



Pengujian Akurasi Data Potensi Kepuasan Pelanggan Kereta Commuterline (KRL) Dengan Algoritma C4.5

Muhammad Fatchan¹, Putri Anggun Sari², Alfiyan³, Sri Nuryati⁴, Irfan Afriantoro⁵,
Endah Yaodah K⁶

Program Studi Teknik Informatika, Universitas Pelita Bangsa
Jl. Inspeksi Kalimalang Tegal Danas Arah DELTAMAS, Cikarang Pusat-Kab. Bekasi,
Indonesia

fatchan@pelitaangsa.ac.id

Abstrak	Informasi Artikel
<p><i>PT. KAI Commuter line (KRL) is a transportation service provider that was founded in 2008. There are many ways that transportation service companies can win the competition, including by increasing transportation customer satisfaction. This study aims to analyze the potential for KRL customer satisfaction by using data mining techniques with the C4.5 algorithm. The research instrument obtained data in the form of a questionnaire, the attributes of potential customer satisfaction in this study included price, facilities, services, and loyalty. This study uses 2 scales, namely the nominal scale in the form of codes or labels, and the interval scale in the form of weights in the answers to questions. In this study, the results obtained from several attributes produce a cause-and-effect relationship in classifying satisfied and dissatisfied customers. This research is hoped to be able to help the KRL party in increasing customer satisfaction to retain customers and increase the profit of the KRL company. The classification results using the C4.5 algorithm obtained an accuracy of 91.67%, which indicates that the C4.5 algorithm is suitable for measuring the potential for KRL customer satisfaction.</i></p>	<p>Diterima: 14-02-2021 Direvisi: 10-03-2021 Dipublikasikan: 27-04-2021</p>
	<p>Keywords <i>Data Mining, Decision Tree, Algoritma C4.5, Potensi Kepuasan Pelanggan, Perusahaan Jasa Transportasi.</i></p>

I. Pendahuluan

Pada saat ini kesadaran konsumen akan pentingnya nilai kualitas pelayanan (*Service Quality*) yang diberikan oleh perusahaan, baik dalam bentuk jasa maupun dalam bentuk barang semakin meningkat. Dari beberapa pengalaman menunjukkan bahwa atas pemberian suatu kualitas jasa/ pelayanan tertentu akan menimbulkan penilaian yang berbeda dari setiap pelanggan, karena tergantung dari bagaimana pelanggan mengharapkan kualitas jasa/ pelayanan tersebut. Pelanggan menilai pelayanan tersebut dengan cara membandingkan pelayanan yang mereka terima dengan pelayanan yang mereka harapkan. Perusahaan yang dapat memuaskan pelanggan akan memiliki pelanggan yang loyal, Sehingga secara tidak langsung pelanggan yang loyal tersebut akan merekomendasikan 2 kepada pelanggan lain tanpa dipaksa mengenai pelayanan yang pernah diterimanya. [1].

Data mining adalah proses pengekstrakan informasi dari kumpulan data yang besar. Pengekstrakan informasi dilakukan berdasarkan metode data mining yang akan digunakan. Algoritma C4.5 adalah algoritma yang digunakan untuk membuat pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi, metode pohon keputusan mengubah fakta menjadi pohon keputusan yang merepresentasikan aturan yang mudah dipahami dengan bahasa alami. [2].

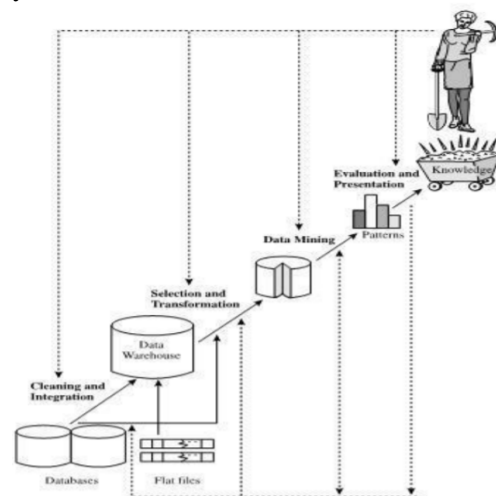
II. Metodologi

2.1 Data Mining

Data mining adalah suatu algoritma di dalam menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (*database*) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui. Analisa data mining berjalan pada data yang cenderung terus membesar dan teknik terbaik yang digunakan kemudian berorientasi kepada data berukuran sangat

besar untuk mendapatkan kesimpulan dan keputusan paling layak. Data mining memiliki beberapa sebutan atau nama lain yaitu: *Knowledge discovery (mining) in databases (KDD)*, *ekstraksi pengetahuan (knowledge extraction)*, *analisa data/pola, kecerdasan bisnis (business intelligence)*, dll. KDD merupakan kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar serta hubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi, dan visualisasi dari pola-pola sejumlah kumpulan data. [3].

Menurut Fayyad dalam bukunya berjudul “Advances in Knowledge Discovery and Data Mining”, tahapan proses dalam data mining secara garis besar dimulai dari data sumber dan berakhir dengan adanya informasi yang dihasilkan dari beberapa tahapan, Adapun beberapa tahapannya yaitu [4]:



Gambar 1 Tahapan Data Mining

- a. Seleksi Data
Pemilihan (seleksi) data baru dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.
- b. Data Cleaning

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses pembersihan pada data yang menjadi fokus KDD. Proses pembersihan mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).

c. Transformation

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses pembersihan pada data yang menjadi fokus KDD. Proses pembersihan mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).

d. Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

e. Interpretasi/Evaluasi

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut dengan *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya. Karena data mining adalah suatu rangkaian proses yang dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat interaktif dimana pemakai terlibat langsung atau dengan perantara knowledge base. Tahap-tahap ini diilustrasikan pada gambar berikut :

2.2 Decision Tree

Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti *Structured Query Language* untuk mencari *record* pada kategori tertentu [5].

Pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan, masing-masing rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain [6].

Pada *decision tree* terdiri dari tiga bagian yaitu [7] :

1. *Root Node*

Node ini merupakan node yang terletak paling atas dari suatu pohon.

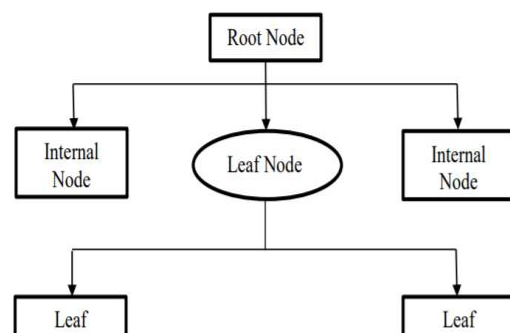
2. *Internal Node*

Node ini merupakan node percabangan, hanya terdapat satu input serta mempunyai minimal dua output.

3. *Leaf Node*

Node ini merupakan node akhir, hanya memiliki satu input, dan tidak memiliki output.

Contoh dari pohon keputusan dapat dilihat dari gambar dibawah ini :



Gambar 2 Model Pohon Keputusan

Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti Structured Query Language untuk mencari record pada kategori tertentu [8].

2.3 Algoritma C4.5

Algoritma C4.5 adalah algoritma yang sudah banyak dikenal dan digunakan untuk klasifikasi data yang memiliki atribut-atribut numerik dan kategorial. Hasil dari proses klasifikasi yang berupa aturan-aturan dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari record yang baru [9]. Algoritma C4.5 adalah salah satu metode untuk membuat decision tree berdasarkan training data yang telah disediakan. Algoritma C4.5 merupakan pengembangan dari ID3. Beberapa pengembangan yang dilakukan pada C4.5 adalah sebagai antara lain bisa mengatasi missing value, bisa mengatasi continui data, dan pruning [10].

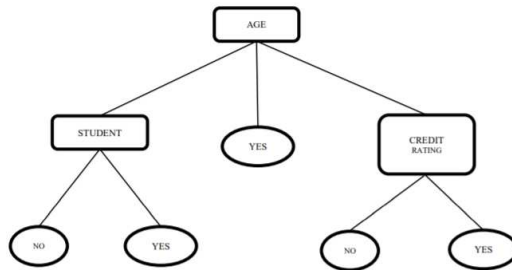
Algoritma C4.5 adalah salah satu metode untuk membuat decision tree berdasarkan training data yang telah disediakan. Algoritma C4.5 merupakan pengembangan dari ID3. Beberapa pengembangan yang dilakukan pada C4.5 adalah sebagai antara lain bisa mengatasi missing value, bisa mengatasi continui data, dan pruning [11].

Penelitian yang sebelumnya dilakukan oleh [12] penelitian ini meneliti tentang pola nasabah menggunakan metode klasifikasi dengan algoritma C4.5 pada Bank BRI Batang dan hasil klasifikasi menggunakan algoritma C4.5 menunjukkan bahwa diperoleh akurasi mencapai 89,5%. Pada penelitian ini menggunakan data primer dengan jumlah 81 data nasabah kredit dengan atribut nama, nomor rekening, umur, status, penghasilan, angsuran, jangka waktu, dan

label. Dari data tersebut diambil 16 data sample yang digunakan sebagai contoh perhitungan manual entropy dan gain, dimana pada data sample memiliki 7 data nasabah lancar dan 9 data nasabah macet. Dari hasil pengujian menggunakan software rapidminer dengan metode decision tree C4.5 dengan jumlah data sebanyak 81 data record, maka menghasilkan 72 data teridentifikasi benar dan 9 data teridentifikasi salah. Berdasarkan evaluasi bab sebelumnya dan dari pengujian yang telah dilakukan, penerapan data mining klasifikasi pola nasabah menggunakan algoritma C4.5 pada Bank BRI Batang menghasilkan akurasi yang tinggi, baik menggunakan aplikasi maupun software rapidminer, yaitu 89,5%. Ibnu Fatchu Rohman dalam penelitiannya menggunakan algoritma C4.5 pada kepuasan pelanggan damri menghasilkan akurasi sebesar 93%, selanjutnya Mochammad Rizki Ilham melakukan penelitian pengujian prediksi kepuasan pelanggan taksi kosti menggunakan metode decision tree C4.5 diperoleh akurasi sebesar 88,01% [13]

Algoritma C4.5 memiliki kelebihan utama karena dapat menghasilkan model berupa pohon atau aturan yang mudah diinterpretasikan dan dikonversi ke aturan-aturan *Structure Query Language* (SQL), memiliki tingkat akurasi yang dapat diterima, dapat menangani atribut bertipe diskret dan numerik, dan efisien dalam menangani atribut bertipe diskret [14]. Dalam pohon keputusan sangat berhubungan dengan Algoritma C4.5, karena dasar Algoritma C4.5 adalah pohon keputusan. Algoritma data mining C4.5 merupakan salah satu algoritma yang digunakan untuk melakukan klasifikasi atau segmentasi atau pengelompokan yang bersifat prediktif. Cabang-cabang pohon keputusan merupakan pertanyaan klasifikasi dan daun-daunnya merupakan kelas-kelas atau segmen-segmennya [15].

Contoh dari pohon keputusan dapat dilihat pada gambar dibawah ini:



Gambar 3 Contoh Pohon Keputusan

2.4 Confusion Matrix

Confusion matrix adalah *tool* yang digunakan untuk evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah. Sebuah matrix dari prediksi yang akan dibandingkan dengan kelas yang asli dari inputan atau dengan kata lain berisi informasi nilai actual dan prediksi pada klasifikasi.

Tabel 1 Confusion Matrix

Classification	Predicted Class	
	Class = Yes	Class = No
Class = Yes	True Positive	True Negative
Class = No	False Positive	False Negative

Rumus untuk menghitung tingkat akurasi pada matriks adalah :

$$Accuracy = \frac{tp+tn}{tp+tn+fp+fn} \times 100\% = \frac{a+d}{a+d+c+b} \times 100\% \dots\dots(1)$$

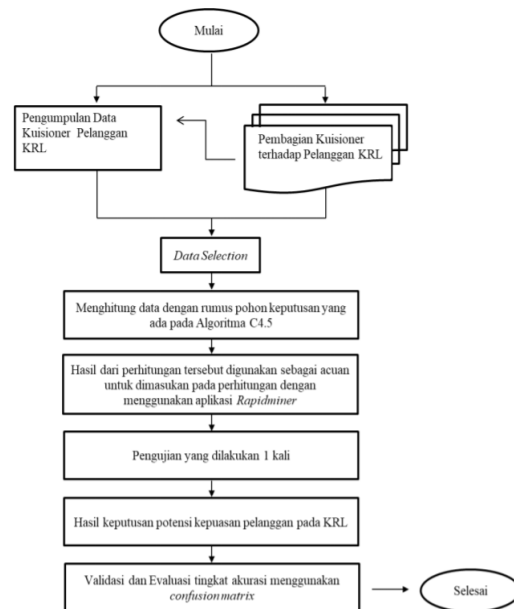
III. Hasil dan Pembahasan

Pada penelitian ini akan dilakukan tahapan-tahapan penelitian sebagai berikut:

- a. Data yang digunakan merupakan dataset yang diambil melalui kuisoner Adapun variable kuisoner terdiri dari Harga, Fasilitas, Pelayanan, dan loyalitas.
- b. Skala pengukuran yang digunakan pada penelitian ini ada 2, yang pertama adalah skala nominal yang paling sederhana dimana angka yang diberikan kepada suatu kategori

lainnya, hanya berupa kode atau label, contoh : gender atau status. Selanjutnya yang kedua yaitu Skala Interval yang memiliki jarak yang tetap antar respon yang ditawarkan, biasanya 1 unit skala, contoh : kepuasan memiliki nilai skala 1-5.

- c. Dalam penelitian ini, dilakukan dengan mengambil objek penelitian 200 data kuisoner potensi kepuasan pelanggan pada KRL.
- d. Dalam penelitian ini dilakukan beberapa tahap penelitian, antara lain sebagai berikut :



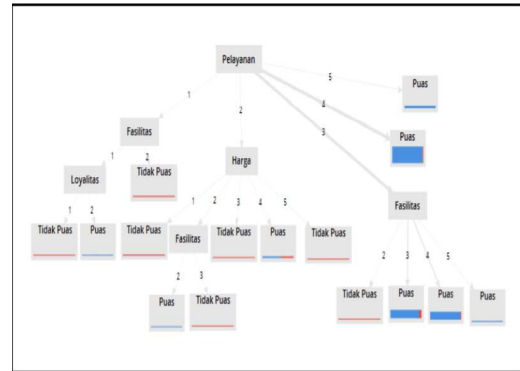
Gambar 4 Tahapan Penelitian

3.1 Pengujian Manual

Langkah awal Algoritma C4.5 adalah mencari nilai *entropy*, selanjutnya yaitu mencari nilai *Gain Ratio*. Berikut adalah table hasil perhitungan nilai *entropy* dan *Gain Ratio*.

Tabel 2 Hasil Perhitungan

Node	Attribute	Nilai	Sum(nilai)	Sum(P)	Sum(TP)	Entropi	Gain
Total Kasus			200	183	17	0.41956	
1	Harga	5	56	54	2	0.22228	0.07115
		4	101	94	7	0.36333	
		3	33	31	2	0.32985	
		2	3	1	2	0.91830	
		1	7	3	4	0.98523	
	Fasilitas	5	18	18	0	0.00000	0.09520
		4	107	103	4	0.23016	
		3	66	59	7	0.48792	
		2	7	2	5	0.86312	
		1	2	1	1	1.00000	
	Pelayanan	5	12	12	0	0.00000	0.14529
		4	88	86	2	0.15649	
		3	85	80	5	0.32276	
		2	11	4	7	0.94566	
		1	4	1	3	0.81128	
	Loyalitas	5	10	10	0	0.00000	0.10165
		4	112	110	2	0.12923	
		3	63	55	8	0.54911	
		2	12	7	5	0.97987	
		1	3	1	2	0.91830	

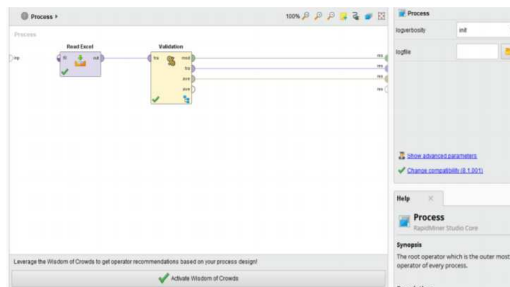


Gambar 6 Hasil Pohon Keputusan

Nilai gain tertinggi dijadikan node akar pertama. Pada table diatas node akar tertinggi adalah *attribute* pelayanan dengan nilai 0,14529.

3.2 Pengujian Rapidminer

Pengujian Kedua yaitu proses pengolahan data menggunakan algoritma C4.5 untuk mendapatkan pohon keputusan dan rule pada Rapidminer. Berikut adalah model yang dijalankan untuk menggunakan *Split Validation* untuk memproses algoritma C4.5.



Gambar 5 Pengujian Algoritma C4.5 Split Validation

Setelah dilakukan metode algoritma C4.5 pada Rapidminer, maka terbentuk pohon keputusan seperti gambar 4.2 :

IV. Kesimpulan

Berdasarkan penelitian yang telah dilakukan algoritma C4.5 didapatkan hasil *Accuracy* sebesar 91,67%, *Precision* sebesar 94,64 %, dan *Recall* sebesar 96,36 %. Atau bisa dilihat pada dibawah ini :

Tabel 3 Evaluasi dan Validasi

<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
91,67 %	94,64 %	96,36 %

Daftar Pustaka

[1] Rohman, Fatchur Ibnu. 2016. "Penerapan Algoritma C4 . 5 Pada Kepuasan Pelanggan Perum DAMRI.":14.mahasiswa.dinus.ac.id/docs/skripsi/jurnal/15396.pdf..

[2] S, Mochamad Rizki Ilham. 2016. "Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Kepuasan Pelanggan Taksi Kosti."

[3] Elmande, Y, and Pp Widodo. 2012. "Pemilihan Criteria Splitting Dalam Algoritma Iterative Dichotomiser 3 (ID3) Untuk Penentuan Kualitas Beras: Studi Kasus Pada Perum Bulog Divre Lampung." Jurnal Telematika Mkom 4(1): 10. <http://demo.pohonkeputusan.com/File s/Pemilihan Criteria Splitting Dalam>

- Algoritma Iterative Dichotomiser 3 (Id3) Untuk Penentuan Kualitas Beras.pdf
- [4] Ayuni, Asistiyasari, and Taufik Baidawi. 2017. "Analisis Penerimaan Karyawan Posisi Field Collector Menggunakan Algoritma C4 . 5 Pada PT. PRISMAS JAMINTARA JAKARTA." *Jurnal Ilmu Pengetahuan dan Teknologi Komputer* 2(2): 1–6. ejournal.nusamandiri.ac.id/ejurnal/index.php/jitk/article/view/223/199.
- [5] D. Hastuti et al., "Algoritma Naïve Baiyes Untuk Prediksi Profesi Berdasarkan Skill Job Seeker," No. April, 2017
- [6] M. Ayu, D. Widyadara, And R. H. Irawan, "Implementasi Metode Naïve Bayes Dalam Penentuan Tingkat Kesejahteraan Keluarga," Vol. 2, No. 1, Pp. 19–24, 2019.
- [7] A. Rane, N. Naik and J. Laxminarayana, "Performance Enhancement of K Nearest Neighbor Classification Algorithm Using 8-Bin Hashing and Feature Weighting," *ACM* 978-1-4503-2908-8/14/08,2014.<http://www.cs.umass.edu/lfw/>, Wild, Labeled Faces in the, 2016.
- [8] Hadi Sucipto, Kusriani, Hanif Al Fatta. 2017. "Sistem Pendukung Keputusan Penerimaan Siswa Baru Dan Penjurusan Pada Sekolah Menengah Kejuruan(SMK)." 5(2). ejournal.polsa.ac.id/index.php/jnet/article/download/76/63.
- [9] Elisa, Erlin. 2017. "Analisa Dan Penerapan Algoritma C4 . 5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT . Arupadhatu Adisesanti." *JOIN (Jurnal Online Informatika)* 2(1): 36–41.
- [10] Hikma Widayu, Surya Darma Nasution, Natalia Silalahi, Mesran. 2017. "Data Mining Untuk Memprediksi Jenis Transaksi Nasabah Pada Koperasi Simpan Pinjam Dengan Algoritma C4.5." 1(2): 32–37. <http://ejournal.stmikbudidarma.ac.id/index.php/mib/article/view/323>.
- [11] Hikma Widayu, Surya Darma Nasution, Natalia Silalahi, Mesran. 2017. "Data Mining Untuk Memprediksi Jenis Transaksi Nasabah Pada Koperasi Simpan Pinjam Dengan Algoritma C4.5." 1(2): 32–37. <http://ejournal.stmikbudidarma.ac.id/index.php/mib/article/view/323>
- [12] Oktaviana, AR. 2016. "Penerapan Datamining Klasifikasi Pola Nasabah."
- [13] Rohman, Fatchur Ibnu. 2016. "Penerapan Algoritma C4 . 5 Pada Kepuasan Pelanggan Perum DAMRI." :1–14. mahasiswa.dinus.ac.id/docs/skripsi/jurnal/15396.pdf
- [14] S, Mochamad Rizki Ilham. 2016. "Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Kepuasan Pelanggan Taksi Kosti."
- [15] Shella, Pradega. 2015. "Sistem Pendukung Keputusan dengan menggunakan Decission Tree Di Sekolah Menengah Pertama (Studi Kasus Di SMPN 2 Rembang).