JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor 2, April 2021, Page 556-563
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i2.2933

Penerapan Algoritma C4.5 Dalam Memprediksi Ketersediaan Uang Pada Mesin ATM

Firman Syahputra*, Hartono, Rika Rosnelly

Fakultas Teknik dan Ilmu komputer, Prodi Magister Ilmu Komputer, Universitas Potensi Utama, Medan, Indonesia
Email: [1,*]24firman@gmail.com, [2]hartonoibbi@gmail.com, [3]rika@potensi-utama.ac.id
Email Penulis Korespondensi: [1]24firman@gmail.com

Abstrak−Penelitian ini bertujuan untuk memberikan evaluasi terhadap ketersediaan uang pada mesin ATM dengan menggunakan data mining. Data mining dengan algoritma C4.5 digunakan untuk meramalkan permintaan uang tunai atau total penarikan uang tunai pada ATM, Untuk menentukan kebutuhan uang ATM berdasarkan data transaksi tunai. Diharapkan dengan adanya peramalan tersebut dapat membantu bagian monitoring dalam mengambil keputusan mengenai kebutuhan uang yang harus dialokasikan ke setiap mesin ATM. Hasil penelitian ini diharapkan dapat membantu unit pengelola ATM dalam mengotimalkan dan monitoring ketersediaan uang pada mesin ATM terhadap kebutuhan uang, sehingga dapat memberikan pelayanan yang optimal kepada nasabah. Algortima C4.5 adalah algoritma yang mampu membentuk sebuah pohon keputusan, dimana pohon keputusan kemudian akan menghasilkan sebuah pengetahuan baru. Hasil dari pengujian diperoleh kecocokan terhadap data ketersediaan uang pada mesin ATM. Hasil dari penerepan metode C4.5 terhadap ketersediaan uang pada mesin ATM dilihat dari waktu tempuh ke lokasi ATM dan juga saldo yang masih tersisa di dalam mesin. Model pohon keputusan yang dihasilkan adalah menjadikan variabel Saldo sebagai akar, Kemudian waktu tempuh sebagai cabang pada Level 1 dengan variabel cepat, sedang, lama, dan Bank menjadi cabang pada Level terakhir (Level 2). Kemudian dilakukan pengujian terhadap algoritma C4.5 menggunkana metode validasi K-Fold Cross validation dengan nilai fold=10 dapat diketahui tingkat akurasi sebesar 85%, nilai Precision sebesar 80% dan nilai Recall sebesar 66,67%. Sedangkan nilai AUC (Area Under Curve) adalah 0.833, hal ini menunjukan bahwa jika nilai AUC mendekati nilai 1 maka tingkat akurasi semakin bagus.

Kata Kunci: Data Mining; Klasifikasi; Algoritma C4.5; Prediksi; Ketersediaan Uang

Abstract−This study aims to provide an evaluation of the availability of money in ATM machines using data mining. Data mining with the C4.5 algorithm is used to predict cash demand or total cash withdrawals at ATMs. To determine the need for ATM cash based on cash transaction data. It is hoped that this forecasting can help the monitoring department in making decisions about the money requirements that must be allocated to each ATM machine. The results of this study are expected to assist the ATM management unit in optimizing and monitoring the availability of money at an ATM machine for cash needs, so that it can provide optimal service to customers. Algortima C4.5 is an algorithm that is able to form a decision tree, where the decision tree will then generate new knowledge. The results of the test matched the data on the availability of money at the ATM machine. The results of implementing the C4.5 method on the availability of money at the ATM machine are seen from the travel time to the ATM location and also the remaining balance in the machine. The resulting decision tree model is to make the balance variable as the root, then the travel time as a branch at Level 1 with the variables fast, medium, long, and the bank becomes a branch at the last level (Level 2). Then the C4.5 algorithm was tested using the K-Fold Cross validation method with the value of fold = 10, it can be seen that the accuracy rate is 85%, the Precision value is 80% and the Recall value is 66.67%. While the AUC (Area Under Curve) value is 0.833, this shows that if the AUC value approaches the value 1, the accuracy level is getting better.

Keywords: Data Mining; Clasification; C4.5 Algorithm; Prediction; Availability of Money

1. PENDAHULUAN

Salah satu fasilitas perbankan yang ditawarkan sebagai salah satu alternative delivery channels dalam bertransaksi dengan bank adalah Anjungan Tunai Mandiri (ATM) yaitu suatu mesin atau alat yang berfungsi sebagai pelayan bank secara elektronik yang melaksanakan fungsi teller secara otomatis. Fasilitas perbankan ini diberikan oleh bank kepada nasabahnya sebagai instrumen dan alat bertransaksi dari produk utamanya yaitu tabungan[1]

Berdasarkan data tahun 2009 sampai 2016 penggunaan alat pembayaran non tunai terus meningkat setiap tahunnya hal ini menunjukan bahwa pembyaran secara non tunai sudah bias diterima masyarakat[2]

Pengontrolan pengisian uang mesin ATM yang di selesaikan dengan teknologi informasi dapat memberikan manfaat bagi perusahaan, teknologi informasi dapat menunjang kinerja perusahaan serta informasi yang dihasilkan menjadi lebih baik. Oleh karena itu kebutuhan akan suatu teknologi informasi yang dapat menunjang proses pengontrolan pengisian uang pada mesin ATM perusahaan sangat diperlukan guna membantu dalam pengambilan keputusan atau sebagai bahan masukan kembali bagi perkembangan perusahaan tersebut. PT. SSI yang bergerak di bidang pengisian uang mesin ATM dalam mendukung aktivitas bisnis saat ini menggunakan aplikasi Microsoft Excel untuk penyimpanan data dan jumlah uang yang diterima dari Bank pemilik ATM, serta informasi ATM mana yang sudah atau belum diisi kembali [3]

Algoritma C4.5 umumnya digunakan untuk pengklasifikasian data, selain algoritma C4.5 algoritma ID3 dan K-Nearest juga dapat digunakan untuk pengklasifikasian data. Studi kinerja K-Nearest Neighbor dan C4.5 sudah dilakukan penelitian dalam menentukan kemungkinan pengunduran diri mahasiswa di STMIK AMIKOM Yogyakarta, hasil penelitian yang diperoleh adalah kinerja algoritma C4.5 lebih cepat dan akurat dibandingkan

Submitted: **18/03/2021**; Accepted: **14/04/2021**; Published: **25/04/2021**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor 2, April 2021, Page 556-563
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i2.2933

dengan algoritma K-Nearest [4]]. Algoritma C4.5 memiliki tingkat ketelitian yang tinggi dalam menghasilkan sebuah keputusan, ketelitiannya hingga 94 % pada tahap pelatihan dan 93 % pada tahap uji coba [5]. Dilakukan pengujian menggunakan dua algoritma klasifikasi yaitu algoritma C4.5 dan Algoritma Naïve Bayes, C4.5 menghasilkan tingkat akurasi yaitu 80%[6]

Menurut penelitian tentang penilaian kinerja karyawan menggunakan metode C4.5 Algoritma ini menggunakan konsep data mining yang melakukan proses penggalian informasi pada keputusan-keputusan sebelumnya, kemudian dijadikan informasi yang membentuk pola pohon keputusan. Proses tersebut dimulai dengan pengolahan data training, kemudian data tersebut dicari prioritas yang paling besar dan dijadikan sebuah rootpada pohon. Proses tersebut diulang dengan mencari gain terbesar sampai bertemu perbandingan yang bersih dan pohon berhenti dibentuk [7]

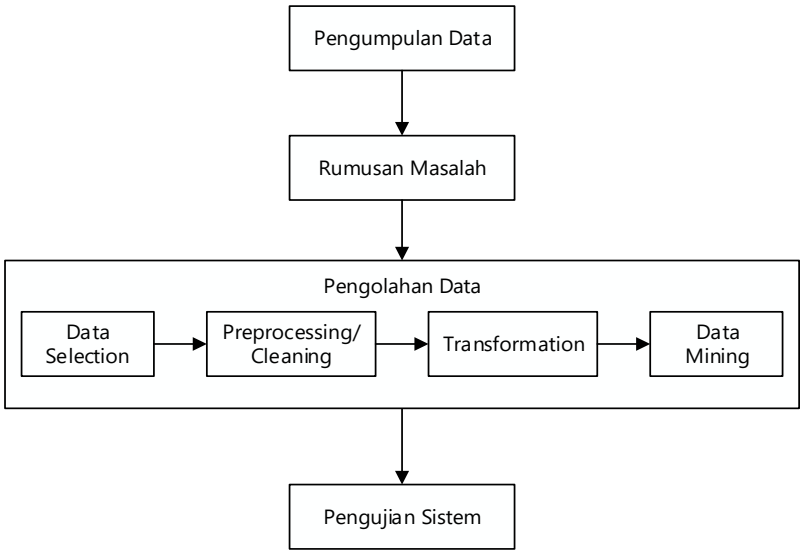
Pada penelitian [8], dilakukan Penelitian menggunakan Algoritma Naïve Bayes untuk klasifikasi waktu kelulusan mahasiswa menghasilkan model klasifikasi dengan nilai akurasi, precision, recall, dan f1-score sebesar 68%, 61.3%, 65.3%, dan 61% yang dihitung menggunakan metode 10-Fold Cross Validaiton, dan Confusion Matrix. Pada penelitian [9], dari pengujian yang telah dilakukan dengan model Cross Validation dengan nilai fold K=5 didapatkan sebuah hasil rata-rata akurasi C4.5 yaitu 79,28% dan Naive bayes 77,58%

Model k-Fold Cross Validation digunakan untuk mengetahui nilai k Optimal[10], Prinsip dasar model cross validation adalah membagi keseluruhan data menjadi data latih dan data uji[11]. Model confusion matrix yang digunakan memperoleh hasil akurasi sebesar 88.01% dengan kriteria akurasi good classification[12]. Pada penelitian ini penulis telah membandingkan dan menganalisis klasifikasi dan algoritma pembelajaran dengan Rapidminer. Rapidminer merupakan software open-source yang bagus untuk data mining dalam melakukan analisis dan simulasi. software ini menyediakan lingkungan pengembangan aplikasi yang cepat dan visualisasi data yang sesuai [13]

Adapun metode problem solving yang digunakan dalam penelitian ini adalah Data Mining. Data Mining yang ditawarkan menggunakan algoritma C4.5. Diharapkan dengan adanya peramalan tersebut dapat membantu bagian monitoring dalam mengambil keputusan mengenai kebutuhan uang yang harus dialokasikan ke setiap mesin ATM. Hasil penelitian ini diharapkan dapat membantu unit pengelola ATM dalam mengotimalkan dan monitoring saldo pada mesin ATM untuk menentukan kebutuhan uang, sehingga dapat memberikan pelayanan yang optimal kepada nasabah.

2. METODOLOGI PENELITIAN

Pada bagian ini akan menjelaskan mengenai kerangka kerja penelitian dimana pada langkah ini menjabarkan langkah-langkah yang digunakan dalam penyelesaian penelitian mulai dari identifikasi masalah sampai dengan pengujian sistem. Adapun kerangka kerja dari penelitian ini dapat digambarkan pada gambar 1 berikut :



Gambar 1. Kerangka Kerja Penelitian

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

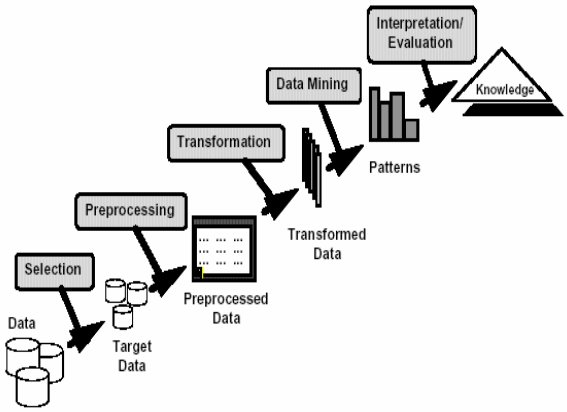
1. Pengumpulan Data

   Pada tahap ini terdapat dua proses yaitu studi lapangan dan studi kepustakaan. Pada proses studi lapangan, untuk memperoleh data yang berhubungan dengan judul yang penulis angkat dengan melakukan wawancara dan sampling. Sedangkan pada proses studi kepustakaan, penulis mengumpulkan data berdasarkan kepustakaan dengan menganalisa dan membaca buku-buku, jurnal-jurnal, serta artikel yang berhubungan dengan judul yang diambil.

Submitted: **18/03/2021**; Accepted: **14/04/2021**; Published: **25/04/2021**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor 2, April 2021, Page 556-563
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i2.2933

2. Rumusan Masalah
   Tahap selanjutnya penulis merumuskan permasalahan berdasarkan latar belakang yang telah penulis uraikan.
3. Pengolahan Data
   Pada tahap pengolahan data mining, ada beberapa proses yang dibutuhkan dalam pengolahan data mining antara lain Data Selection, Preprocessing/Cleaning, Transformation, dan Data mining. Pada Proses Data Selection, penulis melakukan seleksi data yang sesuai dan dibutuhkan dalam melakukan analisa. Selanjutnya pada proses Preprocessing/Cleaning penulis melakukan pembersihan data apabila terdapat data yang tidak relevan, noise dan data yang tidak konsisten serta menghilangkan data apabila terdapat duplikat atau sama. Selanjutnya pada proses Transformation, penulis melakukan perubahan data sesuai dengan yang dibutuhkan. Dan terakhir pada proses Data Mining, penulis melakukan proses analisa menggunakan algoritma C4.5 untuk mencari asosiasi dari permasalahan yang sudah diuraikan di latar belakang. Proses Data mining secara garis besar dapat dilihat pada gambar 2 berikut :



Gambar 2. Aliran Informasi dalam *Data Mining*

Dari gambar 2 dapat dijelaskan tahapan pertama yaitu *Data Selection* merupakan proses penyeleksian data dari data mentah sebelum dilakukan tahap berikutnya untuk penggalian informasi, tahap kedua *Pre-processing / Cleaning* merupakan Hasil dari data seleksi kemudian dilakukan proses cleaning yaitu pembersihan terhadap data yang menjadi focus terhadap penelitian dengan menghilangkan duplikasi dan memeriksa data yang tidak konsisten. Tahap ketiga yaitu 3 *Transformation*, proses transformasi data pada KDD merupakan proses pengklasifikasian data dan sangat bergantung pada jenis data atau pola informasi yang akan dicari Tahap keempat data mining, pada bagian ini merupakan bagian yang paling penting yaitu pross mencari pola terhadap data terpilih dengan menggunakan algoritma tertentu. Dan tahap terakhir *Interpretation / Evaluation*, Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya dan perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pengguna.

4. Pengujian Sistem
   Tahap akhir pada sistem ini adalah tahap uji coba sistem, dimana menyesuaikan hasil dari sistem dengan perhitungan manual dan menguji validasi dari algoritma dengan model cross validation[5]

3. HASIL DAN PEMBAHASAN

Pada bab ini membahas mengenai pengolahan data dengan algoritma C4.5 kemudian dilakukan uji validasi dengan model cross validation untuk menguji tingkat akurasi dari algoritma C4.5

3.1 Pengolahan Data

Pada pengolahan data terdapat beberapa proses yaitu yaitu Data *Selection*, *Cleanning*, *Transformation* dan data mining:

3.1.1 Data *Selection*

Pada tahap ini melakukan seleksi terhadap data yang digunakan dalam uji coba, dan kemudian setelah proses data selection, melakukan proses data cleaning, dimana hasil data cleaning dapat dilihat pada table 1

Tabel 1. Data Cleaning

BANK	SALDO %	WAKTU TEMPUH	KETERANGAN
BNI	< 30 %	< 60 MENIT	ISI
BNI	< 30 %	< 60 MENIT	ISI
BNI	< 30 %	< 60 MENIT	ISI

Submitted: **18/03/2021**; Accepted: **14/04/2021**; Published: **25/04/2021**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor 2, April 2021, Page 556-563
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i2.2933

BANK	SALDO %	WAKTU TEMPUH	KETERANGAN
BNI	30 % - 50 %	< 60 MENIT	TIDAK
BNI	30 % - 50 %	60 - 90 MENIT	ISI
BNI	30 % - 50 %	< 60 MENIT	TIDAK
BNI	30 % - 50 %	60 - 90 MENIT	ISI
BNI	30 % - 50 %	< 60 MENIT	TIDAK
BNI	30 % - 50 %	60 - 90 MENIT	ISI
BNI	50 % - 60 %	> 90 MENIT	ISI
BNI	30 % - 60 %	< 60 MENIT	TIDAK
BNI	50 % - 60 %	> 90 MENIT	ISI
BNI	30 % - 60 %	< 60 MENIT	TIDAK
BNI	50 % - 60 %	> 90 MENIT	ISI
BNI	30 % - 60 %	< 60 MENIT	TIDAK
BRI	< 30 %	< 60 MENIT	ISI
BRI	< 30 %	< 60 MENIT	ISI
BRI	< 30 %	< 60 MENIT	ISI
BRI	30 % - 50 %	< 60 MENIT	ISI

3.1.2 Data Transformation

Transformasi data adalah proses pengkelasan data, adapun data yang akan diklasifikasikan adalah data BANK, Saldo dan Waktu Tempuh. Proses transformasi data berdasarkan beberapa variabel yaitu:

Tabel 2. Klasifikasi Presentase Saldo

Saldo %	Klasifikasi Saldo
< 30 %	Rendah
30% - 50%	Sedang
> 60 %	Tinggi

Tabel 3. Klasifikasi Waktu Tempuh

Waktu Tempuh	Klasifikasi
< 60 Menit	Cepat
60 - 90 Menit	Sedang
> 90 Menit	Lama

Berdasarkan kriteria klasifikasi data dari tabel 2 sampai tabel 3 diperoleh hasil transformasi seperti terlihat pada tabel 4. Data hasil transformasi selanjutnya akan digunakan sebagai data yang akan diproses menggunakan algoritma C4.5 dalam pembuatan pohon keputusan

Tabel 4. Hasil Transformasi

BANK	Saldo	Waktu Tempuh	KETERANGAN
BNI	Rendah	Cepat	ISI
BNI	Rendah	Cepat	ISI
BNI	Rendah	Cepat	ISI
BNI	Sedang	Sedang	ISI
BNI	Sedang	Sedang	ISI
BNI	Sedang	Sedang	ISI
BNI	Sedang	Cepat	TIDAK
BNI	Sedang	Cepat	TIDAK
BNI	Sedang	Cepat	TIDAK
BNI	Tinggi	Lama	ISI
BNI	Tinggi	Lama	ISI
BNI	Tinggi	Lama	ISI
BNI	Tinggi	Cepat	TIDAK
BNI	Tinggi	Cepat	TIDAK
BNI	Tinggi	Cepat	TIDAK
BRI	Rendah	Cepat	ISI
BRI	Rendah	Cepat	ISI
BRI	Rendah	Cepat	ISI
BRI	Sedang	Cepat	ISI

Submitted: **18/03/2021**; Accepted: **14/04/2021**; Published: **25/04/2021**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor 2, April 2021, Page 556-563
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i2.2933

3.1.3 Perhitungan Algoritma C4.5

Data hasil transformasi selanjutnya dianalisa untuk menghasilkan sebuah pohon keputusan dengan menggunakan algoritma C4.5, Algoritma C4.5 menggunakan kriteria penguatan dalam menentukan fitur yang merupakan pemecah simpul pada pohon yang diinduksi[14], secara umum algortima C4.5 untuk membangun pohon keputusan adalah sebagai berikut [15]:

a. Memnghitung Entropy dan Gain
b. Memilih Gain tertinggi sebagai akar (Node)
c. Mengulangi proses perhitungan Entropy dan Gain untuk mencari cabang sampai semua kasus pada cabang memiliki kelas yang sama yaitu pada saat semua variabel telah menjadi bagian dari pohon keputusan atau masing –masing variabel telah memiliki daun atau keputusan.
d. Membuat Rule berdasarkan pohon keputusan

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan persamaan 1 berikut

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

Dimana :

S : Himpunan Kasus
A : Atribut
n : Jumlah Partisi Atribut A
|Si| : Jumlah Kasus pada Partisi ke-I
|S| : Jumlah Kasus dalam S

Sementara itu, perhitungan nilai entropy dapat dilihat pada persamaan 2 berikut

$$Entropy\ (S) = \sum_{i=1}^{n} - pi * log_2\ pi \quad (2)$$

Perhitungan Entropy dan Gain node 1 adalah sebagai berikut:

a. Menghitung Entropy_Total

Entropy(Total) = $\left(-\frac{32}{45} * \text{Log}_2 \frac{32}{45}\right) + \left(-\frac{13}{45} * \text{Log}_2 \frac{13}{45}\right) = 0.867$

b. Menghitung Entropy_Bank

Entropy(Bank, BNI)=15, Entropy(Bank, BRI)=10, Entropy(Bank, BTN)=8, Entropy(Bank, Mandiri)=12

Entropy(Bank, BNI) = 15

$$\left(-\frac{12}{15} * \text{Log}_2 \frac{12}{15}\right) + \left(-\frac{3}{15} * \text{Log}_2 \frac{3}{15}\right) = 0{,}722$$

Entropy(Bank, BRI) = 10

$$\left(-\frac{5}{10} * \text{Log}_2 \frac{5}{10}\right) + \left(- -\frac{5}{10} * \text{Log}_2 \frac{5}{10}\right) = 1$$

Entropy(Bank, BTN) = 8

$$\left(-\frac{4}{8} * \text{Log}_2 \frac{4}{8}\right) + \left(-\frac{4}{8} * \text{Log}_2 \frac{4}{8}\right) = 1$$

Entropy(Bank, Mandiri) = 12

$$\left(-\frac{11}{12} * \text{Log}_2 \frac{11}{12}\right) + \left(-\frac{1}{12} * \text{Log}_2 \frac{1}{12}\right) = 0{,}414$$

c. Menghitung Gain dari masing-masing atribute

Menghitung Gain Total terhadap Bank

$$Gain\ (Total, \text{Bank}) = \text{Entropy(Total)} - \sum_{i=1}^{n} \frac{|Bank|}{|Total|} * Entropy(Bank)$$

Gain (Total, Bank) =

$$0{,}867 - \left(\left(\frac{15}{45} * 0{,}722\right) + \left(\frac{10}{45} * 1\right) + \left(\frac{8}{45} * 1\right) + \left(\frac{12}{45} * 0{,}414\right)\right) = \mathbf{0,116}$$

Submitted: **18/03/2021**; Accepted: **14/04/2021**; Published: **25/04/2021**

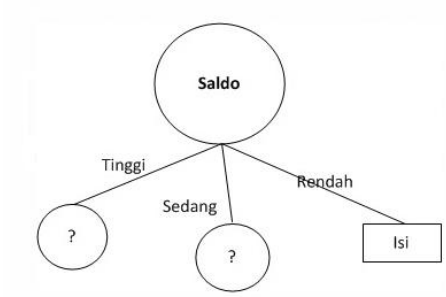
JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor 2, April 2021, Page 556-563
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i2.2933

Hasil perhitungan entropy dan gain untuk mencari akar dapat dilihat pada tabel.5

Tabel 5. Hasil Perhitungan Entropy dan Gain

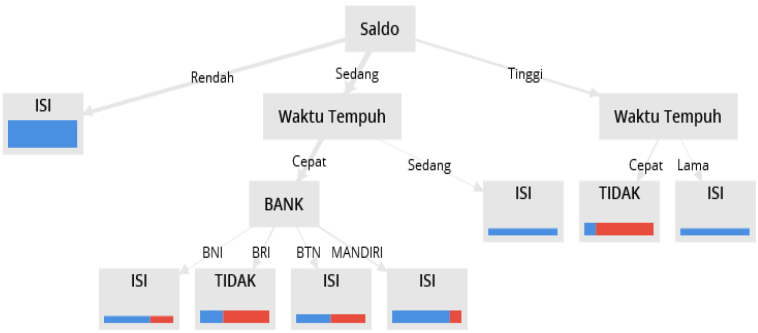
Node	Atribut	nilai	Jumlah Kasus	Isi	Tidak	Total Entropy	Gain
1	**Total**		45	32	13	0,867	
	Bank						**0,116**
		BNI	15	12	3	0,722	
		BRI	10	5	5	1,000	
		BTN	8	4	4	1,000	
		Mandiri	12	11	1	0,414	
	Saldo						**0,207**
		Tinggi	9	4	5	0,991	
		Sedang	22	14	8	0,946	
		Rendah	14	14	0	0,000	
	Waktu Tempuh						**0,071**
		Cepat	39	26	13	0,918	
		Sedang	3	3	0	0,000	
		Lama	3	3	0	0,000	

Berdasarkan hasil perhitungan entropy dan gain untuk mencari akar nilai gain tertinggi adalah variabel Saldo dengan 3 nilai variabel yaitu Tinggi, Sedang dan Rendah. Selanjutnya adalah menggambarkan hasil perhitungan entropy dan gain dalam bentuk pohon keputusan. Pohon keputusan untuk mencari nilai akar dapat dilihat pada gambar 3.



Gambar 3. Pohon Keputusan Hasil Perhitungan Entropy dan Gain Untuk Mencari Akar

Proses perhitungan entropy dan gain akan berulang sampai semua cabang memiliki sebuah keputusan. Pohon keputusan akhir dapat dilihat pada gambar 4.



Gambar 4. Pohon Keputusan Akhir

3.1.4 Membuat Rule dari Pohon Keputusan

Rule yang diperoleh dari pohon keputusan sesuai pada gambar 4 adalah sebagai berikut:
1. If Saldo = Rendah Then PA (Pengisian ATM) = Isi
2. If Saldo = Sedang And Waktu Tempuh = Cepat And Bank = BNI Then PA (Pengisian ATM) = Isi
3. If Saldo = Sedang And Waktu Tempuh = Cepat And Bank = BRI Then PA (Pengisian ATM) = Tidak

Submitted: **18/03/2021**; Accepted: **14/04/2021**; Published: **25/04/2021**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor 2, April 2021, Page 556-563
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i2.2933

4. If Saldo = Sedang And Waktu Tempuh = Cepat And Bank = BTN Then PA (Pengisian ATM) = Isi
5. If Saldo = Sedang And Waktu Tempuh = Cepat And Bank = Mandiri Then PA(Pengisian ATM) = Isi
6. If Saldo = Sedang And Waktu Tempuh = Sedang Then PA (Pengisian ATM) = Isi
7. If Saldo = Tinggi And Waktu Tempuh = Cepat Then PA(Pengisian ATM) = Tidak
8. If Saldo = Tinggi And Waktu Tempuh = Lama Then PA (Pengisian ATM) = Isi

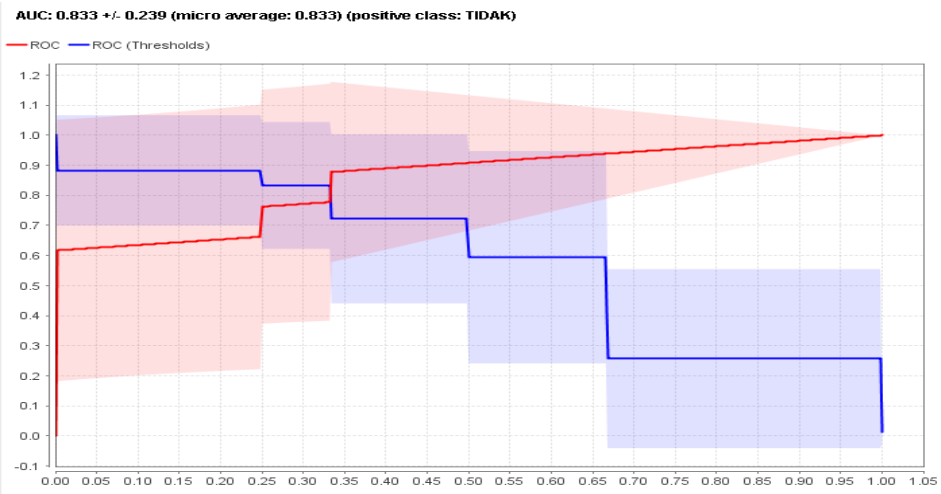
3.2 K-fold Cross Validation

Pada tahap pengujian menggunakan model K-Fold Cross Validation untuk muji tingkat akurasi dari algoritma C4.5 yang digunakan. Cross-validasi atau dapat disebut estimasi rotasi adalah sebuah teknik validasi model untuk menilai bagaimana hasil statistik analisis akan menggeneralisasi kumpulan data independen[16]. K-fold cross validation merupakan salah satu teknik dari validasi silang yang digunakan menghilangkan bias pada data. Pelatihan dan pengujian dilakukan sebanyak k kali [17]. Nilai k-fold cross validation yang digunakan adalah 10-fold. Hasil akurasi confusion matrix terhadap pengujian data dapat dilihat pada tabel 6

Tabel 6. Confision Matrix 10-Fold Cross Validation

Prediction Value	Actual value	
	True ISI	True TIDAK
Pred. ISI	26	4
Pred. Tidak	3	12

Hasil *AUC* (*Area Under Curve)* pada pengujian 10-fold sebagai berikut



Gambar 5. *AUC* (*Area Under Curve)* pada pengujian 10-fold

Hasil AUC adalah 0,833 artinya semakin nilai AUC mendekati angka 1 maka tingkat akurasi dari pengujian semakin bagus. Hasil dari evaluasi dan validasi terhadap pengujian 10-fold cross validation dapat dilihat pada tabel 7.

Tabel 7. Evaluasi dan Validasi Pengujian 10-Fold

K-Fold Validasi	Acuracy	Precision	Recall	AUC
10 - Fold	85.00%	80.00%	66.67%	0.833

4. KESIMPULAN

Berdasarkan pembahasan yang telah dijelaskan disimpulkan bahwa pengujian terhadap algoritma C4.5 menggunkana metode validasi K-Fold Cross validation dengan nilai fold=10 dapat diketahui tingkat akurasi sebesar 85%, nilai Precision sebesar 80% dan nilai Recall sebesar 66,67%. Sedangkan nilai AUC (*Area Under Curve)* adalah 0.833, hal ini menunjukan bahwa jika nilai AUC mendekati nilai 1 maka tingkat akurasi semakin bagus.

UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 5, Nomor 2, April 2021, Page 556-563
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI 10.30865/mib.v5i2.2933

REFERENCES

[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. Lintangsari, 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] M. Di and S. Amikom, "Perbandingan metode nearest neighbor dan algoritma c4.5 untuk menganalisis kemungkinan pengunduran diri calon mahasiswa di stmik amikom yogyakarta," vol. 10, no. 1, 2009.
[5] K. and S. A. Rajesh, "Analysis of SEER Dataset for Breast Cancer Diagnosis using C4.5 Classification Algorithm," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 1, no. 2, pp. 72–77, 2012, [Online]. Available: www.ijarcce.com.
[6] S. Masripah, "Komparasi Algoritma Klasifikasi Data Mining untuk Evaluasi Pemberian Kredit," *Bina Insa. ICT J.*, vol. 3, no. 1, p. 234336, 2016.
[7] W. Julianto, R. Yunitarini, and M. K. Sophan, "Algoritma C4.5 Untuk Penilaian Kinerja Karyawan," *Scan*, vol. Vo. IX, no. No. 2, pp. 33–39, 2014.
[8] I. W. Saputro and B. W. Sari, "Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa," *Creat. Inf. Technol. J.*, vol. 6, no. 1, p. 1, 2020, doi: 10.24076/citec.2019v6i1.178.
[9] D. L. Arisandy, "Analisis Perbandingan Algoritma Naive Bayes dan Algoritma C4.5 untuk Klasifikasi Multi Data," no. 1310651061, 2017.
[10] 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.
[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] P. Mochamad Rizki Ilham, "Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Kepuasan Pelanggan Taksi Kosti," *Simplementasi Data Min. Menggunakan Algoritm. C4.5 Untuk Prediksi Kepuasan Pelangg. Tak. Kosti*, vol. Vol. 4, No, no. 5, p. 11, 2016.
[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] A. Lestari, "Increasing Accuracy of C4 . 5 Algorithm Using Information Gain Ratio and Adaboost for Classification of Chronic Kidney Disease," pp. 32–38, 2020.
[15] F. Harahap, "Penerapan data Mining dalam Pemilihan Mobil Menggunakan Algoritma C4.5," *J. VOI (Voice Informatics)*, vol. 7, no. x, 2018.
[16] 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.
[17] 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.

Submitted: **18/03/2021**; Accepted: **14/04/2021**; Published: **25/04/2021**