleh hasil akurasi sebesar 85,43% dengan dengan nilai *AUC* 0.859 dan termasuk kategori *Baik*.

Berdasarkan pengujian dataset MNC Play diatas menggunakan 2 algoritma yaitu *C4.5* dan *Naive Bayes*, diperoleh hasil penelitian eksperimen yang dapat disimpulkan ke dalam tabel 4.

Tabel 4 : Hasil Pembahasan

Perbandingan	ALGORITMA	
	<i>Naive Bayes</i>	<i>C4.5</i>
Akurasi	77,39 %	85,43 %
AUC	0,50	0,859

Kategori AUC	Klasifikasi gagal	Klasifikasi baik
Tree/Pohon keputusan	-	Tree terbentuk dengan baik

Berdasarkan tabel 4 diantara dua algoritma yang digunakan, algoritma C4.5 yang memiliki *performance* atas *Acuuracy* dan *AUC* yang lebih tinggi yaitu 85,43 % dan 0,859. Sehingga dapat diambil kesimpulan bahwa algoritma C4.5 memiliki *performance* yang lebih baik dibandingkan *Naive Bayes* dalam pengklasifikasian tingkat loyalitas pelanggan di MNC Play Semarang.

4. Conclusion

Berdasarkan hasil dan pembahasan yang telah dilakukan dapat diambil kesimpulan bahwa algoritma C4.5 dan algoritma *Naive bayes* untuk klasifikasi tingkat loyalitas pelanggan MNC Play kota Semarang sebagai berikut:

1. Dengan algoritma C4.5 didapatkan hasil akurasi 85,43 % dan nilai AUC 0,859 termasuk dalam kategori termasuk kategori *good classification*.
2. Dengan algoritma *Naive Bayes* hanya didapatkan nilai akurasi 77,39%, nilai AUC 0,50 termasuk dalam kategori *failure classification*.
3. Komparasi dua algoritma ini menunjukkan bahwa algoritma C4.5 menghasilkan nilai klasifikasi yang lebih akurat dibanding algoritma *Naive Bayes*.
4. Algoritma C4.5 dapat digunakan sebagai acuan untuk klasifikasi pelanggan loyal.permanant dan tidak loyal/termination pada dataset MNC Play kota Semarang.

References

- [1] Mattison,R “*The Telco Churn Management Handbook*,” B.K. Mattison, Ed Oakwood Hills, Illinois, USA: B. K. Mattison, Ed., 2005.
- [2] Santoso, B, “*Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*” (1st ed),” Yogyakarta, Indonesia, 2007.
- [3] Han, J, “*Data Mining Concept and Technique*,” (2nd Edition ed). (A. Stephan, Ed) Champaign, United States of America: Multiscience Press, 2007.
- [4] Berndtsson, M., Hasson, J. Olsson, B., & Lundell, B “*Tesis Projects A Guide for Students in Computer Science and Information System* “(2nd ed). London: Springer. 2008.
- [5] C, D. A., Baskoro, D. A., Ambarwati, L., & Wicaksana, “*Belajar Data Mining dengan RapidMiner*”, Jakarta. 2013.
- [6] Russell, S, J., & Norvig, P, “*Artificial Intelligence: Modern Approach*,” Prentice Hall. 2009.
- [7] Wu, X., & Kumar, V, “*The Top Ten Algorithms in Data Mining*,” Taylor & Francis Group, LLC. 2009.
- [8] Quinlan, J. R,”*Program for Machine Learning*,” Morgan Kaufmann. 2003.
- [9] Kotsiantis, S, Educational Data Mining: A Case Study for Predicting Dropout-Prone Students. Int. J, “*of Knowledge Engineering and Soft Data Paradigms, X*”. 2010.