
Implementasi Data Mining Pada Klasifikasi Ketidakhadiran Pegawai Menggunakan Metode C4.5

Nandang Iriadi¹, Lutfi Setioningtias², Priatno³

¹ Program Studi Teknologi Komputer, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Jl. Kramat Raya No.98, Jakarta Pusat, DKI Jakarta 10450, Indonesia

² Program Studi Sistem Informasi, Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri Jl. Jatiwaringin No. 2, Cipinang Melayu, Makasar, Jakarta Timur - 13620, Indonesia

³ Program Studi Teknologi Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika Jl. Kramat Raya No.98, Jakarta Pusat, DKI Jakarta 10450, Indonesia

e-mail :¹nandang.ndi@bsi.ac.id, ²lutfisetioningtias@gmail.com,
³priatno.prn@bsi.ac.id

Artikel Info : Diterima : 13-01-2021 | Direvisi : 17-01-2021 | Disetujui : 18-01-2021

Abstrak - Tidak semua ketidakhadiran pegawai memiliki alasan yang selalu di cap buruk. Adapula ketidakhadiran yang bisa di toleransi oleh suatu instansi atau perusahaan, dimana alasan tersebut merupakan alasan yang dapat diterima misalnya dikarenakan alasan sakit, atau izin tertentu yang sekiranya dianggap masuk akal dan tentunya memiliki jangka waktu tertentu sehingga tidak begitu menyulitkan pihak tempatnya bekerja. Tujuan penelitian adalah mengetahui sejauh mana algoritma C4.5 dapat membantu perhitungan klasifikasi untuk menemukan solusi agar produktivitas kerja pegawai meningkat. Untuk menemukan solusi agar produktivitas kerja pegawai meningkat membutuhkan teknik dan cara tertentu yaitu dengan klasifikasi data mining menggunakan *Algoritma C4.5*. Supaya pengolahan data mendapatkan hasil yang baik, maka digunakan *Tools Rapid Miner 5*. *Algoritma C4.5* termasuk jenis aturan klasifikasi pada *Data Mining*. Penting tidaknya suatu aturan klasifikasi dapat diketahui dengan dua parameter, yaitu *Entropy* dan *Gain* tertinggi. Setelah dilakukan pengujian, maka hasil yang didapat dari *Algoritma C4.5* memiliki akurasi sebesar 81,08%.

Kata Kunci : Ketidakhadiran, *Algoritma C4.5*, *Data Mining*, *Rapid Miner 5*

Abstract - Not all employee absences have reasons that are always labeled as bad. There is also an absence that can be tolerated by an agency or company, where these reasons are acceptable reasons, for example, due to illness, or certain permits that are considered reasonable and of course have a certain period so that it is not too difficult for the party where he works. The research objective is to determine the extent to which the C4.5 algorithm can help the classification calculation to find solutions so that employee works productivity increases. To find a solution so that employee work productivity increases requires certain techniques and methods, namely by classifying data mining using the C4.5 Algorithm. For data processing to get good results, Rapid Miner 5 Tools are used. The C4.5 algorithm is a type of classification rule in Data Mining. The importance of a classification rule can be determined by two parameters, namely Entropy and the highest Gain. After testing, the results obtained from the C4.5 Algorithm have an accuracy of 81.08%.

Keywords: Absence, C4.5 Algorithm, Data Mining, Rapid Miner 5

PENDAHULUAN

Ketidakhadiran atau kealpaan individu terhadap pekerjaan atau tugas dan kajibannya yang sudah dimiliki oleh masing-masing individu tersebut dan berjalan secara rutin. Secara umum, absensi adalah bagian dari kinerja individual yang dikategorikan sebagai kemalasan dan termasuk dalam pelanggaran antara pegawai terhadap



atasannya dalam ruang lingkup formalitas pekerjaan di dalam suatu instansi atau perusahaan. Kebiasaan tidak hadir di tempat kerja merupakan suatu tindakan moral yang dipandang buruk di mata sosial masyarakat

Berdasarkan permasalahan di atas, penelitian ini mengambil data ketidakhadiran pegawai di tempat kerja selama 3 tahun dari salah satu perusahaan pengiriman di Brazil yang data-datanya di peroleh dari penyimpanan dataset *University of California Irvine (UCI) Machine learning Repository*. Tujuan penelitian adalah untuk mengetahui sejauh mana algoritma C4.5 dapat membantu perhitungan klasifikasi untuk menemukan solusi agar produktivitas kerja pegawai meningkat. Data-data yang tersimpan dan sudah tidak dipergunakan lagi dengan jumlah dataset yang cukup besar, yakni berjumlah 740 dataset ini membutuhkan sebuah teknik atau metode pengolahan data agar data tersebut bisa diolah sebagai suatu informasi yang dibutuhkan dan bermanfaat. Untuk mewujudkannya, maka metode yang digunakan dalam penelitian ini adalah metode Algoritma C4.5.

Pada prosesnya data mining akan mengekstrak informasi yang berharga dengan cara menganalisis adanya pola-pola ataupun hubungan keterkaitan tertentu dari data-data yang berukuran besar. Dengan adanya data mining dapat menemukan pengetahuan yang dapat dipergunakan oleh para praktisi kesehatan agar dapat meningkatkan kualitas pelayanan. Pengetahuan yang ditemukan dapat juga digunakan oleh praktisi medis untuk mengurangi jumlah efek samping dari obat untuk menyarankan lebih murah dalam terapi yang setara dengan berbagai alternatif. Proses mining data lebih dari sekedar analisa data yang meliputi pengklasifikasian, klustering, asosiasi dan prediksi. (M Anbarasi, E Anupriya, 2010)

Data Mining. Absensi adalah tingkat ketidakhadiran pegawai yang berkenaan dengan tanggung jawab. Saat ini banyak pegawai yang melanggar jam kerja, menaati jam kerja merupakan kewajiban seorang pegawai, yaitu masuk kerja dan menaati ketentuan jam kerja. (Ardyan Pratama, 2016)

“Sistem Pendukung Keputusan atau yang disebut dengan DSS (Decision Support System) adalah sistem berbasis komputer yang ditujukan untuk membantu pengambil keputusan dengan memanfaatkan data dan model untuk mengidentifikasi memecahkan masalah dan membuat keputusan”(Ichsan, Syahru, 2013)

Menurut (Christiana, 2013) mengemukakan bahwa: Organisasi telah dibentuk untuk memenuhi kebutuhan masyarakat. Dalam dunia yang kompetitif saat ini, organisasi harus berorientasi pada pertumbuhan. Hal ini memungkinkan ketika produktivitas dipastikan berkenaan dengan kuantitas dan kualitas produk. Ketidakhadiran dan pergantian karyawan memiliki dampak negatif terhadap produktivitas. Karyawan yang absen seringkali tidak dapat berkontribusi terhadap produktivitas dan pertumbuhan organisasi. Demikian pula, pergantian karyawan menyebabkan peningkatan biaya produksi.

Menurut (Hermawati, 2005), mengemukakan bahwa: Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis.

Knowledge Discovery In Databases (KDD) adalah proses non trivial untuk mencari dan mengidentifikasi pola (pattern) dalam data (Wahbeh, 2011)

Data mining menurut David Hand, Heikki Mannila, dan Padhraic Smyth dari MIT adalah analisa terhadap data (biasanya data yang berukuran besar) untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut. (Yuli Mardi, 2014)

Menurut (Widodo et al., 2013) mengemukakan bahwa: Pohon keputusan (*decision tree*) merupakan salah satu teknik terkenal dalam data mining dan merupakan salah satu metode yang populer dalam menentukan keputusan suatu kasus. Hal ini karena metode ini tidak memerlukan proses pengelolaan pengetahuan terlebih dahulu dan dapat menyelesaikan dengan sederhana kasus-kasus yang memiliki dimensi yang besar. Akurasi sangat baik asalkan data yang akan dijadikan patokan merupakan data yang akurat.

Menurut Han dalam (Widodo et al., 2013) mengemukakan bahwa, “Algoritma C4.5 merupakan struktur pohon dimana terdapat simpul yang mende Penelitiankan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Algoritma C4.5 secara rekursif mengunjungi setiap simpul keputusan, memilih pembagian yang optimal, sampai tidak bisa dibagi lagi. Algoritma C4.5 menggunakan konsep *information gain* atau *entropy reduction* untuk memilih pembagian yang optimal”.

Menurut Kusriani dalam (Widodo et al., 2013) Terdapat beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5, yaitu:

1. Menyiapkan data *training*. Data *training* biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah di kelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus:

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

Keterangan:

S = himpunan kasus

n = jumlah partisi S

p_i = proporsi S_i terhadap S

3. Kemudian hitung nilai *gain* menggunakan rumus:

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

* *Entropy* (S_i)

Keterangan:

S = himpunan kasus

A = fitur

n = jumlah partisi atribut A

$|S_i|$ = proporsi S_i terhadap S

$|S|$ = jumlah kasus dalam S

4. Ulangi langkah ke-2 hingga semua *record* terpartisi.
5. Proses partisi pohon keputusan akan berhenti saat:
 - a. Semua *record* dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut di dalam *record* yang dipartisi lagi.
 - c. Tidak ada *record* di dalam cabang yang kosong.

Menurut (Nofriansyah, 2014) mengemukakan bahwa, Adapun kriteria atau ciri-ciri dari sebuah keputusan adalah sebagai berikut:

1. Banyak pilihan atau alternatif.
2. Ada kendala atau surat.
3. Mengikuti suatu pola atau model tingkah laku, baik yang terstruktur maupun tidak terstruktur.
4. Banyak *input* atau variabel.
5. Ada faktor resiko, dibutuhkan kecepatan, ketepatan, dan keakuratan.

Menurut (Mutoi Siregar & Puspabhuana, 2018), *RapidMiner* merupakan perangkat lunak yang bersifat terbuka (*open source*). *RapidMiner* adalah sebuah solusi untuk melakukan analisis terhadap data mining, *text mining* dan analisis prediksi. *RapidMiner* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. *RapidMiner* memiliki kurang lebih 500 operator data mining, termasuk *operator* untuk *input*, *output*, data *preprocessing* dan visualisasi. *RapidMiner* merupakan *software* yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. *RapidMiner* ditulis dengan menggunakan bahasa *java* sehingga dapat bekerja di semua sistem operasi.

Menurut (Sugiyono, 2010) mengemukakan bahwa: *Nonprobability Sampling* adalah teknik pengambilan sampel yang tidak memberi peluang atau kesempatan sama bagi setiap unsur atau anggota populasi untuk dipilih menjadi sampel. Teknik sampel ini meliputi *sampling sistematis*, kuota, insidental, *purposive*, jenuh, dan *snowball*.

Sedangkan teknik pengambilan sampel dari *Nonprobability sampling* yang diambil di dalam penelitian ini adalah *Sampling Sistematis*.

Menurut (Sugiyono, 2010) mengemukakan bahwa, “*Sampling sistematis* adalah teknik pengambilan sampel berdasarkan urutan dari anggota populasi yang telah diberi nomor urut”.

Misalnya anggota populasi yang terdiri dari 100 orang. Dari semua anggota tersebut diberi nomor urut, yaitu nomor urut 1 sampai dengan nomor urut 100. Pengambilan sampel dapat dilakukan dengan nomor ganjil saja, genap saja atau dengan kelipatan dari bilangan tertentu, misalnya kelipatan dari bilangan 5. Untuk itu maka yang diambil sebagai sampel adalah nomor urut 1,5,10,15,20, dan seterusnya sampai dengan 100.

METODE PENELITIAN

Dalam penyusunan Penelitian ini, diperlukan beberapa tahapan untuk mencapai bagian yang telah ditetapkan sebelumnya. Tahapan penelitian yang dilakukan sebagai berikut:

- a. Fase Pemahaman Bisnis (*Business Understanding Phase*).
- b. Fase Pemahaman Data (*Data Understanding Phase*).
- c. Fase Pengolahan Data (*Data Preparation phase*).
- d. Fase Pemodelan (*Modelling Phase*).
- e. Fase Evaluasi (*Evaluation Phase*).
- f. Fase Penyebaran (*Deployment Phase*)

Terdapat 20 Atribut untuk kebutuhan penelitian, diantaranya adalah, Alasan Ketidakhadiran, Bulan Ketidakhadiran, Hari Dalam Seminggu, Musim, Biaya Transportasi, Jarak Tempuh/km, Waktu Pelayanan, Umur, Rata-rata Bekerja/Hari, Target yang Dicapai, Kegagalan Disiplin, Pendidikan, Anak, Peminum, Perokok, Binatang Peliharaan, Berat Badan, Tinggi Badan, Indeks Masa Tubuh, dan Waktu Ketidakhadiran dalam Jam.

Dalam penelitian ini, media pengumpulan datanya adalah berupa data sekunder. Data sekunder yang digunakan di dalam penelitian ini adalah menggunakan data dari *University of California Irvine (UCI) Machine learning Repositor*. Data yang diambil adalah data ketidakhadiran pegawai di tempat kerja selama 3 tahun dari salah satu perusahaan pengiriman di Brazil.

Populasi dalam penelitian ini merupakan semua data karyawan pengiriman sebanyak 740 *record* data dari salah satu data perusahaan di Brazil selama 3 tahun yang data-datanya tersimpan di *University of California Irvine (UCI) Machine learning Repository*.

Sampel yang digunakan di dalam penelitian ini adalah sebanyak 370 *record* data dan teknik pengambilan sampelnya adalah menggunakan salah satu teknik pengambilan sampel dari *Nonprobability Sampling*, yaitu *Sampling Sistematis*. Dikarenakan di dalam teknik pengambilan sampel secara *sampling sistematis* adalah menggunakan nomor urut sebagai acuannya, maka peneliti menggunakan urutan nomor urut ganjil yang dimulai dari nomor urut 1 sampai dengan nomor urut 739 yang pengambilan nomor urutnya tidak boleh lebih dari banyaknya populasi yang diambil, yaitu 740 *record* data.

1. Perhitungan Manual Algoritma C4.5.

Pada proses perhitungan manual, tahap-tahap untuk menentukan pohon keputusan (*Decision Tree*) dengan menggunakan algoritma C4.5 adalah sebagai berikut:

- a. Mengumpulkan dan menyiapkan dataset.
- b. Menghitung Nilai *Entropy*.
- c. Menghitung Nilai *Gain* pada setiap atribut untuk mencari nilai *gain* yang tertinggi dan dijadikan sebagai *Node Akar*.
- d. Proses perhitungan nilai-nilai pada *node* akar dan cabang.

2. Implementasi Pohon Keputusan (*Decision Tree*) dengan *Rapid Miner*.

Pada tahap ini dilakukan proses perhitungan menggunakan aplikasi *Rapid Miner* yang bertujuan untuk membandingkan hasil perhitungan manual dengan hasil perhitungan aplikasi. Pengujian dilakukan agar mendapatkan tingkat keakuratan data dan dapat ditarik kesimpulannya dari hasil tersebut.

HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengetahui sejauh mana algoritma C4.5 dapat membantu perhitungan klasifikasi untuk menemukan solusi agar produktivitas kerja pegawai meningkat. Di dalam menentukan hasil

penelitian ini, peneliti menggunakan dataset sebanyak 740 *record* data dengan sampel penelitian yang diambil adalah sebanyak 370 *record* data.

1. Perhitungan Manual Algoritma C4.5.

Langkah untuk menentukan pohon keputusan dengan menggunakan algoritma C4.5 dengan memakai sampel dataset sebanyak 370, yaitu:

1. Menyiapkan sampel dataset sebanyak 370 data untuk digunakan dalam penelitian ini. Dataset biasanya diambil dari data primer yang sebelumnya sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Hitung nilai *entropy*.

Setelah dilakukan perhitungan *entropy* seperti rumus, di dapat hasil nilai *entropy* sebagai berikut:

Total *Entropy*

$$= \left(\left(\frac{-348}{370} \right) \times \log_2 \left(\frac{348}{370} \right) \right) + \left(\left(\frac{-22}{370} \right) \times \log_2 \left(\frac{22}{370} \right) \right)$$

$$= 0,0832 + 0,2421$$

$$= 0,3253$$

3. Setelah itu hitunglah nilai *gain* untuk setiap atribut, lalu pilih nilai *gain* yang tertinggi untuk dijadikan akar dari pohon. Hitunglah nilai *gain* menggunakan rumus. Misalkan untuk atribut 'Bulan Ketidakhadiran', didapatkan nilai *gain* sebagai berikut:

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

Gain (Bulan Ketidakhadiran)

$$= 0,3253 - \left(\left(\left(\frac{23}{370} \right) \times 0 \right) + \left(\left(\frac{34}{270} \right) \times 0 \right) + \left(\left(\frac{45}{370} \right) \times 0,2623 \right) + \left(\left(\frac{24}{370} \right) \times 0 \right) \right.$$

$$+ \left(\left(\frac{31}{370} \right) \times 0,4587 \right) + \left(\left(\frac{27}{370} \right) \times 0,2285 \right) + \left(\left(\frac{37}{370} \right) \times 0,3034 \right)$$

$$+ \left(\left(\frac{27}{370} \right) \times 0 \right) + \left(\left(\frac{27}{370} \right) \times 0,6913 \right) + \left(\left(\frac{34}{370} \right) \times 0,6024 \right)$$

$$\left. + \left(\left(\frac{36}{370} \right) \times 0,5033 \right) + \left(\left(\frac{25}{370} \right) \times 0 \right) \right)$$

$$= 0,3253 -$$

$$(0+0+0,0319+0+0,0384+0,0167+0,0303+0+$$

$$0,0504+0,0554+0,0490) = 0,3253 - 0,2721$$

$$= 0,0532$$

- a. Proses untuk *node* akar.

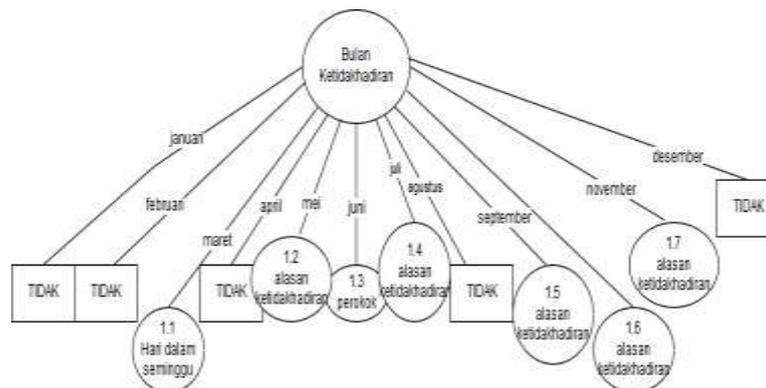
Macam-macam atributnya adalah alasan ketidakhadiran, bulan ketidakhadiran, hari dalam seminggu, musim, biaya transportasi, jarak tempuh/km, waktu pelayanan, umur, rata-rata bekerja/hari, target yang dicapai, kegagalan disiplin, pendidikan, anak, peminum, perokok, binatang peliharaan, berat badan, tinggi badan, indeks masa tubuh, dan waktu ketidakhadiran dalam jam. Setiap atribut memiliki nilai. Sedangkan label kelasnya ada pada kolom keputusan yaitu Tidak dan Ya. Kolom keputusan didapatkan dari nilai atribut "Kegagalan Disiplin" yang dijadikan sebagai label kelas untuk kemudian nilai "Tidak" dan "Ya" dijadikan sebagai kolom keputusan. Nilai "Tidak" berarti "Kegagalan disiplin = Tidak" dan nilai "Ya" berarti "Kegagalan disiplin = Ya". Kemudian data dianalisis, dataset tersebut memiliki 370 kasus yang terdiri dari 348 karyawan mengalami Kegagalan disiplin = Tidak dan 22 kasus karyawan mengalami Kegagalan disiplin = Ya .

Tabel 1. Hasil *gain* tertinggi untuk *Node* Akar

Node	Atribut	Nilai	Jumlah Nilai (S1)	Tidak (S1)	Ya (S1)	Entropy	Gain
	Januari	23	23	0	0		
	Februari	34	34	0	0		
	Maret	45	43	2	0,2623		
	April	24	24	0	0		
	Mei	31	28	3	0,4587		

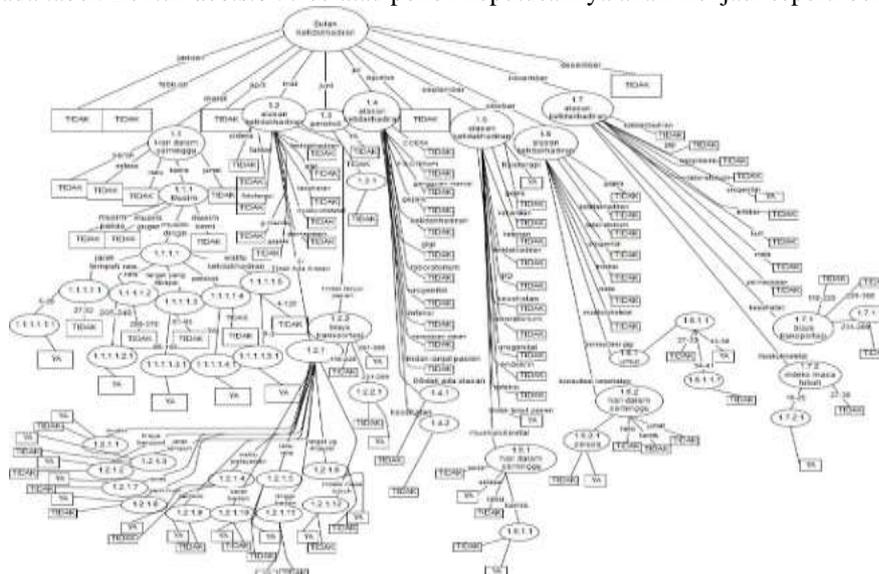
1	Bulan Ketidak hadiran	Juni	27	26	2	0,2285
		Juli	37	35	2	0,3034
		Agustus	27	27	0	0
		September	27	22	5	0,6913
		Oktober	34	29	5	0,6024
		November	36	32	4	0,5033
		Desember	25	25	0	0
		Total	370	348	22	3,0499

Dari hasil tabel 1 setelah dihitung nilai *gain* dari setiap atribut. Maka hasilnya adalah atribut ‘Bulan Ketidakhadiran’ yang mendapat nilai *gain* tertinggi dengan hasil 0,0532. Bentuk *decision tree* atau pohon keputusannya akan menjadi seperti berikut ini:



Gambar 1. Pohon Keputusan Node Akar

Setelah dihitung nilai *gain* dari setiap atribut. Maka hasil akhirnya adalah atribut ‘Indeks Masa Tubuh’ dengan nilai ‘19-25’ yang mendapat nilai *gain* tertinggi dengan hasil 0,4834 dan ditandai oleh kolom berlatar belakang warna merah pada tabel. Bentuk *decision tree* atau pohon keputusannya akan menjadi seperti berikut ini:



Gambar 2. Pohon Keputusan Node 1.7.2.1

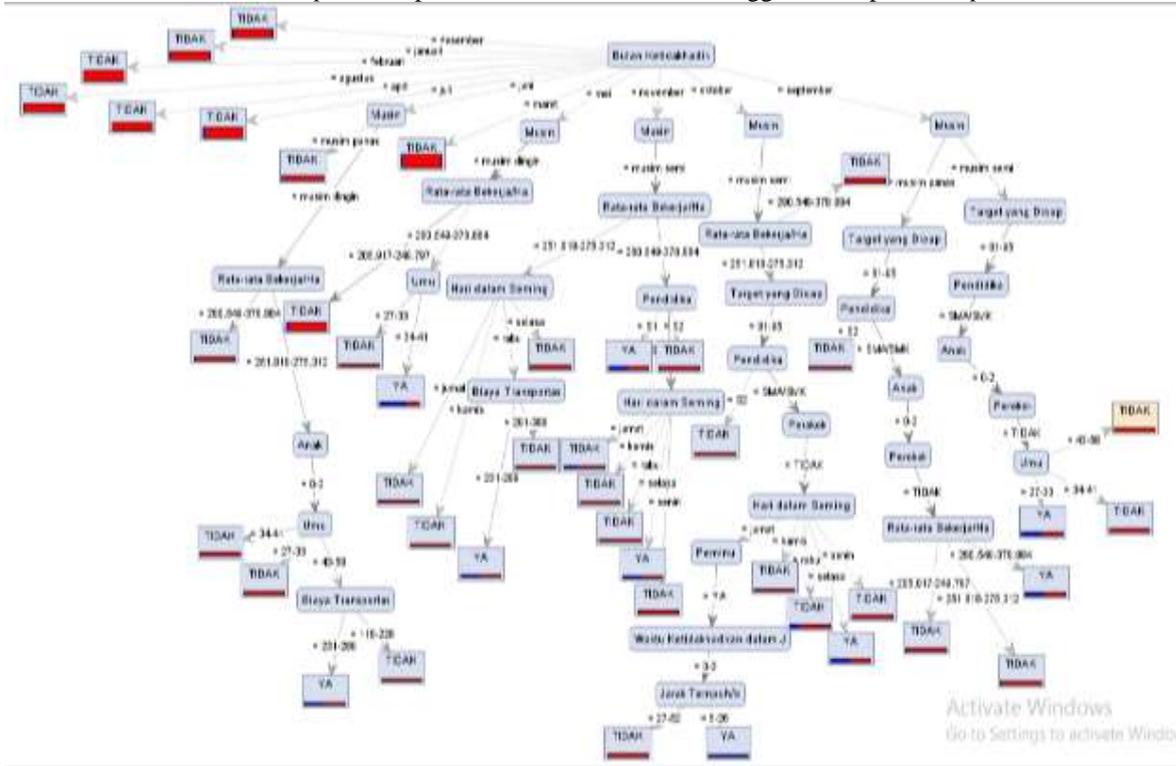
Bentuk aturan IF THEN untuk pohon keputusan dengan algoritma C4.5 adalah sebagai berikut:

1. IF Bulan Ketidakhadiran dengan nilai = Januari THEN Keputusan = Tidak.
2. IF Bulan Ketidakhadiran dengan nilai = Februari THEN Keputusan = Tidak.
3. IF Bulan Ketidakhadiran dengan nilai = April THEN Keputusan = Tidak.
4. IF Bulan Ketidakhadiran dengan nilai = Agustus THEN Keputusan = Tidak.
5. IF Bulan Ketidakhadiran dengan nilai = Desember THEN Keputusan = Tidak.
6. IF Bulan Ketidakhadiran dengan nilai = November AND Alasan Ketidakhadiran = konsultasi kesehatan AND Biaya Transportasi = 291-388 THEN Keputusan = Tidak.

7. *IF* Bulan Ketidakhadiran dengan nilai = November *AND* Alasan Ketidakhadiran = konsultasi kesehatan *AND* Biaya Transportasi = 231-289 *THEN* Keputusan = Tidak.
8. *IF* Bulan Ketidakhadiran dengan nilai = November *AND* Alasan Ketidakhadiran = penyakit pada sistem muskuloskeletal dan jaringan ikat *AND* Indeks Masa Tubuh = 27-38 *THEN* Keputusan = Tidak.
9. *IF* Bulan Ketidakhadiran dengan nilai = November *AND* Alasan Ketidakhadiran = penyakit pada sistem muskuloskeletal dan jaringan ikat *AND* Indeks Masa Tubuh = 19-25 *THEN* Keputusan = Ya.
10. *IF* Bulan Ketidakhadiran dengan nilai = November *AND* Alasan Ketidakhadiran = penyakit pada sistem muskuloskeletal dan jaringan ikat *AND* Indeks Masa Tubuh = 19-25 *THEN* Keputusan = Ya.

2. Implementasi Pohon Keputusan (Decision Tree) dengan *Rapid Miner*.

Berikut ini adalah hasil *pohon* keputusan atau *decision tree* menggunakan aplikasi *rapid miner*.



Gambar 3. Desain Model C4.5

Berdasarkan *rule classification* diatas, untuk mempermudah pemahaman data, maka dibentuk aturan *IF-THEN* yaitu sebagai berikut:

1. *IF* Bulan Ketidakhadiran = September *AND* Musim = Musim Panas *AND* Target yang Dicapai = 81-95 *AND* Pendidikan = S2 *THEN* Keputusan = Tidak *ELSE* Pendidikan = SMA/SMK *AND* Anak = 0-2 *AND* Perokok = TIDAK *AND* Rata-rata Bekerja/Hari = 205.917-249.797 *THEN* Keputusan = Tidak *ELSE* Rata-rata Bekerja/Hari = 251.818-275.312 *THEN* Keputusan = Tidak *ELSE* Rata-rata Bekerja/Hari = 280.549-378.884 *THEN* Keputusan = Ya.
2. *IF* Bulan Ketidakhadiran = September *AND* Musim = Musim Semi *AND* Target yang Dicapai = 81-95 *AND* Pendidikan = SMA/SMK *AND* Anak = 0-2 *AND* Perokok = TIDAK *AND* Umur = 27-33 *THEN* Keputusan = Ya *ELSE* Umur = 34-41 *THEN* Keputusan = Tidak *ELSE* Umur = 43-58 *THEN* Keputusan = Tidak.

Tabel 2. Hasil *Confusion Matrix*.

Accuracy : 81.08%

	True Ya	True Tidak	Class Precision
Pred. Ya	2	19	9.52%
Pred. Tidak	2	88	97.78%
Class Recall	50.00%	82.24%	

Dari hasil evaluasi pada tabel 2. menunjukkan bahwa tingkat akurasi yang dihasilkan adalah sebesar 81,08% atau sebesar 0,8108% yang berarti masuk pada klasifikasi sangat baik untuk hasil klasifikasi pegawai yang mengalami Kegagalan Disiplin = Ya dan Kegagalan Disiplin = Tidak. Berdasarkan hasilnya, dapat dikatakan bahwa pegawai yang mendapatkan Kegagalan Disiplin = Tidak adalah yang terbanyak dibandingkan dengan pegawai yang mendapatkan Kegagalan Disiplin = Ya sehingga akan berdampak baik untuk produktivitas kerja pegawai dan sangat menguntungkan bagi perusahaan tempatnya bekerja.

KESIMPULAN

Berdasarkan hasil penelitian serta pembahasan, maka dapat ditarik beberapa kesimpulan antara lain sebagai berikut:

1. Pemilihan atribut Alasan Ketidakhadiran, Bulan Ketidakhadiran, Hari Dalam Seminggu, Musim, Biaya Transportasi, Jarak Tempuh/Km, Waktu Pelayanan, Umur, Rata-Rata Bekerja/Hari, Target Yang Dicapai, Kegagalan Disiplin, Pendidikan, Anak, Peminum, Perokok, Binatang Peliharaan, Berat Badan, Tinggi Badan, Indeks Masa Tubuh, Dan Waktu Ketidakhadiran Dalam Jam dapat menjadi kriteria penilaian terhadap klasifikasi kegagalan disiplin pegawai yang bernilai “Ya” dan “Tidak”.
2. Penelitian dengan data Ketidakhadiran Pegawai di Tempat Kerja telah dilakukan penelitian menggunakan metode *Neuro Fuzzy Network* yang mana penelitian sebelumnya tersebut menggunakan data yang sama. Penelitian ini berpatokan pada penelitian sebelumnya, namun menggunakan metode yang berbeda, yaitu menggunakan metode Algoritma C4.5.
3. Berdasarkan hasil perhitungan manual dan *rapid miner* ternyata menghasilkan pohon keputusan yang berbeda pada hasil cabang dan daun, namun menghasilkan hasil yang sama pada *Node* akar dan cabang pertama.
4. Sistem yang digunakan dengan *rapid miner* dapat membantu dalam menentukan kriteria data pegawai yang bersifat “Ya, gagal disiplin” atau “Tidak, gagal disiplin”. Sehingga dapat membantu pihak perusahaan dalam pengambilan keputusan terhadap peraturan absensi pegawai.
5. Algoritma C4.5 dapat menghasilkan *rule* atau aturan untuk menggambarkan proses yang terkait dengan kegagalan disiplin.
6. Dari hasil yang diujikan menggunakan *confusion matrix*, di dapatkan hasil bahwa Algoritma *decision tree* C4.5 mengklasifikasikan data dengan tingkat akurasi 81,08% dalam memprediksi jumlah pegawai yang mengalami kegagalan disiplin atau tidak. Nilai *class precision* dan *class recall* di dapatkan hasil masing-masing adalah 9,52% dan 97,78% untuk nilai *class precision* dan 50,00% dan 82,24% untuk nilai *class recall*. Berdasarkan hasil tersebut, dapat diartikan bahwa hasil dari tingkat akurasi sebesar 81,08% mendapatkan hasil yang sangat baik untuk mengklasifikasi Ketidakhadiran pegawai dalam mendapatkan prediksi kegagalan disiplin pegawai.

REFERENSI

- Ardyan Pratama, H. L. (2016). Implementasi Data Mining Menggunakan Metode Deskripsi Untuk Menemukan Pola Absensi Di Badan Kepegawaian Daerah Kota Semarang. *Data Mining*, 15(3), 207–216.
- Christiana, M. B. (2013). Organizational Behaviour. In *Organizational Behaviour*. Educreation Publishing. <https://doi.org/10.4324/9780203857595>
- Hermawati, F. A. (2005). Data Mining Data mining. In *Mining of Massive Datasets* (Vol. 2, Issue January 2013). Andi Offset. https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part
- Ichsan, Syahrul, T. . (2013). Sistem Pendukung Keputusan Pemilihan Penerima Beasiswa Mahasiswa Kurang Mampu Pada Stmik Budidarma Medan Menerapkan Metode Profile Matching. *Pelita Informatika Budi Darma*, 5(1), 2.
- M Anbarasi, E Anupriya, N. C. S. N. I. (2010). Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm. *International Journal of Engineering Science and Technology*, 2(10), 5370–5376.
- Mutoi Siregar, A., & Puspabhuana, A. (2018). *Data Mining: Pengelolaan Data Menjadi Informasi dengan*

RapidMiner (1st ed.). CV. Kekata Group.
<https://books.google.co.id/books?id=rTImDwAAQBAJ&printsec=frontcover&hl=id#v=onepage&q&f=false>

Nofriansyah, D. (2014). *Konsep Data Mining Vs Sistem Pendukung Keputusan* (1st ed.). DEEPUBLISH.

Wahbeh, dkk. (2011). A Comparison Study between Data Mining Tools over some Classification Methods. *International Journal of Advanced Computer Science and Applications*, 1(3).
<https://doi.org/10.14569/specialissue.2011.010304>

Widodo, P. P., Handayanto, R. T., & Herlawati. (2013). Penerapan Data Mining dengan Matlab. In *Rekayasa Sains*. Rekayasa Sains. www.biobses.com

Yuli Mardi. (2014). Analisa Data Rekam Medis Untuk Menentukan Penyakit Terbanyak Berdasarkan International Clasification Of Disease (ICD) menggunakan Decesion Tree C.45 (Studi Kasus RSUD. CBMC Padang. *UPI YPTK Padang*, 213–219.