

ANALISIS PERBANDINGAN ALGORITMA ID3 DAN C4.5 TERHADAP DATA PENGISIAN UANG ATM

COMPARISON ANALYSIS OF ID3 AND C4.5 ALGORITHM ON ATM MONEY CHARGING DATA

Dahri Yani Hakim Tanjung¹,Fujiati²

Fakultas Teknik dan Ilmu komputer, Universitas Potensi Utama, Medan, Indonesia

Email Korespondensi: notashapire@gmail.com, fuji.potensiutama@gmail.com

Abstrak

Penelitian ini bertujuan untuk menganalisis perbandingan dan menguji tingkat akurasi dari algoritma klasifikasi data mining. Perbandingan algoritma bergantung pada berbagai parameter dalam kumpulan data tertentu. Ada sejumlah algoritma pembelajaran dan klasifikasi yang digunakan untuk menganalisis, mempelajari pola, dan mengkategorikan data yang tersedia. Namun masalahnya adalah menemukan algoritma terbaik dan hasil yang diinginkan dengan nilai akurasi tertinggi dalam memprediksi nilai masa depan atau peristiwa dari dataset. Dimana model klasifikasi yang digunakan adalah algoritma ID3 dan C4.5 dengan menggunakan model cross validasi untuk proses pengujian dengan bantuan software Rapidminer yang dapat memvisualisasikan data dengan cepat dan sesuai. Dataset yang digunakan adalah data pengisian ATM dan pengujian yang telah dilakukan dengan model validasi cross validation membuktikan bahwa algoritma C4.5 memiliki performa terbaik dengan nilai akurasi sebesar 96,17%, nilai Precision sebesar 100% dan nilai Recall cenderung sama yaitu sebesar 93,33%. Dari rata-rata nilai tersebut dapat dilihat bahwa algoritma C4.5 terbukti sebagai algoritma pengklasifikasi yang berpotensi efektif dan efisien. Sehingga untuk membentuk pohon keputusan diambil dari algoritma yang memiliki nilai akurasi terbaik yaitu algoritma C4.5 dimana rule dari pohon keputusan yang dihasilkan dapat dijadikan acuan dalam prediksi pengisian ATM kedepannya.

Kata Kunci: Data Mining, Klasifikasi, Algoritma ID3, C4.5, Prediksi, Pengisian Uang

Abstract

This study aims to analyze the comparison and test the accuracy of the data mining classification algorithm. Comparison algorithms depend on various parameters in a given data set. There are a number of learning and classification algorithms used to analyze, study patterns, and categorize the available data. But the problem is finding the best algorithm and the desired result with the highest accuracy value in predicting future values or events from the dataset. Where the classification model used is the ID3 and C4.5 algorithm by using a cross-validation model for the testing process with the help of Rapidminer software which can visualize data quickly and appropriately. The dataset used is ATM filling data and tests that have been carried out with a cross-validation validation model prove that the C4.5 algorithm has the best performance with an accuracy value of 96.17%, a Precision value of 100% and the Recall value tends to be the same, which is 93.33%. From the average value, it can be seen that the C4.5 algorithm is proven to be a classifier algorithm that has the potential to be effective and efficient. So to form a decision tree, the algorithm that has the best accuracy value is the C4.5 algorithm where the rule from the resulting decision tree can be used as a reference in predicting ATM filling in the future.

Keywords: Data Mining, Clasification, ID3 Algorithm, C4.5 Algorithm, Prediction, Money Replenishment

1. PENDAHULUAN

ATM merupakan salah satu fasilitas perbankan yang ditawarkan sebagai salah satu saluran pengiriman alternatif dalam bertransaksi kepada nasabah yang bekerja secara otomatis menggantikan tugas teller [1]. Penggunaan alat pembayaran non tunai tiap tahunnya terus terjadi peningkatan hal ini menunjukkan bahwa masyarakat sudah bisa menerima pembayaran secara non tunai[2]. Pengontrolan pengisian uang mesin ATM yang di selesaikan dengan teknologi informasi dapat memberikan manfaat bagi perusahaan, Oleh karena itu kebutuhan akan suatu teknologi informasi yang dapat menunjang proses pengontrolan pengisian uang pada mesin ATM sangat diperlukan guna membantu dalam pengambilan keputusan atau sebagai bahan masukan kembali bagi perkembangan perusahaan tersebut[3]

Klasifikasi merupakan salah satu teknik data mining yang digunakan untuk memetakan data ke dalam kelompok atau kelas yang telah ditentukan dengan perkembangan yang cukup signifikan[4][5]. Hal ini ditemukan banyaknya penelitian menggunakan teknik klasifikasi data mining yang diimplementasikan dalam bidang ritel, perbankan, militer, dan kesehatan, dan untuk mencapai tujuannya, para peneliti sedang mengembangkan algoritma yang berbeda menggunakan keahlian dari berbagai bidang studi[6][7]

Cross Validasi digunakan untuk menguji keefektifan model pembelajaran mesin dengan teknik validasi dengan membagi seluruh data menjadi dua set pelatihan dan set tes[8]. Model k-Fold Cross Validation digunakan untuk mengetahui nilai k Optimal yang melibatkan pemisahan secara acak kumpulan sampel menjadi serangkaian fold (grup) berukuran sama, di mana k menunjukkan jumlah partisi atau fold dari dataset yang dibagi menjadi data latih dan data uji, Misalnya, jika nilai-k sepuluh digunakan, maka dataset dibagi menjadi sepuluh partisi. Dalam hal ini, sembilan dari partisi digunakan untuk data latih, sedangkan satu partisi sisanya digunakan untuk data uji [9][10][11].

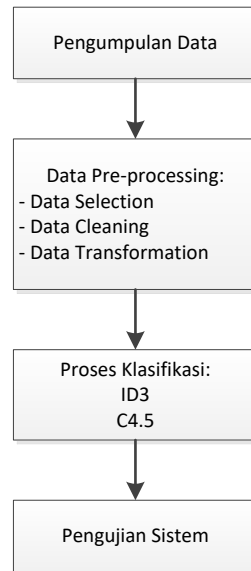
Pada penelitian ini penulis membandingkan dan menganalisis algoritma klasifikasi dengan Rapidminer. Rapidminer merupakan software open-source untuk mendesain alur kerja visual untuk data preparation, deep learning, machine learning, text mining, yang digunakan untuk analisis prediktif dan simulasi[12][13][14]

Adapun metode pemecahan masalah yang digunakan dalam penelitian ini adalah Data Mining. Data Mining yang digunakan adalah metode klasifikasi dengan membandingkan algoritma ID3 dan C4.5. Diharapkan hasil pengujian model algoritma klasifikasi pada dataset yang ada mampu menghasilkan performa yang optimal dan efisien.

2. METODOLOGI PENELITIAN

2.1 Teknik Pengumpulan Data

Pada bab ini menjelaskan tentang metodologi penelitian mengenai perbandingan metode klasifikasi. Metodologi penelitian ini melalui beberapa tahapan yang dapat dilihat pada gambar 1 berikut. :



Gambar 1. Kerangka Kerja Penelitian

Berdasarkan kerangka kerja penelitian pada gambar 1 di atas maka dapat diuraikan sebagai berikut :

1. Pengumpulan Data
Proses pengumpulan data dilakukan dengan mengambil data sampel pengisian uang ATM
2. Pengolahan Data
Pada tahap pengolahan data mining, ada beberapa proses yang dibutuhkan dalam pengolahan data antara lain Data Selection, Data Cleaning, Transformation, Pada Proses Data Selection, dilakukan penyeleksian data yang sesuai dan dibutuhkan dalam melakukan analisa. Selanjutnya pada proses Data Cleaning, dilakukan pembersihan data terhadap data yang tidak relevan, noise dan data yang tidak konsisten serta menghilangkan data apabila terdapat duplikat atau sama. Selanjutnya pada proses Transformation, dilakukan perubahan data sesuai dengan kebutuhan.
3. Proses Klasifikasi
Proses klasifikasi merupakan bagian dari Data Mining, dimana pada proses ini dilakukan perbandingan algoritma klasifikasi yaitu algoritma ID3 dan algoritma C4.5.
4. Pengujian Sistem
Tahap akhir pada sistem ini adalah tahap uji coba system dengan melakukan validasi terhadap proses klasifikasi, Proses validasi yang dilakukan menggunakan metode k-fold cross validation.

2.2 Algoritma ID3

ID3 adalah algoritma pembelajaran pohon keputusan sederhana yang dikembangkan oleh Ross Quinlan (1983). Pada tahun 1975, ID3 pertama kali dipresentasikan dalam sebuah buku Quinlan. Algoritma Concept Learning System (CLS) adalah dasar dari algoritma ID3. Dengan menambahkan pemilihan fitur heuristik ID3 untuk meningkatkan CLS. Atribut dari instance pelatihan adalah dicari melalui ID3 dan atribut yang paling baik memisahkan contoh yang diberikan diekstraksi olehnya. ID3 berhenti jika atribut dengan sempurna mengklasifikasikan set pelatihan; jika tidak, secara rekursif beroperasi pada n , di mana n adalah jumlah kemungkinan nilai atribut dari himpunan bagian yang dipartisi untuk mendapatkan atribut "terbaik" mereka. Selain itu, algoritma ID3 memilih atribut sebagai informasi entropy yang akan dihitung berdasarkan probabilitas, dan perolehan informasi untuk mengembangkan pohon keputusan [15][16]. Untuk membangun pohon keputusan, hitung information gain untuk setiap atribut dan pilih atribut dengan information gain tertinggi untuk dijadikan sebagai akar. Beri label pada atribut sebagai akar. Information Gain merupakan suatu ukuran korelasi pada model parametrik yang menggambarkan ketergantungan antara dua peubah acak X dan Y . Terdapat beberapa tahapan dalam membentuk pohon keputusan dengan algoritma ID3 [17] yaitu:

1. Pilih atribut sebagai akar, dengan menghitung nilai entropy. Dimana rumus untuk menghitung nilai entropy dapat dilihat pada persamaan 1 berikut:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (1)$$

2. Buat cabang untuk setiap atribut yang memiliki nilai gain tertinggi, dimana untuk menghitung nilai gain ratio dapat dilihat pada persamaan 2 berikut:

$$Gain(S, A) = Entropy(S) - \sum_{v \in \text{nilai}(A)} \left| \frac{S_v}{S} \right| Entropy(S_v) \quad (2)$$

3. Ulangi langkah 1 dan 2 untuk setiap cabang hingga semua kasus memiliki daun atau keputusan
4. Buat rule berdasarkan pohon keputusan

2.3 Algoritma C4.5

Algoritma ini diusulkan pada tahun 1993, oleh Ross Quinlan yang merupakan bagian algoritma supervised learning yang digunakan untuk klasifikasi yang digambarkan seperti diagram alur yang masing-masing simpul internal menunjukkan tes pada atribut, setiap cabang mewakili hasil tes, dan setiap daun atau terminal node memegang label kelas dan menghasilkan pohon keputusan melalui pembelajaran dari satu set pelatihan, di mana setiap contoh adalah terstruktur dalam untuk memperbaiki atribut-nilai.[18][19][16]. Algoritma C4.5 perkembangan dari algoritma ID3 yang menggunakan Information Gain kemudian diperbaharui menjadi Gain Ratio yang bertujuan untuk mengurangi bias atribut yang memiliki banyak cabang biasanya [5]. Terdapat beberapa tahapan dalam membentuk pohon keputusan dengan algoritma C4.5 [17] yaitu:

1. Hitung nilai Entropy dan Gain

$$Entropy(S) = - \sum_{i=1}^n p_i + \log_2 p_i \quad (3)$$

Hasil dari nilai Entropy kemudian digunakan untuk menghitung nilai Gain

$$Gain(S, A) = Entropy(S_i) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (4)$$

5. Hitung nilai Split Info

$$Split\ Information(S, A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} + \log_2 \frac{|S_i|}{|S|} \quad (5)$$

6. Hasil dari nilai Gain dan Split Info digunakan untuk mencari nilai Gain Ratio

$$Gain\ Ratio(S, A) = \frac{Gain\ Information(S, A)}{Split\ Information(S, A)} \quad (6)$$

7. Nilai Gain Ratio tertinggi dijadikan sebagai akar
8. Ulangi langkah 1 dan 2 untuk setiap cabang hingga semua kasus memiliki daun atau keputusan
9. Buat rule berdasarkan pohon keputusan

2.4 Cross Validation

Cross Validation merupakan suatu metode tambahan dari teknik data mining pengembangan model split validation yang bertujuan untuk memperoleh hasil akurasi yang maksimal dimana validasinya mengukur training error dengan menguji dengan test data atau data uji.[20] Dari hasil percobaan dengan pengujian sebanyak n kali dengan k-fold cross validation akan dicatat nilai evaluasi performa dari model tersebut dengan menggunakan confusion matrix. Pengujian pada algoritma ID3 dan C4.5 dilakukan dengan confusion matrix untuk mengetahui tingkat Sensitivity (recall),

PPV (positive predictive value) atau precision, dan akurasi dari metode klasifikasi yang dibuat [17]:

Untuk menghitung recall dengan rumus :

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

Recall bertujuan untuk mengukur akurasi/kemampuan sebuah model prediksi untuk memilih instance dari kelas tertentu dari kumpulan data.

Untuk perhitungan precision menggunakan rumus :

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

Precision bertujuan untuk mengukur proporsi jumlah kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. Presisi bertujuan untuk mengukur akurasi yang berasal dari kelas tertentu yang telah diprediksi. Mencoba untuk menentukan proporsi positif identifikasi yang sebenarnya benar.

Akurasi dihitung dengan rumus :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

Akurasi bertujuan untuk menjumlah prediksi penerima hibah pemasangan air minum yang benar.

Dimana :

True Positive (TP) : proporsi positif yang terdapat dalam data set yang diklasifikasikan positif.

False Negative (FN) : proporsi negatif yang terdapat dalam data set yang diklasifikasikan negatif.

False Positive (FP) : proporsi negatif yang terdapat dalam data set yang diklasifikasikan positif.

True Negatif (TN) : proporsi positif yang terdapat dalam data set yang diklasifikasikan negative

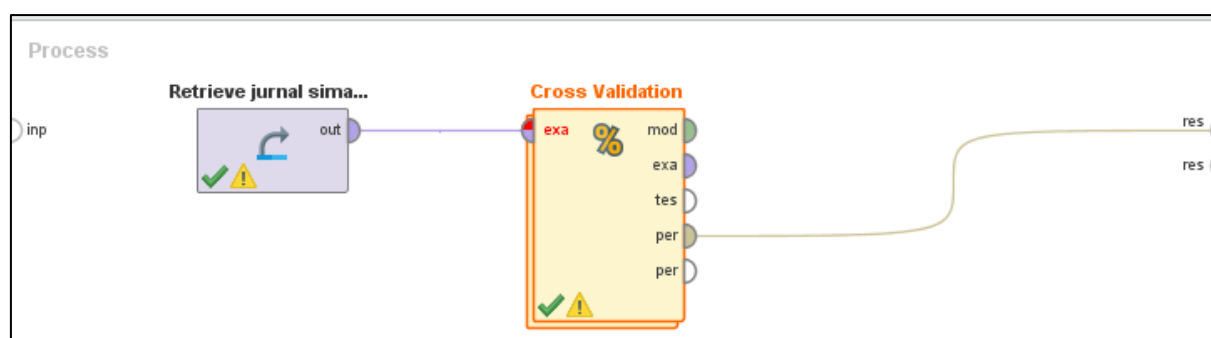
3. HASIL DAN PEMBAHASAN

Didalam bab ini penulis akan membahas mengenai data diuji dan dianalisis dengan dua algoritma klasifikasi yaitu algoritma ID3 dan algoritma C4.5 (menggunakan cross validasi). Data yang digunakan adalah data pengisian ATM yang diambil untuk analisis penelitian ini. Dimana data yang digunakan dapat dilihat di tabel 1 di bawah ini

Tabel 1. Dataset

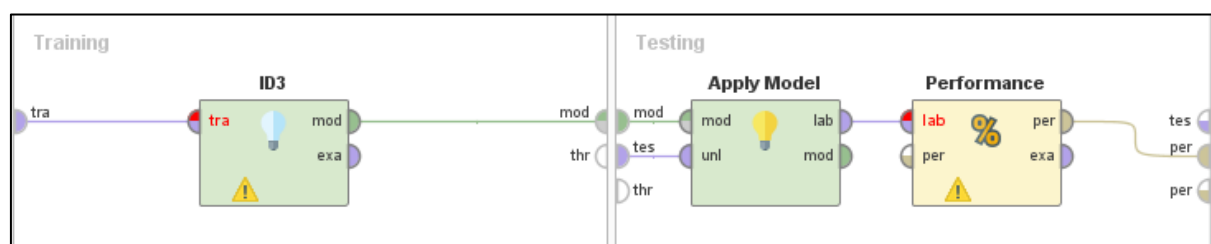
ID ATM	CIB %	Lokasi	Status ATM	Restock	CashOut	EndCash	Status Pengisian
SIAMDNO8BF	Besar	Jauh	Perbaikan	Sedang	Banyak	Sedang	Tidak Isi
SIAMDNO9XN	Besar	Sedang	Baik	Sedang	Sedikit	Banyak	Tidak Isi
SIAMDNI12L	Kecil	Sedang	Baik	Sedang	Banyak	Sedikit	Isi
SIAMDNA002	Sedang	Sedang	Baik	Banyak	Banyak	Sedang	Isi
SIAMDNA004	Besar	Sedang	Baik	Sedang	Banyak	Sedang	Tidak Isi
SIAMDNA020	Besar	Dekat	Baik	Sedang	Sedikit	Banyak	Tidak Isi
SIAMDNA027	Besar	Dekat	Perbaikan	Sedang	Banyak	Sedang	Tidak Isi
SIAMDNA039	Kecil	Jauh	Baik	Sedang	Banyak	Sedikit	Isi
SIAMDNA043	Sedang	Jauh	Baik	Sedang	Banyak	Sedikit	Isi
SIAMDNA046	Besar	Dekat	Perbaikan	Sedang	Banyak	Sedikit	Tidak Isi
SIAMDNA047	Kecil	Dekat	Baik	Sedang	Banyak	Sedikit	Isi
SIAMDNA048	Besar	Sedang	Baik	Sedang	Sedikit	Banyak	Tidak Isi
SIAMDNA054	Sedang	Jauh	Baik	Sedang	Banyak	Sedikit	Isi
SIAMDNA055	Sedang	Jauh	Baik	Sedang	Banyak	Sedikit	Isi
SIAMDNA061	Kecil	Sedang	Baik	Sedang	Sedang	Sedang	Isi
SIAMDNA096	Sedang	Dekat	Baik	Sedang	Banyak	Sedikit	Tidak Isi
SIAMDNA099	Kecil	Dekat	Baik	Sedang	Banyak	Sedikit	Isi
SIAMDNR012	Sedang	Dekat	Perbaikan	Banyak	Banyak	Sedang	Isi
SIATTIA020	Besar	Sedang	Baik	Banyak	Banyak	Banyak	Tidak Isi
SIATUSU01NF	Besar	Dekat	Baik	Sedang	Sedikit	Sedang	Tidak Isi
SIATUSU01NX	Sedang	Dekat	Baik	Sedang	Sedikit	Sedang	Tidak Isi
SIATUSU04AL	Sedang	Sedang	Perbaikan	Sedang	Banyak	Sedikit	Isi

Kemudian dilakukan transformasi data untuk tahap selanjutnya, hasil dari transformasi data tersebut akan diuji menggunakan desain model algoritma ID3 dan C4.5 dengan rapidminer. Pada gambar 2 merupakan model yang dirancang di Rapidminer



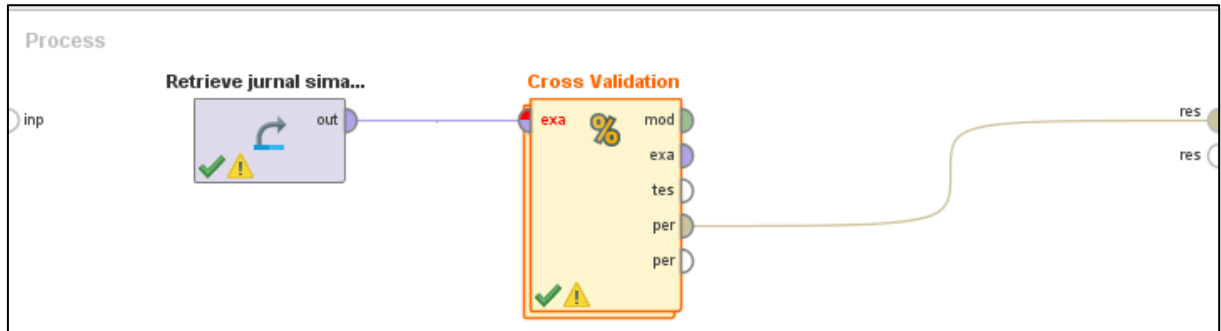
Gambar 2. Model Cross Validation Algoritma ID3

Data input pada model diatas disimpan dalam format xls dan ada juga set data pelatihan yang diberikan agar sistem dapat belajar dari data yang diberikan



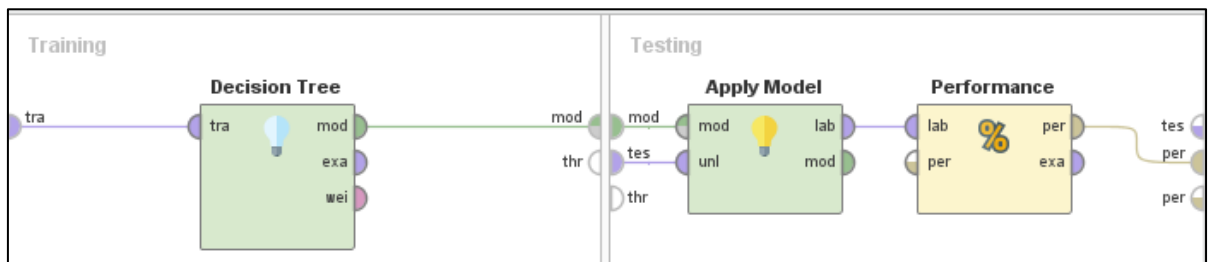
Gambar 3. Pengujian Cross Validation ID3

gambar 3 merupakan model dari algoritma ID3 yang diterapkan dalam rapidminer, dimana operator dan parameter yang digunakan akan berpengaruh terhadap akurasi dan model yang terbentuk.



Gambar 4. Model Cross Validation Algoritma C4.5

Pada gambar 4 diatas sama halnya pada Algoritma sebelumnya yaitu algoritma ID3, dimana dataset dihubungkan dengan operator cross validation yang didalamnya terdapat proses seperti gambar 5 berikut.



Gambar 5. Pengujian Cross Validation Algoritma C4.5

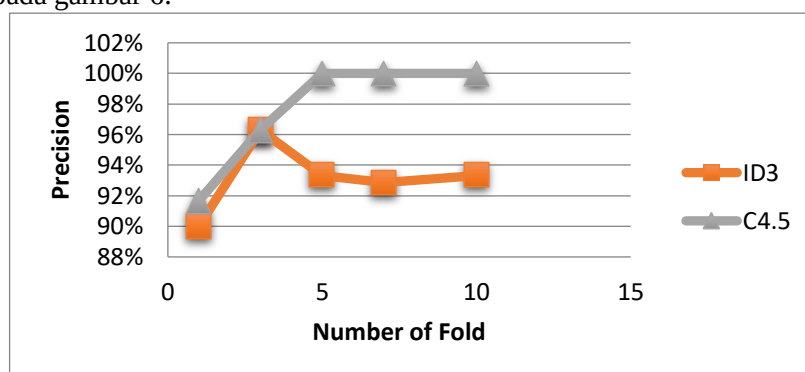
Gambar 5 merupakan model pengujian dengan cross validasi, dimana proses dalam membangun model C4.5 dengan menggunakan operator dan parameter. Model yang digunakan juga akan berpengaruh terhadap akurasi yang dihasilkan.

Selanjutnya tahap pengujian yang dilakukan menggunakan model K-Fold Cross Validation untuk menguji tingkat akurasi dari model algoritma klasifikasi yang digunakan. Cross-validasi merupakan sebuah teknik validasi model untuk menilai bagaimana hasil statistik analisis akan menggeneralisasi kumpulan data independen[21]. K-fold cross validation merupakan salah satu teknik dari cross validasi yang digunakan untuk menghilangkan bias pada data. Pelatihan dan pengujian dilakukan sebanyak k kali [22]. Nilai k-fold cross validation yang digunakan adalah 1, 3, 5, 7 dan 10-fold. Hasil validasi terhadap pengujian data dapat dilihat pada tabel 2

Tabel 2. Nilai *Presi*si perbandingan Algoritma dengan Cross Validation

Algorithm Classification	Number of Fold	Precision
ID3	1	90%
	3	96,30%
	5	93,33%
	7	92,86%
	10	93,33%
C4.5	1	91,67%
	3	96,30%
	5	100%
	7	100%
	10	100%

Perbandingan dari kedua algoritma klasifikasi dengan dengan nilai *presi*si ditunjukkan dalam grafik yang terdapat pada gambar 6.

Gambar 6. Grafik *Presi*si

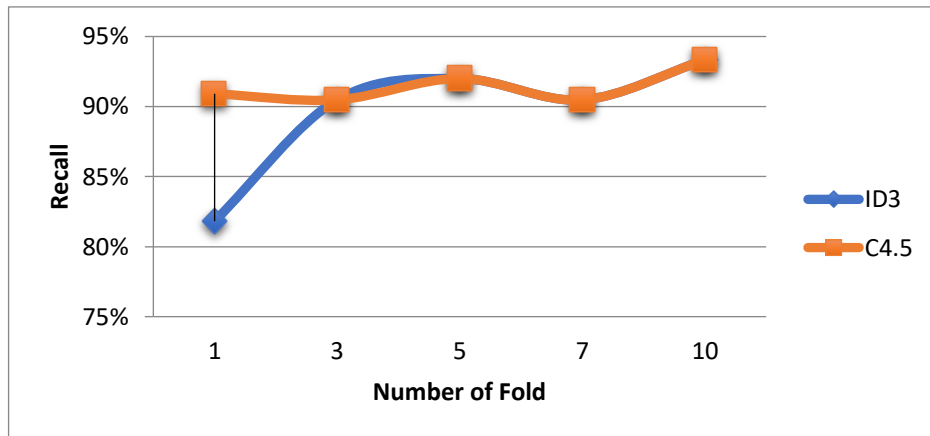
Hasil pengujian dengan model cross validation menunjukkan bahwa Algoritma C4.5 memiliki nilai precision 100%, sedangkan pada Algoritma ID3 nilai precision tertinggi sebesar 93,33% dengan nilai fold = 10. Sehingga algoritma C4.5 menunjukkan nilai precision lebih tinggi dari pada algoritma ID3. Sedangkan hasil *Recall* algoritma ID3 dan C4.5 dapat dilihat pada tabel 3 berikut

Tabel 3. Nilai *Recall* perbandingan Algoritma dengan Cross Validation

Algorithm Classification	Number of Fold	Recall
ID3	1	81,82%
	3	90,48%
	5	92%
	7	90,48%
	10	93,33%
C4.5	1	90,91%
	3	90,48%
	5	92%
	7	90,48%

10 93,33%

Perbandingan dari kedua algoritma klasifikasi dengan dengan nilai *Recall* ditunjukkan dalam grafik yang terdapat pada gambar 7.



Gambar 7. Grafik *Recall*

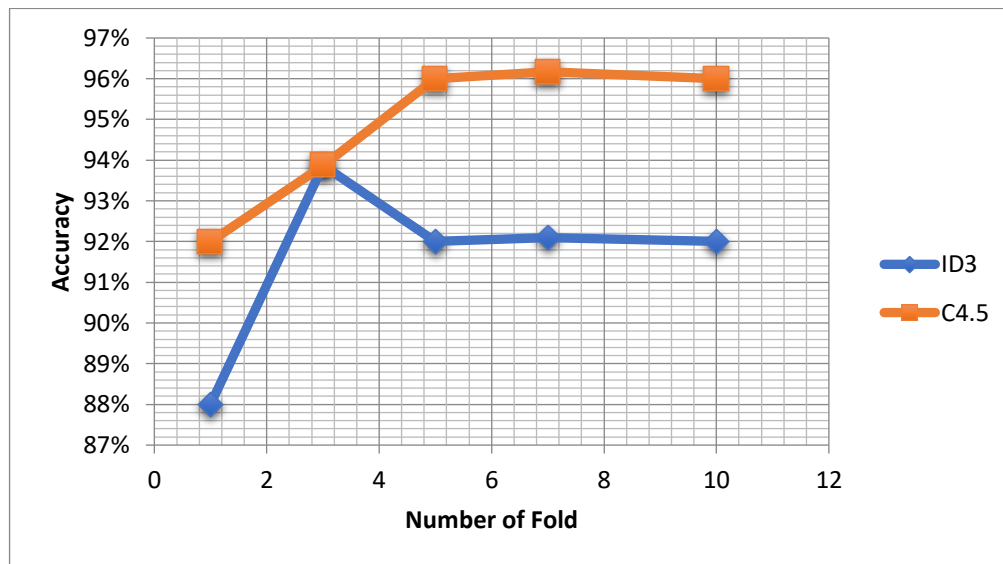
Nilai *Recall* yang dihasilkan dari pengujian dengan model cross validation menunjukkan bahwa Algoritma C4.5 dan algoritma ID3 memiliki nilai yang sama yaitu sebesar 93,33%, dan cenderung memiliki nilai recall yang sama. Yang membedakan pada nilai fold = 1 dimana algoritma C4.5 memiliki nilai recall lebih tinggi dibandingkan dengan algoritma ID3 yaitu sebesar 90,91% sedangkan algoritma ID3 dengan nilai 81,82%. meskipun pada jumlah data yang lain nilai recall yang dihasilkan sama, tetapi nilai rata rata yang dihasilkan pada algoritma C4.5 lebih tinggi.

Selanjutnya merupakan hasil akurasi dari perbandingan algoritma ID3 dan C4.5 yang dapat dilihat pada tabel 4 berikut

Tabel 4. Nilai *Akurasi* perbandingan Algoritma dengan Cross Validation

Algorithm Classification	Number of Fold	Accuracy
ID3	1	88%
	3	93,87%
	5	92%
	7	92,09%
	10	92%
C4.5	1	92%
	3	93,87%
	5	96%
	7	96,17%
	10	96%

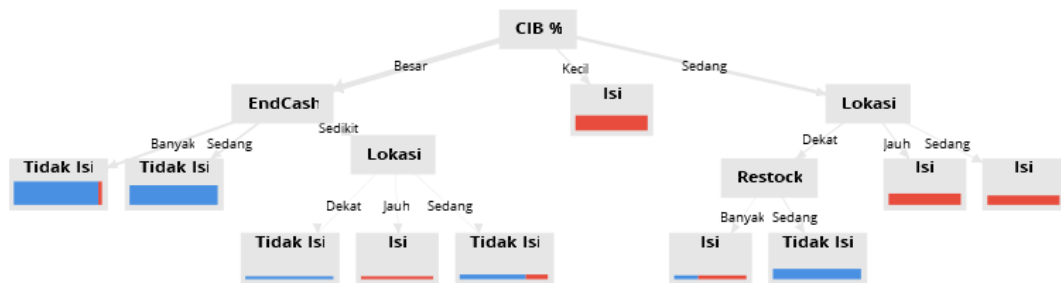
Perbandingan dari kedua algoritma klasifikasi dengan dengan nilai *Akurasi* ditunjukkan dalam grafik yang terdapat pada gambar 8



Gambar 6. Grafik Akurasi

Dari hasil pengujian validasi dengan beberapa nilai fold dapat diketahui bahwa algoritma C4.5 memiliki nilai akurasi lebih tinggi dibandingkan dengan algoritma ID3 dengan nilai akurasi 96,17% sedangkan algoritma ID3 dengan nilai 92,09% pada nilai fold = 7. Dilihat dari hasil k-fold yang lain juga menunjukkan hasil akurasi dari algoritma C4.5 lebih tinggi daripada algoritma ID3.

Oleh sebab itu, dapat disimpulkan bahwa algoritma C4.5 memiliki tingkat akurasi yang lebih tinggi berdasarkan pengujian jumlah data. Sehingga untuk membentuk pohon keputusan berdasarkan hasil akurasi tertinggi yaitu algoritma C4.5. Dimana hasil pohon keputusan dapat dijadikan acuan untuk prediksi pengisian ATM kedepannya yang dapat dilihat pada gambar 7



Rule yang dihasilkan dari pohon keputusan pada gambar 7 menggunakan tools rapid miner adalah sebagai berikut :

CIB % = Besar

| EndCash = Banyak: Tidak Isi {Tidak Isi=23, Isi=1}

| EndCash = Sedang: Tidak Isi {Tidak Isi=20, Isi=0}

| EndCash = Sedikit

```

| | Lokasi = Dekat: Tidak Isi {Tidak Isi=2, Isi=0}
| | Lokasi = Jauh: Isi {Tidak Isi=0, Isi=2}
| | Lokasi = Sedang: Tidak Isi {Tidak Isi=3, Isi=1}
CIB % = Kecil: Isi {Tidak Isi=0, Isi=15}
CIB % = Sedang
| Lokasi = Dekat
| | Restock = Banyak: Isi {Tidak Isi=1, Isi=2}
| | Restock = Sedang: Tidak Isi {Tidak Isi=10, Isi=0}
| Lokasi = Jauh: Isi {Tidak Isi=0, Isi=11}
| Lokasi = Sedang: Isi {Tidak Isi=0, Isi=9}

```

4. KESIMPULAN

Berdasarkan pembahasan yang telah dilakukan dapat disimpulkan bahwa

1. Pengujian terhadap algoritma C4.5 memiliki tingkat akurasi yang tinggi yaitu sebesar 96,17%, nilai Precision sebesar 100% dan nilai Recall cenderung sama yaitu sebesar 93,33%.
2. Berdasarkan hasil pengujian yang dilakukan dengan 5 nilai K-Fold Cross Validation yang berbeda dapat dilihat bahwa pada nilai fold = 10 hasil pengujian memiliki nilai paling tinggi untuk algoritma C4.5 dibandingkan dengan algoritma ID3. Hal ini menunjukkan bahwa semakin tinggi nilai k-fold maka tingkat akurasi juga semakin tinggi.

5. SARAN

Adapun Saran yang dapat diberikan untuk penelitian selanjutnya adalah

1. Dapat dilakukan penelitian lebih lanjut dengan menggunakan algoritma yang lain agar dapat dilihat algoritma mana yang paling akurat dalam melakukan prediksi terhadap data pengisian ATM
2. Dapat dilakukan pengujian terhadap model validasi yang lain seperti model visual X – Validation

DAFTAR PUSTAKA

- [1] R. N. Damaris, M. Sitanggang, and R. P. Simanjuntak, "Sistem Pengendalian Intern Atas Transaksi Penerimaan Dan Pengeluaran Kas Anjungan Tunai Mandiri (Atm) Pt Bank Central Asia, Tbk," *J. Ilm.*, vol. 18, no. 2, pp. 54–63, 2014.
- [2] N. N. Lintang Sari, N. Hidayati, Y. Purnamasari, H. Carolina, and W. F. Ramadhan, "Analisis Pengaruh Instrumen Pembayaran Non-Tunai Terhadap Stabilitas Sistem Keuangan Di Indonesia," *J. Din. Ekon. Pembang.*, vol. 1, no. 1, p. 47, 2018, doi: 10.14710/jdep.1.1.47-62.
- [3] R. Hafsari and S. Mawlan, "Sistem Informasi Monitoring Pengisian Uang Pada Mesin ATM oleh PT . Advantage Palembang," 2008.
- [4] K. Adhatrao, A. Gaykar, A. Dhawan, R. Jha, and V. Honrao, "Predicting Students' Performance Using ID3 and C4.5 Classification Algorithms," *Int. J. Data Min. Knowl. Manag. Process*, vol. 3, no. 5, pp. 39–52, 2013, doi: 10.5121/ijdkp.2013.3504.
- [5] M. Yusa, E. Utami, and E. Luthfi. Taufiq, "Evaluasi Performa Algoritma Klasifikasi Decision Tree ID3, C4.5, dan CART Pada Dataset Readmisi Pasien Diabetes," *Infosys (Information Syst. J.)*, vol. 4, no. 1, pp. 23–34, 2016.
- [6] I. D. Mienye, Y. Sun, and Z. Wang, "Prediction performance of improved decision tree-based algorithms: A review," *Procedia Manuf.*, vol. 35, pp. 698–703, 2019, doi: 10.1016/j.promfg.2019.06.011.
- [7] P. Ahmad, S. Qamar, and S. Qasim Afser Rizvi, "Techniques of Data Mining In Healthcare: A Review," *Int. J. Comput. Appl.*, vol. 120, no. 15, pp. 38–50, 2015, doi: 10.5120/21307-4126.
- [8] N. Darapureddy, N. Karatapu, and T. K. Battula, "Research of machine learning algorithms

- using k-fold cross validation,” *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6 Special issue, pp. 215–218, 2019, doi: 10.35940/ijeat.F1043.0886S19.
- [9] I. dan A. Mutiara, “Penerapan K-Optimal Pada Algoritma Knn Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan Ip Sampai Dengan Semester 4,” *Klik - Kumpul. J. Ilmu Komput.*, vol. 2, no. 2, pp. 159–173, 2015.
- [10] C. A. Ramezan, T. A. Warner, and A. E. Maxwell, “Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification,” *Remote Sens.*, vol. 11, no. 2, 2019, doi: 10.3390/rs11020185.
- [11] I. A. M. SUPARTINI, I. K. G. SUKARSA, and I. G. A. M. SRINADI, “Analisis Diskriminan Pada Klasifikasi Desa Di Kabupaten Tabanan Menggunakan Metode K-Fold Cross Validation,” *E-Jurnal Mat.*, vol. 6, no. 2, p. 106, 2017, doi: 10.24843/mtk.2017.v06.i02.p154.
- [12] B. Devipriya and Y. Kalpana, “Evaluation of sentiment data using classifier model in rapid miner tool,” *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 2966–2972, 2019, doi: 10.35940/ijeat.A1323.109119.
- [13] T. k and M. Wadhawa, “Analysis and Comparison Study of Data Mining Algorithms Using Rapid Miner,” *Int. J. Comput. Sci. Eng. Appl.*, vol. 6, no. 1, pp. 9–21, 2016, doi: 10.5121/ijcsea.2016.6102.
- [14] V. Pynam, R. R. Spanadna, and K. Srikanth, “An extensive study of data analysis tools,” *SSRG Int. J. Comput. Sci. Eng. (SSRG – IJCSE)*, vol. 5, no. 9, pp. 4–11, 2018.
- [15] A. K. Sharma and S. Sahni, “A comparative study of classification algorithms for spam email data analysis,” *Int. J. Comput. Sci. Eng.*, vol. 3, no. 5, pp. 1890–1895, 2011, [Online]. Available: <http://www1.ics.uci.edu/~mllearn/MLRepository.html>.
- [16] J. Dattetrya, “Improved C4 . 5 Decision Tree Classifier Algorithm,” no. 10, pp. 18–24, 2017.
- [17] D. melina Agustina and Wijanarto, “Analisis Perbandingan Algoritma ID3 Dan C4 . 5 Untuk Klasifikasi Penerima Hibah Pemasangan Air Minum pada PDAM Kabupaten Kendal,” *J. Appl. Intell. Syst.*, vol. 1, no. 3, pp. 234–244, 2016.
- [18] B. Hssina, A. Merbouha, H. Ezzikouri, and M. Erritali, “A comparative study of decision tree ID3 and C4 . 5,” no. July, 2014, doi: 10.14569/SpecialIssue.2014.040203.
- [19] N. Nofriani, “Comparations of Supervised Machine Learning Techniques in Predicting the Classification of the Household’s Welfare Status,” *J. Pekommas*, vol. 4, no. 1, p. 43, 2019, doi: 10.30818/jpkm.2019.2040105.
- [20] F. Syahputra, H. Hartono, and R. Rosnelly, “Penerapan Algoritma C4.5 Dalam Memprediksi Ketersediaan Uang Pada Mesin ATM,” *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 556, 2021, doi: 10.30865/mib.v5i2.2933.
- [21] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, “Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung,” *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 81–86, 2020, doi: 10.33096/ilkom.v12i2.507.81-86.
- [22] F. Tempola, M. Muhammad, and A. Khairan, “Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, p. 577, 2018, doi: 10.25126/jtiik.201855983.