

Klasifikasi Ketidakhadiran Karyawan Menggunakan Metode Algoritma Decision Tree C4.5

¹Rita Sarina Siburian,²Oktafanny Rozalia,³Pina Alpianita,⁴Aulia Agung Dermawan

^{1,2,3,4}Program Studi Manajemen Rekayasa, / fakultas teknologi industri/
institut teknologi batam/, The Vitka City Complex, Tiban, Jl. Gajah Mada, Kota Batam, Kepulauan Riau
29425, /indonesia

e-mail: ¹2212033@student.iteba.ac.id, ²2212005@student.iteba.ac.id, ³2212031@student.iteba.ac.id,
⁴agung.dermawan29@gmail.com

Abstract - Not every employee's absence has a negative connotation attached to it. Additionally, there are certain permits that are deemed appropriate and have a certain duration that won't make it too difficult for the parties to work, or there are absences that can be tolerated by an organization or firm if they have a valid reason such as illness. The purpose of the study is to determine how much the C4.5 algorithm can aid in the categorization computation process in order to identify a way to boost worker productivity Using Algorithm C4.5 to classify data mining is one of the approaches and methods needed to identify solutions that will increase employee productivity in the workplace. For optimal results, data processing is carried out using Rapid Miner 5 Tools. One type of data mining classification rule is the C4.5 algorithm. After the test, Algorithm C4.5 yielded results with 60% accuracy.

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

Abstrak- Tidak setiap karyawan absen memiliki konotasi negatif yang melekat padanya. Selain itu, ada izin tertentu yang dianggap tepat dan memiliki jangka waktu tertentu yang tidak akan membuat pihak terlalu sulit untuk bekerja, atau ada absen yang dapat ditoleransi oleh organisasi atau perusahaan jika mereka memiliki alasan yang sah - seperti sakit. Tujuan dari studi ini adalah untuk menentukan berapa banyak algoritma C4.5 dapat membantu dalam proses komputasi kategorisasi untuk mengidentifikasi cara untuk meningkatkan produktivitas pekerja Menggunakan Algoritma C4.5 untuk mengklasifikasikan penambahan data adalah salah satu pendekatan dan metode yang diperlukan untuk mengidentifikasi solusi yang akan meningkatkan produktivitas karyawan di tempat kerja. Untuk hasil yang optimal, pengolahan data dilakukan dengan menggunakan Rapid Miner 5 Tools. Salah satu jenis aturan klasifikasi penambahan data adalah algoritma C4.5. Setelah pengujian, Algoritma C4.5 menghasilkan hasil dengan akurasi 60%.

Kata kunci : Ketidakhadiran, Algoritma C4.5, Data Mining, rapid miner.

Diterima : Juni 2024
Disetujui : Juni 2024
Dipublikasi : Juni 2024

Pendahuluan

Kekurangan atau kelalaian pekerjaan atau kewajiban seseorang, serta yang sudah menjadi kewajiban setiap orang dan dilakukan secara teratur. Ketidakhadiran, secara umum, adalah komponen dari kinerja individu yang diklasifikasikan sebagai kemalasan dan termasuk pelanggaran formalitas pekerjaan di dalam organisasi atau perusahaan oleh bawahan terhadap atasan mereka. Tidak adanya pekerjaan secara teratur adalah pelanggaran moral yang dikecam oleh

masyarakat. Penambangan data adalah metode yang menganalisis sejumlah besar data untuk menemukan pola atau hubungan yang dapat digunakan untuk mengekstrak informasi yang berguna. Penambangan data dapat digunakan untuk mengungkapkan informasi yang dapat digunakan oleh para profesional medis untuk meningkatkan standar perawatan. Para profesional medis juga dapat memanfaatkan informasi tersebut untuk menyediakan pilihan perawatan yang lebih murah yang sebanding dengan pilihan yang berbeda sambil meminimalkan jumlah reaksi obat yang merugikan. Analisis data, yang mencakup asosiasi, prediksi, pengelompokan, dan klasifikasi, hanya salah satu aspek dari proses penambangan data (2010) Anbarasi, M., Anupriya, E. Tingkat ketidakhadiran karyawan yang terkait dengan penambangan data adalah tidak adanya pekerja yang bertanggung jawab. Banyak pekerja dewasa ini tidak menaati jam kerja; namun, berpaut pada hukum jam kerja dan melapor untuk bertugas adalah kewajiban pekerjaan. Pratama, Ardyana(2016).

Menurut Widodo dkk. (2013), pohon keputusan banyak digunakan untuk pengambilan keputusan kasus dan merupakan salah satu pendekatan penambangan data yang paling terkenal. Hal ini disebabkan oleh fakta bahwa pendekatan ini dapat menyelesaikan instans langsung dengan banyak dimensi tanpa perlu prosedur manajemen pengetahuan awal. Selama data yang di benchmark akurat, akurasi yang luar biasa. Dalam (Widodo et al., 2013), Menurut Han, "Algoritme C4.5 adalah struktur pohon di mana setiap daun mewakili kelas, setiap cabang mewakili hasil dari kualitas yang diuji, dan tidak ada node yang menunjukkan atribut Penelitian. Setiap node keputusan terlihat iteratif oleh algoritma C4.5, yang memilih pembagian terbaik sampai tidak dapat dibagi lagi. Algoritme C4.5 memilih pembagian terbaik dengan memanfaatkan ideal keuntungan informasi atau pengurangan entropi.

Pembentukan organisasi ini dimotivasi oleh kebutuhan masyarakat. Organisasi harus berorientasi pada pertumbuhan di dunia kompetitif yang kita tinggali saat ini. Ketika produktivitas diverifikasi sehubungan dengan jumlah dan kualitas produk, hal ini dapat dilakukan. Absen dan penggantian karyawan memiliki efek merugikan pada output. Ketidakhadiran karyawan sering menghambat kemampuan organisasi untuk beroperasi secara efisien dan berkembang. Dengan cara yang sama, biaya produksi yang lebih tinggi dihasilkan dari pergantian personel (Christiana, 2013).

Untuk membuat pohon keputusan, algoritma pohon keputusan C4.5 secara iteratif mengunjungi setiap node keputusan dan memilih split terbaik sampai tidak ada lagi split yang diizinkan. Ini adalah tambahan Quinlan untuk teknik ID3 (Larose, 2005). Ide dasar algoritma pohon keputusan C4.5 adalah untuk mengubah data menjadi pohon keputusan dan aturan keputusan. Teknik yang efektif untuk masalah penambangan data dan klasifikasi adalah algoritme pohon keputusan C4.5. Berlaku untuk klasifikasi baru, algoritma pohon keputusan C4.5 mengklasifikasikan nilai atribut (Xindong, 2009).

Absensi adalah daftar orang dari sekelompok orang yang tergabung dalam sebuah instansi resmi dengan peraturan, pembatasan, dan pedoman. dan siapa pun yang berpartisipasi di dalamnya tunduk pada aturan ini. Sebaliknya, absensi, sebagaimana dimaksud oleh Hasibuan (2001), adalah tidak adanya karyawan yang bekerja pada hari kerja karena adanya izin, sakit, alpa, atau cuti. Akibatnya, tidak adanya karyawan dari pekerjaan pada hari kerja dan untuk jumlah waktu yang telah ditentukan sebelumnya untuk alasan seperti sakit, otorisasi, cuti, atau alpa dikenal sebagai absensi (Juhartini, 2020).

Tujuan penelitian ini adalah untuk memastikan tingkat bantuan algoritma C4.5 dalam perhitungan klasifikasi untuk mengidentifikasi cara untuk meningkatkan produktivitas pekerja dan untuk memastikan temuan dan analisis berdasarkan data yang digunakan Rapid Miner.

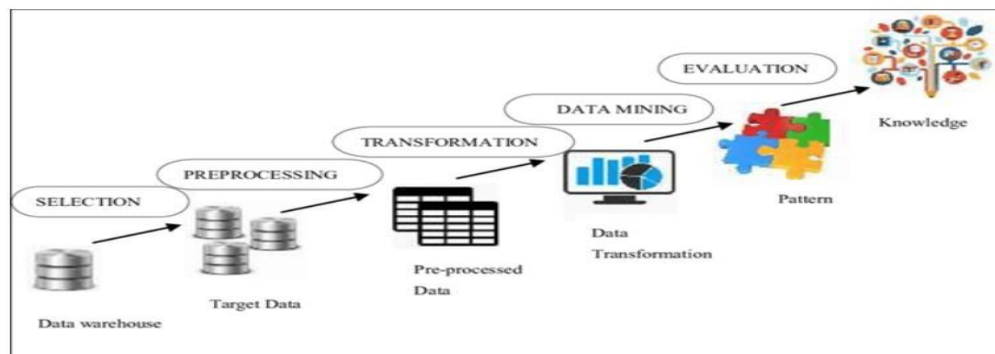
Mampu mengenali tren atau elemen yang menyebabkan ketidakhadiran pekerja sehingga dapat dilakukan tindakan atau pengaturan yang sesuai. Strategi atau metode pemrosesan data diperlukan untuk memproses data menjadi informasi yang diperlukan dan relevan ketika datang ke data yang disimpan dan tidak digunakan lagi, dengan dataset yang cukup besar sekitar 100 set data.

Metode Penelitian

Klasifikasi berasal dari kata Latin "classis". Tindakan klasifikasi memerlukan pengumpulan objek atau entitas yang sebanding dan membagi tubuh atau makhluk yang berbeda. Batasan klasifikasi secara luas didefinisikan sebagai upaya untuk mengatur alam semesta pengetahuan secara metodis, klasifikasi adalah pengaturan metodis dari berbagai konsep, buku, item, atau objek lain ke dalam kelas atau kelompok tertentu berdasarkan karakteristik bersama. Dalam kehidupan sehari-hari, klasifikasi adalah sesuatu yang telah banyak dilakukan manusia. Sebagai contoh, perhatikan pengecer yang mengelompokkan produk serupa di pasar, supermarket, atau toko buku. Tujuan dari hal ini adalah untuk membuatnya lebih sederhana bagi pelanggan untuk memilih persyaratan penting. Sebuah Pengantar Ilmu Perpustakaan, Sulisty-Basuki, 1991.

- **Data Mining**

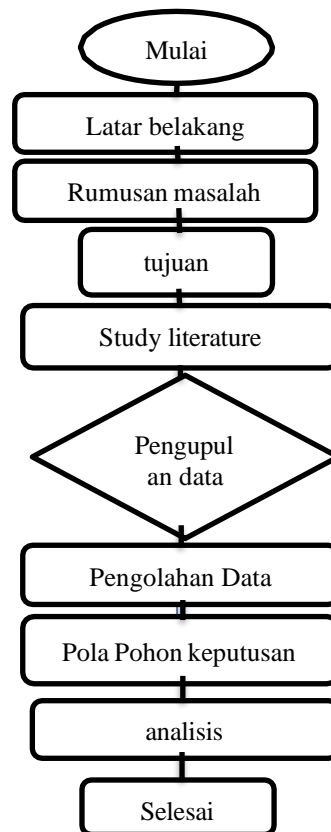
Data mining adalah suatu penambangan data pemeriksaan observasional terhadap kumpulan data untuk mengungkap asosiasi tak terduga dan menyajikan data secara segar dan ramah pengguna bagi pemilik data Menurut Larose (2006). penambangan data adalah prosedur semi-otomatis yang mengambil dan mengidentifikasi bahan pengetahuan yang berpotensi berharga dari data besar dengan menerapkan pendekatan statistik, matematika, kecerdasan buatan, dan pembelajaran mesin Menurut Turban (2005),



Gambar 1. KDD

- **Flowchart**

Untuk tiba di bagian yang ditunjuk dari penyelidikan ini, diperlukan berbagai langkah. Flowchart adalah penggambaran visual dari serangkaian tindakan atau proses yang ditunjukkan dengan simbol grafis. Flowchart atau Diagram alir digunakan untuk secara visual mewakili alur kerja atau metode sistem. Tahap-tahap dari pratikum ini adalah sebagai berikut:



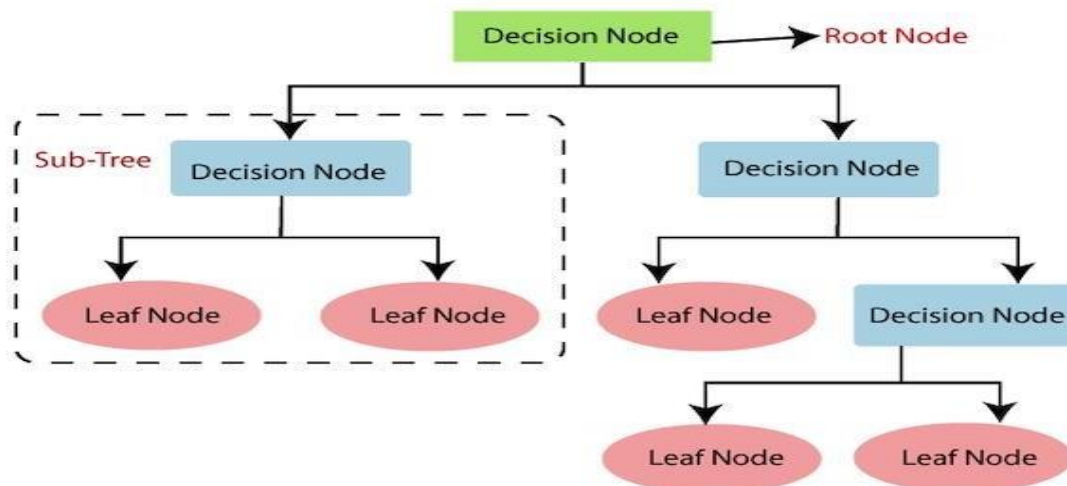
Gambar 2. Flowchart

- **Decision tree**

Decision tree adalah Pendekatan pembelajaran mesin yang diawasi disebut pohon keputusan digunakan untuk mengklasifikasikan atau meramalkan serangkaian pertanyaan atau tugas yang belum terjawab. Karena pohon keputusan dapat memilih fitur optimal untuk mengoptimalkan akuisisi informasi untuk pengambilan keputusan, mereka terdaftar sebagai beberapa algoritma yang paling kuat. Selain itu, untuk melatih subset data pelatihan variabel dan menghasilkan hasil yang lebih akurat, hutan acak menggunakan pendekatan ini.

Untuk membuat pohon keputusan, algoritma pohon keputusan C4.5 secara iteratif mengunjungi setiap node keputusan dan memilih split terbaik sampai tidak ada lagi split yang diizinkan. Ini adalah tambahan Quinlan untuk teknik ID3 (Larose, 2005). Ide dasar algoritma pohon keputusan C4.5 adalah untuk mengubah data menjadi pohon keputusan dan aturan keputusan. Teknik yang efektif untuk masalah penambahan data dan klasifikasi adalah algoritme pohon keputusan C4.5. Berlaku untuk klasifikasi baru, algoritma pohon keputusan C4.5 mengklasifikasikan nilai atribut (Xindong, 2009).

Terdapat atribut dalam data klasifikasi ketidakhadiran karyawan yaitu nama, emp, tanggal absen, keterangan dan usia. Dalam penelitian ini, kami mengumpulkan datanya adalah berupa data sekunder. Data sekunder yang digunakan di dalam penelitian ini adalah menggunakan data dari salah satu Perusahaan di Batam.



Gambar 3. Decision tree

- **Perhitungan Manual Algoritma C4.5.**

Berikut ini adalah proses yang terlibat dalam prosedur perhitungan manual untuk menentukan pohon keputusan (Pohon Keputusan) menggunakan algoritma C4.5:

- Mengumpulkan dan menyiapkan data
- Menemukan Nilai Entropi.
- Menentukan Nilai Gain untuk setiap properti, mencari nilai tertinggi untuk ditetapkan sebagai Node Akar.
- Prosedur untuk menghitung nilai pada node cabang dan root.

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

Pada tahap aplikasi Rapid Miner digunakan untuk melakukan prosedur penghitungan dengan tujuan membandingkan hasil perhitungan aplikasi dengan hasil komputasi manual. Tujuan pengujian adalah untuk mendeteksi kebenaran data dan membuat referensi dari temuan.

Hasil dan Pembahasan

Berikut ini adalah data yang sudah di kumpulkan serta sumber data pada pratikum ini adalah ketidakhadiran karyawan di sebuah Perusahaan, data yang digunakan yaitu 100 data yang dibagi menjadi 2 sampel data, yaitu data sampel pertama 70 data dan data sampel kedua 30 data dengan menggunakan metode klasifikasi:

Tabel 1. Data ke 1

NO	NAMA	Emp. No.	JENIS KELAMIN	TANGGAL ABSEN	KETERANGAN
1	RUDI PARNINGTON PURBA	7971	L	14-Jun-24	SAKIT
2	ALFIAN YAKOP SITOANG	8041	L	12-Jun-24	SAKIT
3	RIZKI AMRIADI	8488	L	18-Jun-24	CUTI
4	DEBY MELIANA SIHOMBING	8581	P	11-Jun-24	CUTI
5	MEMORI PUTRI MENDROFA	8993	P	24-May-24	SAKIT
6	YOKO A. MANURUNG	9329	L	18-Jun-24	CUTI
7	ABDUSSY AKUR	9959	L	18-Jun-24	SAKIT
8	DIAN RAHMAWATI	9968	P	22-May-24	CUTI
9	JUPRIONO PURBA	10027	L	21-May-24	CUTI
10	DEWIN SIMANUNGKALIT	10046	L	12-Jun-24	SAKIT
11	DAVID PARDAMEAN SIANIPAR	10073	L	30-May-24	SAKIT
12	KAMSINA INA LIPAT	10089	P	18-Jun-24	CUTI
13	VERA KRISDAYATI HUTASOIT	10111	P	13-Jun-24	SAKIT
14	DANIEL ANTO SIMAMORA	10177	L	03-Jun-24	CUTI
15	ERNI JULITA	10246	P	13-Jun-24	SAKIT
16	LIEN RIKA SINAGA	10325	P	28-May-24	CUTI
17	MARDIANTO	10327	L	21-May-24	CUTI
18	NOVA P SITUMORANG	10388	P	30-May-24	SAKIT
19	ILGA AFDIRAHAYU	10442	P	28-May-24	SAKIT
20	SAMUEL SIREGAR	10469	L	03-Jun-24	CUTI
21	ASTUTI DOHONA	10480	P	12-Jun-24	SAKIT
22	REGI G SYAHPUTRA	10502	L	05-Jun-24	SAKIT
23	DEA S RAMADANI PUTRI	10517	P	04-Jun-24	CUTI
24	NORA MAGDALENA	10587	P	14-Jun-24	CUTI
25	DENI PASARIBU	10625	L	18-Jun-24	SAKIT
26	RIZKI SYAHPUTRA	10698	L	03-Jun-24	CUTI
27	ARNETA	10711	P	18-Jun-24	CUTI
28	FIRNANDA RAMADANI	10743	P	13-Jun-24	SAKIT
29	IAN WIDIANTO	10748	L	06-Jun-24	CUTI
30	HAMIDAH ROHIMAH LUBIS	10777	P	30-May-24	SAKIT
31	NIKA ROSANTA SIBURIAN	10856	P	10-Jun-24	SAKIT
32	RATA K SIMANGUNSONG	10859	P	13-Jun-24	CUTI
33	IRLIANA OKTAVANI IDRIS	10912	P	29-May-24	CUTI
34	RIANI SIMANJUTAK	11019	P	18-Jun-24	SAKIT
35	VICO DAMARA	11106	L	31-May-24	SAKIT
36	ELYSON S SIMANGUNSONG	11205	L	11-Jun-24	SAKIT

37	SHARON NATAL SIMAMORA	11244	P	13-Jun-24	CUTI
38	ZUHRIANNUR RANGKUTI	11324	P	28-May-24	SAKIT
39	IWAN MARTUA TARIHORAN	11386	L	21-May-24	CUTI
40	RIDHO M PANGABEAN	11396	L	24-May-24	SAKIT
41	DOMINGGO LUMBAN TORUAN	11847	L	14-Jun-24	CUTI
42	DELIMA DAMERIA PANJAITAN	11990	P	12-Jun-24	SAKIT
43	JOHANES NATHAN	11997	L	18-Jun-24	SAKIT
44	MONICA MIERANTY	12155	P	12-Jun-24	CUTI
45	MARIA ANGELY LBN GAOL	12197	P	30-May-24	CUTI
46	M. NUR HIIDAYAT	12251	L	03-Jun-24	CUTI
47	JEFRI	12333	L	18-Jun-24	SAKIT
48	ENDANG SAPRIKA TAMPUBOLON	12346	P	13-Jun-24	CUTI
49	ERINA YULDRI YANI SIMANJORANG	12347	P	12-Jun-24	CUTI
50	LASTIURMA LB TOBING	12356	P	03-Jun-24	SAKIT
51	OLIJAH TANJUNG	12362	P	07-Jun-24	CUTI
52	PANI IDA KUSUMA	12363	P	13-Jun-24	SAKIT
53	LESTINA TURNIP	12400	P	24-May-24	SAKIT
54	MAHDIN SIHOTANG	12402	L	04-Jun-24	CUTI
55	MELYANA BR SARAGIH	12405	P	31-May-24	CUTI
56	SUARMAN GEA	12418	P	14-Jun-24	SAKIT
57	ANNISA SHINDI NUUR AINII	12440	P	18-Jun-24	SAKIT
58	JOISCE LARASATI SIMANUNGKILAT	12450	P	18-Jun-24	SAKIT
59	MINYIK MUNTE	12454	P	18-Jun-24	SAKIT
60	MUHAMMAD ARKAN TAMIR	12455	L	22-May-24	CUTI
61	MUHAMMAD HABIBI	12456	L	18-Jun-24	SAKIT
62	FANDI HARDIKA	12497	L	14-Jun-24	CUTI
63	JUVENTIA ARDANA PUTRI	12504	P	12-Jun-24	CUTI
64	RISKI DESRI WANDI	12518	L	10-Jun-24	CUTI
65	RIZKY ARYA PRATAMA	12521	L	22-May-24	SAKIT
66	SERLINA TAMBUNAN	12526	P	18-Jun-24	CUTI
67	AGUSTINA FITRIA SIREGAR	12542	P	18-Jun-24	SAKIT
68	ARION L MAULANA PARDEDE	12546	L	31-May-24	SAKIT
69	DAVID BEKHAM IRAMAN NAZARA	12551	L	05-Jun-24	CUTI
70	MARIA PORENSIA YUNITA	12568	P	18-Jun-24	CUTI

Berikut ini adalah data ke 2 dari data yang kami ambil, yang digunakan sebagai perbandingan untuk data pertama.

Tabel 2. Data ke 2

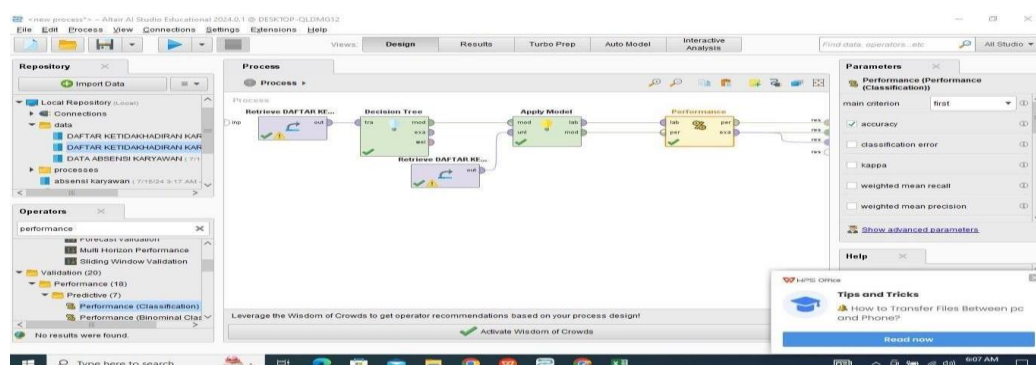
NO	NAMA	Emp. No.	JENIS KELAMIN	TANGGAL ABSEN	KETERANGAN
1	MICELLA PUTRI SARI	12571	P	03-Jun-24	SAKIT
2	PERIANTO SITANGGANG	12574	L	21-May-24	SAKIT
3	RANGGA PUTRA HERMAWAN	12576	L	30-May-24	CUTI
4	RIYANDA PUTRA SIREGAR	12580	L	06-Jun-24	CUTI
5	RIZKI ULINA PERANGIN-ANGIN	122581	L	18-Jun-24	SAKIT
6	ROBERTO ISKANDAR TAMBA	12582	L	18-Jun-24	CUTI
7	SHIFA MELLYANI	12587	P	05-Jun-24	SAKIT
8	SIDETIAN	12588	P	18-Jun-24	CUTI
9	SISKA PUTRI	12590	P	19-Jun-24	CUTI
10	SUSI INDRIYANTI IRMA	12592	P	06-Jun-24	SAKIT
11	TRIDANINGSIH SEPRIANTI	12593	P	21-May-24	SAKIT
12	WILLIE SINAGA	12628	L	18-Jun-24	CUTI
13	AISYAH MUSTAPA ALIA	12635	P	14-Jun-24	SAKIT
14	ASTUR SALIM SIMANJUTAK	12638	L	04-Jun-24	CUTI
15	FIFI FITRIASARI	12652	P	19-Jun-24	SAKIT
16	MARIO PENANDRE	12664	L	21-May-24	CUTI
17	RAMA HIDAYAT	12681	L	18-Jun-24	CUTI
18	SYIFA SYAHRANI	12697	P	18-Jun-24	SAKIT
19	SEKAR DWIJAYANI	12758	P	18-Jun-24	SAKIT
20	GINO GILANG PERMANA	12838	L	14-Jun-24	CUTI
21	HASAN NUDDIN SIREGAR	12842	L	11-Jun-24	SAKIT
22	NENGSIH	12899	P	04-Jun-24	SAKIT
23	SAIFUL ALAMSAH	12943	L	13-Jun-24	CUTI
24	TUTI LESTARI NAZARA	12973	P	14-Jun-24	CUTI
25	YUNI LESTARI	12987	P	12-Jun-24	SAKIT
26	AHMAD SYAFIDI	12989	L	18-Jun-24	CUTI
27	BUE SITOMPUL	12998	L	13-Jun-24	CUTI
28	DOLA MILEN	13007	P	14-Jun-24	SAKIT
29	WINDA SIAHAAN	13077	P	10-Jun-24	CUTI
30	RESTIKA ISTIFANY	13135	P	18-Jun-24	SAKIT

Tujuan dari penelitian ini adalah untuk menentukan berapa banyak algoritma C4.5 dapat membantu dalam komputasi klasifikasi untuk mengidentifikasi solusi yang akan meningkatkan produktivitas pekerja. 100 data digunakan dalam kumpulan data oleh para peneliti untuk menentukan kesimpulan penelitian. Untuk memusatkan penggambaran pohon keputusan selanjutnya pada tingkat ketidakhadiran karyawan, berbagai perubahan dilakukan pada parameter selama proses data Rapidminer. Misalnya, atribut pilihan peran operator diubah menjadi "KET persentase IPM" dan peran target diubah menjadi "label".

Hasil dari pembahasan klasifikasi ketidakhadiran karyawan menggunakan RapidMiner dengan metode Decision Tree menunjukkan akurasi sebesar 53.33%. Data yang digunakan terdiri dari 740 dataset yang dibagi menjadi dua sampel, yaitu 70 dan 30 data. Precision untuk kategori 'Sakit' adalah 52.00% dan untuk kategori 'Cuti' adalah 60.00%, dengan recall masing-masing 86.67% dan 20.00%. Penggambaran pohon keputusan dengan algoritma C4.5 memperlihatkan klasifikasi ketidakhadiran berdasarkan tanggal absen dan beberapa kriteria lainnya. Performa model yang kurang dari 60% menunjukkan bahwa model ini tidak cukup akurat, kemungkinan disebabkan oleh kurangnya atribut dalam data yang digunakan, sehingga memerlukan perbaikan lebih lanjut. dengan hasil pengolahan data rapid miner ketidakhadiran karyawan dengan metode klasifikasi. Data yang digunakan yaitu 100 data yang dibagi menjadi 2 sampel data, yaitu data sampel pertama 70 data dan data sampel kedua 30 data. Dari hasil pengolahan data rapid miner pada parameter decision tree digunakan criteria information gain. Dari hasil pengolahan data rapid miner dengan klasifikasi di dapatkan hasil akurasi yaitu 53.33%. Dimana hasil yang kami dapatkan tidak akurat yang kemungkinan disebabkan kurang atribut pada data yang kami olah.

Hasil

Berikut adalah penjelasan hasil kerja menggunakan RapidMiner berdasarkan gambar pohon keputusan yang diberikan dan asumsi dari analisis ketidak hadiran karyawan:



Gambar 4. Operator

Pada gambar diatas decision tree di pergunakan untuk eksplorasi data atau menemukan hubungan antara variabel input dengan variabel target. Pada apply model pada rapitminer menggunakan untuk menerapkan model yang telah di latih sebelumnya menggunakan data testing. Performance pada rapidminer digunakan sebagai evaluasi kinerja model untuk memberikan daftar nilai kriteria secara otomatis sesuai dengan tugas yang diberikan.

- a. Hasil data split dalam studi klasifikasi pendekatan RapidMiner dengan Decision Tree tentang ketidakhadiran karyawan menunjukkan akurasi model 53,33%. 52,00% presisi dan 60,00% presisi untuk kategori "Put" dan "Pain", masing-masing, dengan 86,67% dan 20,00% recall. Kinerja model kurang dari 60% menyiratkan bahwa data yang digunakan dalam model mungkin kurang dalam fitur tertentu, yang membuat model tidak cukup akurat.
- b. Model klasifikasi absen karyawan pendekatan RapidMiner with Decision Tree dilakukan dengan akurasi 53,33%. 52,00% presisi dan 60,00% presisi untuk kategori "Put" dan "Pain", masing-masing, dengan 86,67% dan 20,00% recall. Jika kinerja model kurang dari 60%, itu tidak cukup tepat dan harus ditingkatkan. Dengan menggunakan informasi ini, label fokus dengan peringkat kinerja berikutnya diproduksi.
- c. Digunakan untuk menerapkan model pra-pelatihan pada data tanpa label (data testing) dengan menggunakan data pelatihan.
- d. Untuk mengklasifikasikan tingkat ketidakhadiran karyawan berdasarkan pohon keputusan dalam gambar yang digunakan untuk menilai apakah pohon keputusan ada atau tidak ada, hasil data dalam pohon keputusan dengan deskripsi pohon keputusan dihasilkan menggunakan algoritma C4.5. mengkategorikan absen karyawan berdasarkan tanggal absen dan sejumlah parameter tambahan. Setiap node adalah kondisi yang memisahkan data menjadi cabang sesuai standar tertentu. Ketidakhadiran karyawan pada akhirnya diklasifikasikan sebagai penyakit atau cuti, seperti yang ditunjukkan oleh hasil akhir pada daun pohon keputusan.



Gambar 5. Hasil decision tree.

Pada proses pengolahan data pada RapidMiner, dilakukan perubahan pada parameternya, seperti operators set role diubah select atribut nya menjadi “KET presentase IPM” dan target role menjadi “label”. Hal ini dilakukan untuk memfokuskan penggambaran Decision Tree nantinya dalam angka ketidakhadiran karyawan. Pada pengolahan data RapidMiner ketidakhadiran karyawan dengan metode klasifikasi, data yang digunakan yaitu 100 data yang dibagi menjadi 2 sampel data, yaitu data sampel pertama 70 data dan data sampel kedua 30 data. Dari hasil pengolahan data rapid miner pada parameter decision tree digunakan criteria information gain. Dari hasil pengolahan data rapid miner dengan klasifikasi di dapatkan hasil akurasi yaitu 53.33%. Dimana hasil yang kami dapatkan tidak akurat yang kemungkinan disebabkan kurang atribut pada data yang kami olah.

	true SAKIT	true CUTI	class precision
pred. SAKIT	13	12	52.00%
pred. CUTI	2	3	60.00%
class recall	86.67%	20.00%	

Gambar 6. Hasil performance

Hasil performa model klasifikasi ketidakhadiran karyawan menggunakan RapidMiner dengan harus metode Decision Tree dapat diketahui bahwa precision untuk kategori 'Sakit' adalah 52.00% dan untuk kategori 'Cuti' adalah 60.00%, dengan recall masing-masing 86.67% dan 20.00%. Performa model yang kurang dari 60% menunjukkan bahwa model ini tidak cukup akurat dan memerlukan perbaikan lebih lanjut.

Pembahasan

Klasifikasi pada data ketidakhadiran karyawan dengan RapidMiner menunjukkan bahwa dengan menggunakan metode algoritma C4.5 dapat memprediksi ketidakhadiran karyawan. Dengan menganalisis Decision Tree (pohon keputusan) dapat diketahui tingkat akurasi kehadiran karyawan. Dengan hasil Decision Tree (pohon keputusan) dari penelitian ini adalah bahwa meskipun metode Decision Tree dapat mengidentifikasi kriteria utama yang mempengaruhi ketidakhadiran karyawan, model ini memerlukan perbaikan lebih lanjut untuk meningkatkan akurasi dan analisis yang lebih mendalam untuk memahami faktor-faktor yang berkontribusi terhadap ketidakhadiran. Namun, performansi yang kami hasil dari model ini kurang dari 60% sehingga tidak dapat diandalkan untuk membuat prediksi yang akurat. Hal ini kemungkinan disebabkan oleh kurangnya atribut pada data yang digunakan. Untuk mendapatkan hasil yang lebih akurat, perlu dilakukan analisis ulang dengan memperbaiki model atau menggunakan metode klasifikasi lain yang lebih efektif. RapidMiner digunakan untuk pengolahan data, dan pendekatan klasifikasi digunakan. Hasil berikut dicapai dari pemrosesan data yang dilakukan menggunakan algoritma C4.5 dan metode Pohon Keputusan: presisi model: 53,33 persen 'Pain' kategori presisi: 52,00% 'Pain' kategori presisi: 60,00% 'Pain' kategori recall rate: 86,67% 20,00% recall untuk kategori "Pain" Namun, kinerja model ini kurang dari 60%, yang berarti bahwa itu tidak dapat diandalkan untuk menghasilkan perkiraan yang tepat. Tidak adanya atribut dalam data yang digunakan mungkin menjadi penyebabnya. Hal ini diperlukan untuk menganalisis ulang dengan meningkatkan model atau menggunakan teknik kategorisasi lain yang lebih efisien untuk mendapatkan temuan yang lebih akurat.

Kesimpulan

1. Menerapkan Metode Analitika Data: bisa menerapkan metode analitika data untuk mengklasifikasikan ketidakhadiran karyawan di Kota Batam. Hal ini menunjukkan bahwa teknik analitika data efektif dalam menganalisis pola ketidakhadiran.
2. Menentukan faktor Utama: memiliki kemampuan untuk menentukan faktor utama yang mempengaruhi bagaimana absen karyawan diklasifikasikan di Kota Batam. Kriteria ini membantu dalam pemahaman variabel yang paling mempengaruhi ketidakhadiran. Untuk melakukan pemeriksaan lebih menyeluruh terhadap elemen yang menyebabkan ketidakhadiran, identifikasi elemen Penyebab. Misalnya, menentukan apakah karakteristik

3. demografi seperti usia, jenis kelamin, penyakit, atau cuti berhubungan dengan ketidakhadiran.
4. Metode C4.5 menawarkan pemeriksaan menyeluruh terhadap variabel yang mempengaruhi kinerja, sehingga berguna untuk perhitungan kategorisasi yang meningkatkan produktivitas pekerja.. Dengan membangun pohon keputusan yang mengklasifikasikan data berdasarkan atribut penting, C4.5 membantu dalam mengidentifikasi faktor kunci dan pola yang berdampak pada produktivitas. Hasilnya, organisasi dapat membuat keputusan yang lebih terinformasi dan strategis untuk mengoptimalkan proses kerja dan menerapkan solusi yang meningkatkan kinerja pegawai secara keseluruhan.

Daftar Pustaka

- 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
- Sulistyo-Basuki. 1991. Pengantar Ilmu Perpustakaan. Jakarta: Gramedia Pustaka Utama.
- Christiana, M. B. (2013). Organizational Behaviour. In *Organizational Behaviour*. Educreation Publishing. <https://doi.org/10.4324/9780203857595>
- Hasibuan, Malayu S.P. (2001). Manajemen Sumber Daya Manusia. Jakarta PT. Bumi Aksara)
- Juhartini. (2020). Sistem Informasi Absensi Siswa Menggunakan Php Dan Mysql Berbasis Web Pada Man 2 Unggulan Mataram. *Explore*, 10(1), 60-64.
- Larose, D. T. (2005). *Discovering knowledge in data*. A John Wiley & Sons, Inc. New Jersey.
- 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.
- Particle Swarm Optimization: Application and Resources. IEEE. Developments. Efrain Turban, J. F. A. & T.-P. I.. (2005). *Decision Support Systems and Intelligent*
- Widodo, P. P., Handayanto, R. T., & Herlawati. (2013). Penerapan Data Mining dengan Matlab. In *Rekayasa Sains*. Rekayasa Sains. www.biobses.com
- Xindong, W. K. V. (2009). *The Top Ten Algorithms in Data Mining*. USA: Taylor & Francis Group, LLC.