



Penerapan Decision Tree untuk Prediksi Kepuasan Pengguna Bus Transjakarta

Reza Hikmatulloh¹, Hanna Mahaerani P², Qurrotul Aini³

^{1,2,3}Universitas Islam Negeri Syarif Hidayatullah, Jl. Ir. H. Djuanda No. 95, Ciputat, Tangerang Selatan, Banten

¹Reza.Hikmatulloh16@mhs.uinjkt.ac.id, ²Hanna.Mahaerani16@mhs.uinjkt.ac.id, ³Qurrotul.aini@uinjkt.ac.id

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 22-08-2020

Revisi Akhir: 17-10-2020

Diterbitkan Online: 20-10-2020

KATA KUNCI

Kepuasan Pengguna,
Bus Transjakarta,
Decision tree,
Algoritma C4.5

KORESPONDENSI

Telepon: +6281210600746

E-mail: reza.hikmatulloh16@mhs.uinjkt.ac.id

A B S T R A C T

Transjakarta buses are one of the public transportation facilities used by residents of the capital city of Jakarta. The existence of Transjakarta Buses, transportation access is easier and easier for its users. With so many users, there are several things that must be considered in the facilities and infrastructure provided by the Transjakarta Bus. User satisfaction is an assessment given by a user of a service, whether satisfied or not. This research was conducted to analyze the satisfaction of Transjakarta Bus users who are students of Syarif Hidayatullah State Islamic University Jakarta. The attributes to be taken in the study include the rates, facilities, and services provided by the Transjakarta Bus where the questionnaire results will be processed using the C4.5 Decision Tree method. The results showed that an accuracy of 92.86% was obtained with accuracy in Good Classification using Confusion Matrix.

1. PENDAHULUAN

Transportasi umum dapat mengangkut banyak orang dalam satu kali jalan merupakan salah satu solusi untuk mengurangi kemacetan yang terjadi di ibu kota. Karena semakin banyaknya pengguna transportasi umum, maka perlu diperhatikan masalah sarana dan prasarana yang menunjang agar pengguna nyaman. Beberapa moda transportasi umum yang digunakan oleh masyarakat, diantaranya: kereta, pesawat, angkot, taksi, dan bus.

Bus merupakan transportasi umum yang sering digunakan oleh banyak orang. Dengan harga yang terjangkau, masyarakat dapat menempuh jarak yang jauh dengan transportasi ini. Transportasi umum bus dapat melayani rute yang berbeda-beda tergantung dengan tujuan pengguna. Transportasi umum bus di Jakarta banyak yang sudah tidak layak jalan, dipakai kebut-kebutan yang dapat menimbulkan kecelakaan dan merugikan banyak orang. Salah satu cara pemerintah agar dapat meningkatkan kualitas pelayanan transportasi adalah dengan menetapkan kebijakan dengan merintis jalur khusus angkutan umum jalan raya sejak 15 Januari 2004 dengan nama Transjakarta dan jalurnya disebut Busway [1].

Bus Transjakarta merupakan sistem transportasi *Bus Rapid Transit* (BRT) yang mempunyai pelayanan yang berbeda dari transportasi umum bus lainnya, dimana Bus Transjakarta memiliki jalur khusus, supirnya diberikan

gaji, hanya dapat menaikkan atau menurunkan penumpang di halte yang di sediakan, dan mempunyai petugas *on board* dimana petugas ini ada di dalam Bus Transjakarta untuk memandu dan melayani pengguna. Petugas ini sebelumnya sudah diberikan pelatihan khusus agar dapat maksimal dalam memberikan pelayanan. Pelayanan yang diberikan merupakan salah satu hal yang cukup penting untuk membuat konsumen menjadi loyal dalam menggunakan jasa Bus Transjakarta. Salah satu cara agar dapat memprediksi kepuasan pengguna Bus Transjakarta adalah dengan menggunakan teknik *data mining*.

Data Mining merupakan pengumpulan atau penambangan informasi baru dengan tujuan untuk mendapatkan pola atau aturan tertentu dari sejumlah data dalam jumlah data yang besar [2]. Keluaran *data mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan [3]. Ada beberapa metode dan algoritma dalam *data mining* yang dapat digunakan untuk mengekstrak informasi antara lain: metode klasifikasi, metode asosiasi, metode *clustering*, metode prediksi, dan metode estimasi.

Pada penelitian ini, teknik yang digunakan adalah teknik klasifikasi. Klasifikasi merupakan suatu proses yang menemukan properti-properti yang sama pada sebuah himpunan obyek di dalam sebuah basis data dan mengklasifikasikannya ke dalam kelas-kelas yang berbeda

menurut model klasifikasi yang ditetapkan. Tujuan dari teknik klasifikasi adalah menemukan model dari *training set* yang membedakan atribut ke dalam kategori atau kelas yang sesuai, model tersebut kemudian digunakan untuk mengklasifikasi atribut yang kelasnya belum diketahui sebelumnya. Teknik klasifikasi terbagi menjadi beberapa teknik yang diantaranya adalah *decision tree* [4].

Decision tree atau pohon keputusan adalah salah satu metode klasifikasi yang kuat dan terkenal. Metode *Decision Tree* mengubah fakta besar menjadi pohon keputusan yang mewakili aturan, sehingga aturan tersebut dapat dengan mudah dipahami oleh manusia. *Decision Tree* juga berfungsi untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah variabel input dan variabel tujuan [5]. Metode *decision tree* memiliki algoritma, algoritma C4.5 adalah salah satu dari algoritma yang memiliki *decision tree*.

Algoritma C4.5 merupakan metode yang dapat digunakan untuk melakukan pembentukan pohon keputusan [6]. Algoritma C4.5 dan pohon keputusan merupakan dua model yang tak terpisahkan, karena untuk membangun sebuah pohon keputusan, dibutuhkan algoritma C4.5. Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5, yaitu [7]:

- 1) Menyiapkan data training. Data training biasanya dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan 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*-nya.

Pada penelitian sebelumnya tentang penggunaan Algoritma *Decision Tree* (C4.5) untuk Memprediksi Kepuasan Mahasiswa Terhadap Kinerja Dosen Politeknik TEDC Bandung, menggunakan beberapa indikator yaitu bukti langsung, keandalan, daya tanggap dan empati. Hasil penelitian menunjukkan bahwa indikator tangible (bukti langsung) memiliki pengaruh paling besar pada penelitian ini [8]. Penelitian selanjutnya berjudul Analisa Kepuasan Konsumen Menggunakan Klasifikasi *Decision Tree* Di Restoran Dapur Solo, menghasilkan sebuah sistem yang dapat menghasilkan suatu informasi yang siap digunakan untuk membantu menentukan langkah strategis yang harus diambil dan dapat mengetahui sejauh mana kepuasan konsumen dan bidang apa yang harus ditingkatkan untuk mencapai kepuasan dari para konsumennya [9]. Penelitian selanjutnya yang berjudul Prediksi Kepuasan Mahasiswa Terhadap Tingkat Pelayanan Menggunakan Algoritma C4.5, menggunakan beberapa atribut untuk menganalisa tingkat kepuasan mahasiswa terhadap pelayanan yang terdapat di Fakultas Teknik Universitas Hamzanwadi, antara lain: biaya, fasilitas, pelayanan dan loyalitas. Dari hasil penelitian, mendapatkan hasil akurasi sebesar 81.07 % [5].

Berdasarkan hasil evaluasi dari penelitian sebelumnya menunjukkan bahwa Algoritma *Decision Tree* C4.5 akurat diterapkan untuk pengukuran kepuasan pengguna dimana konsep yang digunakan oleh *decision tree* adalah mengubah data menjadi suatu keputusan pohon dan aturan-aturan keputusan (*rules*). Oleh karena itu, dalam penelitian ini menggunakan metode Algoritma *Decision Tree* C4.5

yang bertujuan untuk mengetahui apakah pelayanan yang diberikan dari transportasi umum Bus Transjakarta sudah memuaskan dan dapat diterima oleh pelanggan. Dengan metode yang diambil yaitu *decision tree*, diharapkan mendapatkan data yang relevan, akurat dan terpercaya. Teknik pengambilan data yang dilakukan dengan menggunakan kuesioner yang akan disebar ke responden pengguna Bus Transjakarta koridor Ciputat - Kp. Rambutan yang berasal dari kampus UIN Syarif Hidayatullah Jakarta. Adapun atribut yang akan diambil dalam kuesioner meliputi tarif, fasilitas, dan pelayanan yang diberikan oleh Bus Transjakarta.

2. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah deskriptif kuantitatif yang bertujuan untuk menggambarkan, meringkas berbagai kondisi, berbagai situasi atau berbagai variabel yang timbul di masyarakat. Yang menjadi objek dalam penelitian ini adalah pengguna Bus Transjakarta Koridor Ciputat – Kp. Rambutan. Proses analisa data menggunakan tools RapidMiner Studio.

2.1. Pemilihan Sampel

1) Populasi

Populasi yang digunakan pada penelitian ini adalah pengguna transportasi umum Bus Transjakarta koridor Ciputat – Kp. Rambutan yang merupakan mahasiswa Universitas Islam Syarif Hidayatullah Jakarta. Teknik yang digunakan dalam pengambilan sampel adalah *Simple Random Sampling*, yaitu teknik pengambilan sampel yang memberikan kesempatan yang sama kepada setiap anggota yang ada dalam suatu populasi untuk dijadikan sampel.

2) Responden Penelitian

Pemilihan sampel dengan metode yang tepat dapat menggambarkan kondisi populasi sesungguhnya yang akurat, dan dapat menghemat biaya penelitian secara efektif. Idealnya, sampel haruslah benar-benar menggambarkan atau mewakili karakteristik populasi yang sebenarnya.

Menurut Sugiyono, ukuran sampel yang layak dalam penelitian adalah antara 30 sampai dengan 500 [10]. Oleh karena itu, pada penelitian ini jumlah sampel yang akan diambil dari populasi sebanyak 100 orang pengguna transportasi umum Bus Transjakarta koridor Ciputat – Kp. Rambutan yang akan mengisi kuesioner yang akan diberikan.

2.2. Pengumpulan Data

1) Studi Pustaka

Studi pustaka dilakukan dengan mempelajari teori-teori terkait dan hasil penelitian sebelumnya yang dapat mendukung pemecahan masalah penelitian. Pencarian informasi metodologi pustaka ini dengan membaca dari buku-buku dan *website* yang membahas tentang penerapan data *mining* dengan menggunakan *decision tree* untuk prediksi kepuasan pengguna.

2) Studi Literatur

Dalam melengkapi data-data yang diperoleh, penulis mengumpulkan data-data yang dibutuhkan menggunakan studi literatur dengan cara membaca dan mempelajari

buku-buku, jurnal-jurnal dan tugas akhir yang dijadikan sebagai referensi yang mendukung topik yang akan dibahas dalam penelitian ini.

3) Kuesioner

Metode kuesioner digunakan untuk menggali informasi dari pengguna tentang pelayanan, fasilitas, dan biaya yang diberikan dari Bus Transjakarta. Dengan metode ini diharapkan dapat memperoleh data-data yang relevan, akurat, dan terpercaya. Kuesioner merupakan teknik pengumpulan data yang dilakukan dengan cara memberi seperangkat pertanyaan atau pernyataan tertulis kepada responden untuk dijawabnya. Metode ini dilakukan dengan menyebarkan daftar pernyataan atau pertanyaan dimana hasilnya nanti kemudian diolah untuk mendapatkan nilai rata-rata dari variabel yang akan diolah pada proses *data mining* dengan teknik *decision tree* menggunakan algoritma C4.5.

2.3. Pengukuran Data

Untuk mendapatkan hasil penerapan *data mining* untuk memprediksi kepuasan pengguna Bus Transjakarta, maka diperlukan sebuah metode pengukuran. Skala yang digunakan dalam kuesioner untuk memberikan sejumlah pertanyaan kepada responden dengan menggunakan skala likert. Skala likert digunakan untuk mengukur sikap, pendapat dan persepsi seseorang atau sekelompok orang tentang fenomena sosial. Pengolahan data berdasarkan penilaian responden pada kuesioner menggunakan Skala Likert yang dapat dilihat pada Tabel 1.

TABEL 1. SKALA LIKERT

Jawaban	Akronim	Skor
Sangat Setuju	SS	5
Setuju	S	4
Ragu-ragu	RG	3
Tidak Setuju	TS	2
Sangat Tidak Setuju	STS	1

2.4. Analisis Data

Data yang digunakan dalam penelitian ini adalah data berdasarkan pada kuesioner yang diberikan kepada pelanggan pengguna Bus Transjakarta. Metode yang diusulkan yaitu menggunakan metode Algoritma C4.5 dengan *tools* Rapidminer. Untuk mengetahui keakuratan dari hasil perhitungan menggunakan Rapidminer, maka dilakukan pengujian sebanyak tiga kali dengan tingkat acuan pengujian yang berbeda. Setelah dilakukan perhitungan, hasil dari evaluasi tersebut baru dapat diketahui sampai sejauh mana kepuasan pelanggan pada Bus Transjakarta.

Tahap yang dilakukan untuk melakukan pengujian ini adalah dengan menghitung nilai akurasi (persamaan 1), *precision* (persamaan 2), *recall* (persamaan 3) dan nilai AUC (*Area Under Curve*), seperti berikut:

$$\text{Akurasi} = \frac{tp+tn}{tp+tn+fp+fn} \times 100\% \quad (1)$$

$$\text{Precision} = \frac{tp}{tp+fn} \times 100\% \quad (2)$$

$$\text{Recall} = \frac{tp}{tp+fp} \times 100\% \quad (3)$$

Pada persamaan di atas, *tp* merupakan nilai jumlah benar positif, *tn* merupakan nilai jumlah benar negatif, *fp* merupakan nilai jumlah salah positif, dan *fn* merupakan nilai jumlah salah negatif.

3. HASIL DAN PEMBAHASAN

3.1. Pengolahan Data

Data yang digunakan pada penelitian ini adalah data berupa kuesioner yang diberikan kepada pengguna transportasi Bus Transjakarta koridor Ciputat – Kp. Rambutan sebanyak 100 kuesioner dengan atribut harga, fasilitas dan pelayanan. Item pertanyaan kuesioner adalah sebagai berikut:

- 1) Tarif harga Bus Transjakarta relatif terjangkau
- 2) Fasilitas didalam Bus Transjakarta seperti AC, Radio, tempat duduk, dan gantungan bus sudah memadai dan baik digunakan.
- 3) Bus Transjakarta bersih akan sampah.
- 4) Tempat duduk yang nyaman.
- 5) Tersedianya tempat duduk prioritas bagi ibu hamil, lansia, membawa anak, dan disabilitas.
- 6) Petugas dan Karyawan seperti supir dan kenek nya berpakaian rapih dan bersih.
- 7) Petugas dan karyawan memberikan respon dan pelayanan yang baik untuk para penumpang.
- 8) Para petugas selalu memberikan tiket untuk satu kali perjalanan untuk penumpang
- 9) Kemampuan petugas dalam menjawab tempat pemberhentian yang dituju
- 10) Bus yang digunakan memiliki mesin yang baik sehingga jarang mengalami kerusakan
- 11) Kecepatan bus berjalan sesuai dengan aturan yaitu 50 km/jam
- 12) Bus Transjakarta selalu mematuhi aturan lalu lintas
- 13) Bus Transjakarta selalu datang tepat waktu sesuai dengan jadwal yang di tentukan atau jam-jam operasional
- 14) Bus Transjakarta menjamin keselamatan penumpang
- 15) Bus Transjakarta menyediakan asuransi untuk penumpang
- 16) Bagaimana kepuasan akan tarif, pelayanan, dan fasilitas yang disediakan oleh Bus Transjakarta koridor Ciputat – Kp. Rambutan ? Puas/Tidak Puas

Data kuesioner yang terkumpul kemudian diolah menggunakan Microsoft Excel untuk mengambil rata-rata dari tiap atribut. Hasil perhitungan rata-rata ditunjukkan pada Tabel 2.

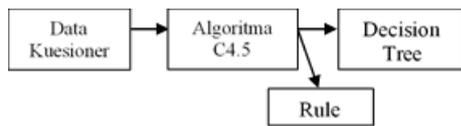
TABEL 2. HASIL RATA-RATA

Nama	Harga (x1)	Fasilitas (x2)	Pelayanan (x3)	Hasil (Y)
Polynomial	Numeric	Numeric	Numeric	Binominal
Id	Atributte	Atributte	Atributte	Label
Indah Kusuma	5	5	4.4	Puas
Syarifah	4	4	4.4	Puas
Noni	5	4	3.9	Puas
Eljo	5	4.75	4.4	Puas
Mifta Namira	5	5	3.7	Puas
Hanif Aulia	4	3.75	4	Puas
Firda	2	4	3.8	Puas
Pandu Mahesa	4	4.5	4.1	Puas
....

Hitung rata-rata dari hasil kuesioner x1 (Harga) yang di tetapkan oleh Bus Transjakarta, rata-rata x2 (Fasilitas) yang disediakan, dan x3 (pelayanan) yang di berikan, dan Y (Hasil) dari perhitungan rata-rata kuesioner. Dari hasil kuesioner yang di sebarakan kepada 100 orang pengguna Bus Transjakarta, diperoleh hasil 93 merasa PUAS, dan 7 orang merasa TIDAK PUAS.

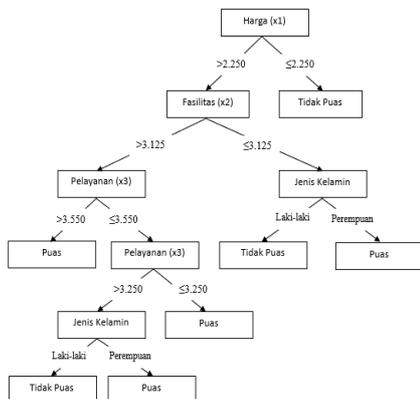
3.2. Proses Data Mining

Pada tahap selanjutnya, data akan diproses menggunakan tools Rapidminer dimana pengujiannya menggunakan Cross Validation yang sudah ada pada Rapidminer. Gambar proses pengolahan data menggunakan algoritma C4.5 untuk mendapatkan pohon keputusan dan rules pada Rapidminer dapat dilihat pada Gambar 1.



Gambar 1. Proses pengolahan data pada Rapidminer

Setelah melakukan proses pengolahan data diatas, akan menghasilkan suatu pohon keputusan yang akan ditunjukkan pada Gambar 2.



Gambar 2. Pohon Keputusan

Pohon keputusan yang dihasilkan terdiri dari 5 leave dimana 4 berada di kiri dan 1 berada di sisi kanan. Dari gambaran yang diberikan oleh diagram pohon terlihat bahwa penentu utama faktor kepuasan adalah faktor fasilitas. Pohon keputusan tersebut menghasilkan rules yang dapat dilihat pada Tabel 3.

TABEL 3. RULES

RULES UNTUK HASIL = PUAS	
1.	IF nilai Harga (X1) >2.250 dan nilai Fasilitas >3,125 dan nilai Pelayanan (X) >3.550 Then Hasil = PUAS.
2.	IF nilai Fasilitas (X2) <=3.125 dan Jenis Kelamin Perempuan Then Hasil = PUAS.
3.	IF nilai Pelayanan(X)>3.250 dan Jenis Kelamin Perempuan Then nilai Hasil = PUAS.
4.	IF nilai Pelayanan (X) <=3.250 Then Hasil = PUAS
RULES UNTUK HASIL = TIDAK PUAS	
1.	IF nilai Harga (X1) <= 2.250 Then nilai Hasil = TIDAK PUAS.
2.	IF nilai Harga (X1) >2.250 dan nilai Fasilitas (X2) <=3.125 dan Jenis Kelamin Laki – laki Then Hasil = TIDAK PUAS.
3.	IF nilai Pelayanan (X) >3.250 dan Jenis kelamin Laki – laki Then nilai Hasil = TIDAK PUAS.

3.3. Pengujian Rules

Pengujian rules dilakukan untuk mengetahui kinerja dari algoritma C4.5 dalam melakukan klasifikasi terhadap kelas yang telah ditentukan dalam uji coba. Pengujian yang dilakukan dalam penelitian ini sebanyak 3 kali dengan menggunakan metode K-Fold Validation yang berbeda pada operator cross validation. Pengujian terdiri atas K-Fold Validation 10, K-Fold Validation 6, dan K-Fold Validation 3.

Pada model Algoritma C4.5 juga melakukan suatu pengujian terhadap data-data yang telah dibagi oleh cross validation menjadi dua bagian process yaitu training dan testing. Training terdiri dari Pohon keputusan, sedangkan testing terdiri dari apply model dan performance.

1) Pengujian 1

Pada pengujian pertama dilakukan dengan membagi 10 bagian pada 100 data kuesioner yang akan di uji dalam penelitian ini. 10 bagian ini terdiri dari 9 data yang akan di training dan 1 bagian testing.

Dalam pengujian ini akan menghasilkan nilai akurasi precision, recall, dan AUC (Area Under Curve). Hasil akurasi Confusion Matrix pada pengujian 1 dapat dilihat pada Tabel 4.

TABEL 4.HASIL AKURASI CONFUSION MATRIK PADA PENGUJIAN 1

	Actual	
	True PUAS	True TIDAK PUAS
Prediksi Puas	91	7
Prediksi Tidak Puas	0	0

Jumlah benar positif (tp) sebanyak 91 data, salah positif (fp) sebanyak 0 data, jumlah benar negative (tn) sebanyak 0 data dan jumlah salah negative (fn) sebanyak 7 data. Berdasarkan perhitungan, nilai akurasi yang di dapat dari persamaan 1 adalah 92.86%, nilai precision yang di dapat dari persamaan 2 adalah 92.86% dan nilai recall yang di dapat dari persamaan 3 adalah 100%. Hasil dari AUC (Area Under Curve) pada pengujian 1 mendapatkan hasil

0.978 yang berarti pengujian 1 termasuk pengujian *good classification*, yang dapat dilihat pada Gambar 3.



Gambar 3. Pengujian AUC (Arean Under Curve) ke 1

Berdasarkan hasil perhitungan di atas, hasil evaluasi dan validasi keseluruhan pada pengujian ke 1 dapat dilihat pada Tabel 5.

TABEL 5. EVALUASI DAN VALIDASI KE-1

K – Fold Validation	Akurasi	Precision	Recall	AUC
10	92.86%	92.86%	100%	0.987

2) Pengujian 2.

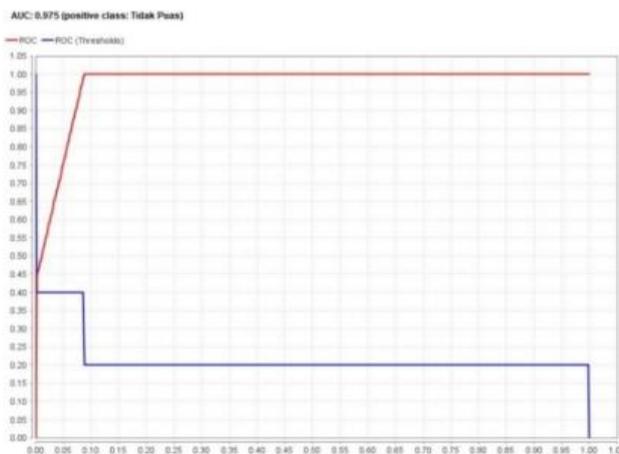
Pengujian ini dilakukan dengan membagi sebanyak 5 bagian pada 100 data kuesioner yang akan di uji 5 bagian ini terdiri dari 4 data *training* dan 1 data *testing*. Dalam pengujian ini akan menghasilkan nilai akurasi *precision*, *recall*, dan AUC (*Area Under Curve*). Hasil akurasi *Confusion Matrix* pada pengujian 2 dapat dilihat pada Tabel 6.

TABEL 6. HASIL AKURASI CONFUSION MATRIK PADA PENGUJIAN 2

	Actual	
	True PUAS	True TIDAK PUAS
Prediksi Puas	91	6
Prediksi Tidak Puas	0	1

Jumlah benar *positif* (tp) sebanyak 91 data, salah *positif* (fp) sebanyak 0 data, jumlah benar *negative* (tn) sebanyak 1 data dan jumlah salah *negative* (fn) sebanyak 6 data.

Berdasarkan perhitungan, nilai akurasi yang di dapat dari persamaan 1 adalah 93.81%, nilai *precision* yang di dapat dari persamaan 2 adalah 93.81% dan nilai *recall* yang di dapat dari persamaan 3 adalah 100%. Hasil dari AUC (*Area Under Curve*) pada pengujian 2 mendapatkan hasil 0.975 yang berarti pengujian 2 termasuk pengujian *good classification*, yang dapat dilihat pada Gambar 4.



Gambar 4. Pengujian AUC (Arean Under Curve) ke 2

Berdasarkan hasil perhitungan di atas, hasil evaluasi dan validasi keseluruhan pada pengujian ke 2 dapat dilihat pada Tabel 7.

TABEL 7. EVALUASI DAN VALIDASI KE-2

K – Fold Validation	Akurasi	Precision	Recall	AUC
5	93.81%	93.81%	100%	0.975

3) Pengujian 3

Pengujian ini dilakukan dengan membagi sebanyak 3 bagian pada 100 data kuesioner yang akan di uji, 3 bagian tersebut terdiri dari 2 bagian data *training* dan 1 bagian data *testing*. Dalam pengujian ini akan menghasilkan suatu nilai akurasi, *precision*, *recall*, dan AUC (*Area Under Curve*). Hasil akurasi *Confusion Matrix* pada pengujian 3 dapat dilihat pada Tabel 8.

TABEL 8. HASIL AKURASI CONFUSION MATRIK PADA PENGUJIAN 3

	Actual	
	True PUAS	True TIDAK PUAS
Prediksi Puas	91	5
Prediksi Tidak Puas	0	2

Jumlah benar *positif* (tp) sebanyak 91 data, salah *positif* (fp) sebanyak 0 data, jumlah benar *negative* (tn) sebanyak 2 data dan jumlah salah *negative* (fn) sebanyak 5 data.

Berdasarkan perhitungan, nilai akurasi yang di dapat dari persamaan 1 adalah 94.90%, nilai *precision* yang di dapat dari persamaan 2 adalah 94.79% dan nilai *recall* yang di dapat dari persamaan 3 adalah 100%. Hasil dari AUC (*Area Under Curve*) pada pengujian 3 mendapatkan hasil 0.988 yang berarti pengujian 3 termasuk pengujian *good classification*, yang dapat dilihat pada Gambar 5.



Gambar 5. Pengujian AUC (Area Under Curve) ke 3

Berdasarkan hasil perhitungan di atas, hasil evaluasi dan validasi keseluruhan pada pengujian ke 3 dapat dilihat pada Tabel 9.

TABEL 9. EVALUASI DAN VALIDASI KE-3

K - Fold Validation	Akurasi	Precision	Recall	AUC
3	94.90%	94.79%	100%	0.988

4) Hasil Pengujian Keseluruhan

Hasil dari pengujian 1,2, dan 3 menghasilkan akurasi, *precision*, *recall* dan AUC (*Area Under Curve*) akan disajikan pada Tabel 10.

TABEL 10. TABEL KESELURUHAN HASIL EVALUASI DAN VALIDASI

K - Fold Validation	Akurasi	Precision	Recall	AUC
10	92.86%	92.86%	100%	0.987
5	93.81%	93.81%	100%	0.975
3	94.90%	94.79%	100%	0.988

Berdasarkan tabel hasil keseluruhan dari *testing* dan *data training* untuk melihat keakuratan model, maka dapat diketahui bahwa dalam pengujian K-Fold Validation K10 mendapatkan nilai akurasi sebesar 92.86%, menggunakan K5 mendapatkan nilai akurasi sebesar 93.81% sedangkan dengan K3 mendapatkan nilai akurasi 94.90%. Nilai akurasi yang di dapat semakin meningkat dengan nilai *Recall* 100%. Hal ini menunjukkan bahwa nilai hasil yang hampir seratus persen akurat, presisi yang hampir akurat, dan nilai *recall* yang sempurna menyimpulkan bahwa peneliti berhasil dalam mengimplementasikan algoritma klasifikasi C4.5 dengan sangat baik dan dapat membantu Trans Jakarta Koridor Kp Rambutan - Ciputat dalam merekomendasikan dalam meningkatkan fasilitas dan pelayanan di dalam bus.

4. KESIMPULAN

Kesimpulan yang dapat diambil dari penelitian yang dilakukan adalah bahwa kepuasan pengguna Bus Transjakarta dapat di prediksi dan di evaluasi dengan memanfaatkan teknik data mining menggunakan algoritma *decision tree* C4.5 untuk memperdiksi dari kepuasan pelanggan dengan data training yang telah diperoleh. Setelah melakukan 3 kali proses *data training* dan *testing* untuk kepuasan pengguna Bus Transjakarta koridor

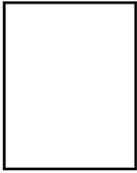
Ciputat – Kp. Rambutan dengan menggunakan *tools* penambangan data Rapidminer Studio menggunakan metode *decision tree* C4.5, diketahui bahwa percobaan 1,2 dan 3 dengan data training dikatakan berhasil, karena nilai akurasi yang terus meningkat dan semakin akurat. Dari hasil perhitungan diperoleh akurasi sebesar 92,86% dengan kriteria akurasi *Good Classification* menggunakan *Confusion Matrix*.

Penelitian ini hanya sampai penerapan metode, lebih baik untuk penelitian selanjutnya di kembangkan lagi sampai pada tahapan implementasi system metode klasifikasi. Untuk penelitian selanjutnya lebih baik menggunakan data yang lebih banyak lagi untuk menghasilkan *rules* yang akurat dan dengan jumlah kuisioner yang perlu ditambah, sehingga dapat diperoleh hasil akurasi yang lebih baik lagi dan akurat. Penelitian selanjutnya dapat dikembangkan dengan menggunakan atau membandingkan dengan algoritma klasifikasi lain untuk mendapatkan hasil prediksi yang lebih baik.

DAFTAR PUSTAKA

- [1] P. Suharso, "Kualitas Pelayanan Transjakarta," *War. Penelit. Perhub.*, vol. 26, no. 6, pp. 321–332, 2014.
- [2] S. Haryati, A. Sudarsono, and E. Suryana, "Implementasi Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
- [3] M. S. Efendi and H. A. Wibawa, "Prediksi Penyakit Diabetes Menggunakan Algoritma ID3 dengan Pemilihan Atribut Terbaik," *JUITA*, vol. VI, pp. 29–35, 2018.
- [4] Rismayanti, "Decision Tree Penentuan Masa Studi Mahasiswa Prodi Teknik Informatika (Studi Kasus : Fakultas Teknik dan Komputer Universitas Harapan Medan)," *QUERY J. Sist. Inf.*, vol. 02, no. 01, pp. 16–24, 2018.
- [5] B. A. C. P., "Prediksi Kepuasan Mahasiswa Terhadap Tingkat Pelayanan Menggunakan Algoritma C4.5 (Decision Tree) (Studi Kasus : Fakultas Teknik Universitas Hamzanwadi)," *Infotek J. Inform. dan Teknol.*, vol. 1, no. 1, pp. 32–39, 2018.
- [6] Oktafianto, "Analisis Kepuasan Mahasiswa Terhadap Pelayanan Akademik Menggunakan Metode Algoritma C4.5," *TIM Darmajaya*, vol. 02, no. 01, pp. 1–11, 2016.
- [7] E. T. L. Kusriani, "Algoritma data mining," *Yogyakarta Andi Offset*, 2009.
- [8] A. Yuliana and D. B. Pratomo, "Algoritma Decision Tree (C4.5) Untuk Memprediksi Kepuasan Mahasiswa Terhadap Kinerja Dosen Politeknik TEDC Bandung," in *Seminar Nasional Inovasi Teknologi*, 2017, pp. 377–384.
- [9] A. Shiddiq, R. K. Niswatin, and I. N. Farida, "Analisa Kepuasan Konsumen Menggunakan Klasifikasi Decision Tree Di Restoran Dapur Solo (Cabang Kediri)," *Gener. J.*, vol. 2, no. 1, pp. 9–18, 2018.
- [10] D. Sugiyono, "Metodologi Penelitian Bisnis," *Alf. Bandung*, 2010.

BIODATA PENULIS



Reza Hikmatulloh
Universitas Islam Negeri Syarif
Hidayatullah



Hanna Mahaerani P
Universitas Islam Negeri Syarif
Hidayatullah



Qurrotul Aini
Universitas Islam Negeri Syarif
Hidayatullah