

Penerapan Algoritma C4.5 untuk Prediksi Kepuasan Penumpang Bus Rapid Transit (BRT) Trans Semarang

Anggita Safitri Febriarini¹, Erna Zuni Astuti²

Program Studi Teknik Informatika, Universitas Dian Nuswantoro

Jalan Imam Bonjol 207 Semarang, (+6224) 3517261

e-mail: ¹anggitasafitri31@gmail.com, ²erna.zuni.astuti@dsn.dinus.ac.id

Abstrak

Sejumlah aktivitas perkotaan yang ada di Kota Semarang menjadikan tingginya peningkatan yang tentunya membutuhkan dukungan sistem angkutan umum yang andal, cepat, dan efisien. Sebagai bagian dari upaya memecahkan masalah kemacetan, Pemerintah Pusat melalui Departemen Perhubungan membuat Bus Rapid Transit (BRT) yang saat ini diterapkan di berbagai kota di Indonesia. Konsep BRT merupakan sistem angkutan massal yang terintegrasi di setiap koridor, yang bertujuan untuk memenuhi kebutuhan masyarakat akan transportasi dalam kota. Tujuan penelitian ini adalah untuk mengetahui dan menganalisis kepuasan penumpang BRT Trans Semarang dengan menggunakan teknik data mining dengan metode decision tree tepatnya algoritma C4.5. Dalam penelitian ini variabel pengukuran kepuasan penumpang yang digunakan adalah harga, fasilitas, dan pelayanan. Dari penelitian ini atribut yang digunakan menghasilkan hubungan sebab-akibat dalam mengklasifikasikan penumpang puas dan tidak puas. Penelitian ini diharapkan dapat membantu Pemerintah dalam meningkatkan kepuasan penumpang agar masyarakat beralih menggunakan BRT Trans Semarang dan dapat memberikan pengaruh terjadinya peningkatan jumlah penumpang. Berdasarkan hasil klasifikasi dengan menggunakan algoritma C4.5 menunjukkan adanya peningkatan di tiap pengujian yang telah dilakukan dengan hasil akurasi akhir yang cukup baik sebesar 95% yang menunjukkan bahwa algoritma C4.5 cocok digunakan untuk mengukur tingkat kepuasan penumpang BRT Trans Semarang.

Kata kunci: Data Mining, Decision Tree, Algoritma C4.5, Kepuasan Penumpang.

Abstract

A number of urban activities in Semarang City make a high increase, which of course require the support of a public transport system that is reliable, fast, and efficient. As part of the effort to solve the congestion problem, the Central Government through the Transportation Department made Bus Rapid Transit (BRT), which is currently applied in various cities in Indonesia. The BRT concept is an integrated mass transit system in every corridor, which aims to meet the needs of the urban transport community. The purpose of this research is to know and analyze passenger satisfaction of BRT Trans Semarang by using data mining technique with decision tree method exactly C4.5 algorithm. In this study, the variable measurement of passenger satisfaction used is price, facilities, and service. From this study, the attributes used to result in causal relationships in classifying passengers satisfied and not satisfied. This research is expected to assist the Government in increasing passenger satisfaction so that people switch to using BRT Trans Semarang and can give influence to the increase of passenger. Based on the result of classification using C4.5 algorithm showed an increase in each test that has been done with a good accurate end result of 95% indicating that C4.5 algorithm suitable to be used to measure the level of passenger satisfaction BRT Trans Semarang.

Keywords: Data Mining, Decision Tree, Algorithm C4.5, Passenger Satisfaction.

1. Pendahuluan

Kota Semarang terletak di Provinsi Jawa Tengah dengan aktivitas perkotaan yang besar. Aktivitas yang dilakukan manusia untuk memenuhi kebutuhan dan keperluan hidup sehari-hari seperti bekerja, bersekolah, berbelanja, dan berjalan-jalan di objek wisata terkenal di Kota Semarang. Berdasarkan data dari Badan Pusat Statistika Kota Semarang, peningkatan jumlah kendaraan pribadi hingga tahun 2014 sebanyak 247.936 untuk sepeda motor dan 56.453 mobil pribadi atau dinas. Adanya peningkatan jumlah kendaraan bermotor dapat menimbulkan kemacetan. Penyebaran penduduk juga masih belum merata, masih terpusat

di kota sedangkan masyarakat yang di pinggiran kota membutuhkan akses transportasi untuk melakukan segala aktivitasnya. Maka Pemerintah khususnya Dinas Perhubungan Kota Semarang meluncurkan Trans Semarang untuk memberikan layanan transportasi yang lebih cepat dan efisien. Adanya bus Trans Semarang ini diharapkan dapat menjadi solusi transportasi yang tepat dan jauh lebih baik dari angkutan kota yang ada dan mampu beralih dari menggunakan kendaraan pribadi menggunakan transportasi publik [1].

Bus Trans Semarang diluncurkan oleh Pemerintah Kota Semarang pada 2 Mei 2009. Bus Trans Semarang adalah layanan transportasi massal berbasis Bus Rapid Transit (BRT) yang beroperasi di Semarang. Bus Trans Semarang memiliki tempat duduk yang memanjang dan saling berhadapan, dan di dalamnya menggunakan *air conditioner* (AC) yang membuat penumpang merasa nyaman tidak merasakan kepanasan seperti bus-bus lain pada umumnya. Bus Trans Semarang dapat menampung penumpang dalam jumlah yang cukup banyak dan memiliki beberapa koridor. Bus Trans Semarang direncanakan akan memiliki 12 koridor, tetapi untuk saat ini Bus Trans Semarang hanya memiliki 6 koridor yang sedang beroperasi [2].

Penumpang bus Trans Semarang dapat membeli tiket di halte transit, bila di halte selain halte transit pembelian tiket dapat dibeli saat di dalam bus Trans. Pembelian tiket dibedakan menjadi 2 yaitu Rp 1.000,00 untuk pelajar dan Rp.3.500,00 untuk umum. Pengelola bus Trans Semarang adalah Badan Layanan Umum (BLU) yang sebelumnya dikelola oleh Unit Pelaksana Teknis Dinas Daerah (UPTD) Terminal Mangkang. Bus Trans Semarang beroperasi setiap hari mulai jam 05.30 sampai pukul 18.30 WIB [2].

Penanganan keluhan saling berkaitan dengan kepuasan konsumen, maka sangatlah penting untuk diperhatikan, diperbaiki, dan dipahami oleh pihak Pemerintah. Kepuasan konsumen akan tercapai apabila kinerja yang diberikan sesuai dengan apa yang diharapkan. Tetapi, apabila kinerja tidak sesuai dengan yang diharapkan maka kepuasan konsumen tidak dapat dicapai. Kepuasan konsumen sangat perlu dipertahankan dan ditingkatkan. Maka, kepuasan konsumen merupakan awal dari terciptanya kepercayaan konsumen terhadap Pemerintah. Selain itu, untuk meningkatkan kepuasan, Pemerintah harus memberikan mereka pelayanan yang lebih dari yang diharapkan, sehingga konsumen dapat bertahan [3].

Seorang konsumen akan menilai tingkat kepuasan yaitu dengan melihat baik atau tidaknya kualitas dari sebuah pelayanan. Konsumen akan membandingkan pelayanan yang mereka terima. Pada umumnya kepuasan tiap konsumen berbeda-beda, tingkat kepuasan yang berbeda ini adalah indikator yang baik untuk mengukur tingkat kualitas pelayanan yang diberikan. Ada salah satu cara untuk meningkatkan kepuasan konsumen yaitu berusaha untuk menjalin hubungan yang baik dengan konsumen dan juga mau mendengar semua keluhan konsumen tentang pelayanan yang diberikan kepada kepuasan jasa tersebut. Konsumen yang merasa puas dengan pelayanan jasa yang diberikan kepada Pemerintah maka konsumen akan datang kembali. Sehingga konsumen yang merasa puas akan berbagi pengalaman tentang pelayanan yang diterima kepada konsumen lain [4].

Algoritma C4.5 merupakan salah satu contoh metode untuk membuat pohon keputusan dan tepat digunakan untuk mengklasifikasi data yang sangat besar ke dalam *class* tertentu menurut bentuk data yang ada. Untuk *data mining* dan pembelajaran mesin (*machine learning*), algoritma C4.5 digunakan untuk mempelajari data dengan jumlah yang besar dan menggunakan model pembelajaran berupa pohon keputusan yang bisa dipraktikkan untuk memprediksi data yang belum muncul. Pohon keputusan adalah alternatif pemecahan masalah yang diambil dari masalah tersebut [5].

Ade Yuliana, Duwi Bayu Pratomo [6] dengan penelitian algoritma *decision tree* (C4.5) untuk memprediksi kepuasan Mahasiswa terhadap kinerja dosen Politeknik TEDC Bandung dan dari keseluruhan hasil pengujian menggunakan algoritma C4.5 memperoleh tingkat akurasi sebesar 94,62%. Selanjutnya Eni Irfani [7] dengan penelitian prediksi keluhan pelanggan pada apartemen di Jakarta dan dari hasil validasi prediksi diperoleh sebesar 75%. Terakhir Fandi Ramadhan Wiratama [8] dengan meneliti tentang analisis kepuasan pelanggan Bank Danamon Kantor Cabang Utama Semarang kepada nasabah menggunakan Algoritma C4.5 dengan sampel data yang digunakan untuk menguji penelitian ini adalah sebanyak 150 data yang didapat dari menyebar kuesioner dan memperoleh hasil dengan tingkat akurasi sebesar 83%.

Kepuasan setiap orang berbeda-beda ada yang merasa puas dan merasa tidak puas. Agar mengetahui apakah Pemerintah yang sudah memberikan fasilitas transportasi umum seperti Bus Trans Semarang sudah sesuai dengan keinginan masyarakat, maka perlu dilakukan survei untuk mengetahui kepuasan setiap orang dari pelayanan yang sudah diberikan. Dari pelayanan yang sudah diberikan akan menjadi tolak ukur kepuasan seseorang. Oleh karena itu berdasarkan permasalahan di atas, maka peneliti mengambil judul: “Penerapan Algoritma C4.5 Untuk Prediksi Kepuasan Penumpang Bus Rapid Transit (BRT) Trans Semarang”.

2. Metode Penelitian

2.1. Metode Pengumpulan Data

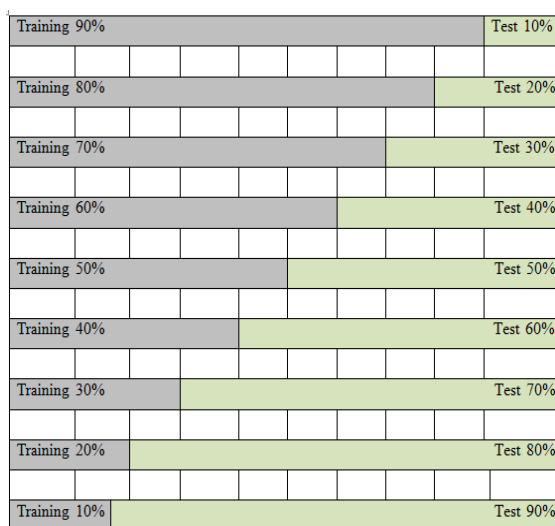
Penelitian dilakukan pada bulan November - Januari 2018. Bus Trans Semarang dipilih sebagai objek penelitian, karena Bus Trans Semarang merupakan jenis transportasi umum yang sudah beroperasi selama 8 tahun dan masih bertahan di tengah munculnya jenis transportasi berbasis *online* seperti Gojek dan Grab. Dan untuk pengumpulan data penumpang Bus Trans Semarang menggunakan metode kuisisioner.

Melalui metode ini diharapkan peneliti mendapatkan data-data yang relevan, akurat, dan terpercaya. Kuisisioner adalah suatu teknik pengumpulan informasi dari serangkaian pertanyaan tertulis yang harus diisi oleh responden. Pada penelitian ini peneliti membagikan sebanyak 200 kuisisioner yang dibagikan kepada responden yaitu penumpang Bus Trans Semarang. Kuisisioner dibagikan berupa pertanyaan dengan pilihan jawaban-jawaban yang disediakan oleh peneliti yang digunakan untuk memperoleh data dari responden. Beberapa atribut yang terdapat di kuisisioner adalah tarif pada Bus Trans, fasilitas yang diberikan, dan pelayanan dari petugas Bus Trans.

2.2. Teknik Pengujian Instrumen

- Uji Validasi

Metode yang digunakan untuk uji validasi menggunakan *Split Validation*. *Split validation* merupakan teknik validasi untuk membagi data *training* dan data *testing*. Dengan melakukan percobaan *training* berdasarkan *Split Ratio* yang sudah ditentukan sebelumnya dengan menggunakan *Split Validation*, kemudian sisa dari *Split Ratio* data *testing* adalah data yang akan menjadi data pengujian kebenaran atau keakurasian hasil pembelajaran dan belum pernah digunakan dalam pembelajaran, ditunjukkan pada Gambar 1.



Gambar 1. Ilustrasi *split validation*.

2.3. Evaluasi

Setelah data diuji menggunakan RapidMiner, maka akan terbentuk pohon keputusan dari data tersebut. Kemudian akan dilakukan penghitungan akurasi, *precision*, *recall*, selanjutnya dilakukan evaluasi hasil dihitung menggunakan rumus akurasi, *precision* berikut ini:

$$\text{Akurasi} = \frac{TP+TN}{(TP+FP+TN+FN)} \tag{1}$$

$$\text{Precision} = \frac{TP}{(TP+FN)} \tag{2}$$

$$\text{Recall} = \frac{TP}{(TP+FP)} \tag{3}$$

Keterangan:

TP = *True Positive*

FP = *False Positive*
 TP = *True Negative*
 FN = *False Negative*

2.4. Teknik Analisis Data

Untuk penelitian ini data yang akan digunakan adalah data dengan membuat kuesioner yang dibagikan kepada 200 penumpang Bus Trans Semarang dari survei yang dilakukan langsung di lapangan. Dan metode yang digunakan adalah metode algoritma C4.5.

Tahap penelitian kepuasan penumpang Bus Trans Semarang:

1. Penulis membuat kuesioner yang sudah ditentukan yaitu sebanyak 200 kuesioner dan dibagikan kepada para penumpang Bus Trans Semarang. Untuk menentukan jumlah puas dan tidak puas pada persentase kepuasan penumpang Bus Trans Semarang, kuesioner tersebut akan dijadikan sebagai data awal.
2. Data-data akan diolah dengan menggunakan Microsoft Excel untuk mendapatkan rata-rata dari setiap atributnya.
3. Menghitung hasil data tersebut menggunakan rumus algoritma C4.5 untuk memperoleh pohon keputusan.
4. Menyiapkan RapidMiner yang akan digunakan untuk menghitung data kuesioner dengan menggunakan algoritma C4.5.
5. Pengujian dilakukan sebanyak 3 kali dengan tingkat acuan pengujian yang berbeda, untuk mengetahui keakuratan dan hasil dari perhitungan menggunakan RapidMiner.
6. Dilakukan validasi dan mengukur keakuratan hasil yang dicapai oleh model menggunakan beberapa teknik yang ada pada *framework* RapidMiner yaitu dengan menghitung nilai akurasi, *precision*, dan *recall* untuk mengukur tingkat akurasi model dan *split validation* untuk validasi.
7. Dari penelitian ini diperoleh hasil agar bisa diketahui sejauh mana kepuasan penumpang Bus Trans Semarang.

Mengimplementasikan *rule* yang sudah didapat dari pengolahan data RapidMiner dengan menggunakan Sublime Text untuk membuat program.

3. Hasil dan Pembahasan

3.1. Pengolahan Data

Data yang dikumpulkan adalah data kuesioner yang dibagikan kepada penumpang Bus Trans Semarang pada bulan November – Januari 2018 dari 200 data dengan atribut nama, jenis kelamin, umur, harga, fasilitas, pelayanan, kepuasan.

Jika data sudah diolah menggunakan Microsoft Excel, maka pengolahan data kuesioner akan dilakukan kembali dengan mengambil atribut nilai rata-rata Harga, Fasilitas, Pelayanan, dan Hasil, seperti gambar 2 di bawah ini:

| 1 | NAMA | Harga | Fasilitas | Pelayanan | Hasil |
|----|-------------------|-------|-----------|-----------|------------|
| 2 | Sugiyono | 5 | 3,6 | 4 | Puas |
| 3 | Fauzan T. | 4 | 3,8 | 4,3 | Puas |
| 4 | Aliba Arazi | 2 | 2 | 2,7 | Tidak Puas |
| 5 | Thalita Putra | 3 | 2 | 2,2 | Tidak Puas |
| 6 | Inge Satna | 5 | 4,2 | 4,4 | Puas |
| 7 | Brilia Crisping | 5 | 3,4 | 3,5 | Puas |
| 8 | Sariyatun | 4 | 4,2 | 3,9 | Puas |
| 9 | Satyana Listyanti | 1 | 1,4 | 2,4 | Tidak Puas |
| 10 | Anjas Setiawan | 5 | 4,6 | 4 | Puas |
| 11 | Rizky Putra | 4 | 3,8 | 3,9 | Puas |
| 12 | Sarah Amalia | 5 | 3,6 | 3,5 | Puas |
| 13 | Naya Maharani | 4 | 4 | 4 | Puas |
| 14 | Rycho Purnama | 4 | 4,6 | 4,2 | Puas |
| 15 | Reni Puji Lestari | 4 | 3,6 | 4,1 | Puas |
| 16 | Naufal Bagus | 1 | 1,4 | 1,8 | Tidak Puas |
| 17 | Partejo | 2 | 2 | 1,9 | Tidak Puas |
| 18 | Nizar | 4 | 4 | 3,8 | Puas |
| 19 | Agus Hendry | 4 | 4,4 | 4,1 | Puas |
| 20 | Siti Atikah | 5 | 4 | 4,3 | Puas |
| 21 | Bayu Fajar | 4 | 3,2 | 3,9 | Puas |
| 22 | Novia Nazwa | 5 | 4,8 | 4,7 | Puas |

Gambar 2. Setelah pengambilan atribut.

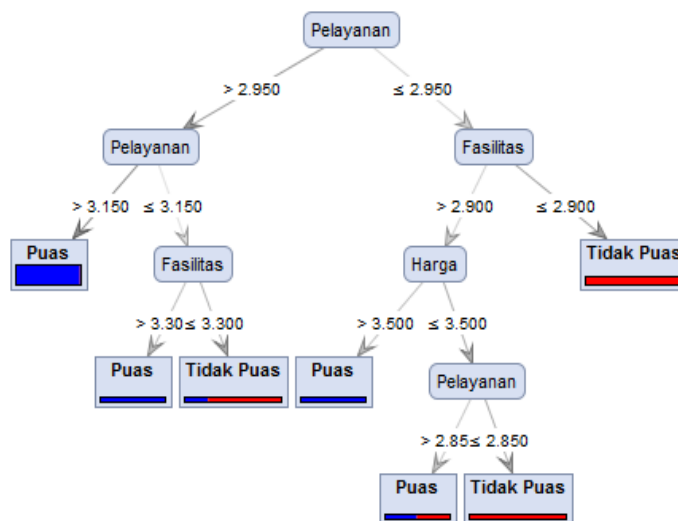
3.2. Proses Data Mining pada RapidMiner

Selanjutnya data akan diproses menggunakan *tools* RapidMiner dan untuk pengujiannya menggunakan *Split Validation* yang terdapat pada RapidMiner, maka data akan dapat dilihat dari pohon keputusan, nilai akurasi, dan *rule*. Pada Gambar 3 data yang akan di-*import* ke RapidMiner sebagai berikut:

| No | NAMA | Harga | Fasilitas | Pelayanan | Hasil |
|----|----------------|-------|-----------|-----------|------------|
| 1 | Sugiyono | 5 | 3.600 | 4 | Puas |
| 2 | Fauzan T. | 4 | 3.800 | 4.300 | Puas |
| 3 | Aliba Arazi | 2 | 2 | 2.700 | Tidak Puas |
| 4 | Thalita Putri | 3 | 2 | 2.200 | Tidak Puas |
| 5 | Inge Satna | 5 | 4.200 | 4.400 | Puas |
| 6 | Brilia Crispir | 5 | 3.400 | 3.500 | Puas |
| 7 | Sariyatun | 4 | 4.200 | 3.900 | Puas |
| 8 | Satyana List | 1 | 1.400 | 2.400 | Tidak Puas |
| 9 | Anjas Setiaw | 5 | 4.600 | 4 | Puas |
| 10 | Rizky Putra | 4 | 3.800 | 3.900 | Puas |
| 11 | Sarah Amali | 5 | 3.600 | 3.500 | Puas |
| 12 | Naya Mahar | 4 | 4 | 4 | Puas |
| 13 | Rycho Purne | 4 | 4.600 | 4.200 | Puas |
| 14 | Reni Puji Le | 4 | 3.600 | 4.100 | Puas |
| 15 | Naufal Bagu | 1 | 1.400 | 1.800 | Tidak Puas |

Gambar 3. Data yang di-*import* ke RapidMiner.

Setelah melakukan perhitungan dengan metode algoritma C4.5 pada RapidMiner, maka akan membentuk pohon keputusan seperti Gambar 4:



Gambar 4. Pohon keputusan hasil pengolahan dengan RapidMiner.

Pohon keputusan di atas akan memperoleh *rule* yang nantinya diimplementasikan dalam program. Di bawah ini adalah *rule* yang diperoleh pada Tabel 1 sebagai berikut:

Tabel 1. *Rules*.

| Rule untuk Hasil = PUAS |
|---|
| 1. IF nilai pelayanan >2.950 and nilai pelayanan >3.150 THEN hasil = PUAS |
| 2. IF nilai pelayanan >2.950 and nilai pelayanan <=3.150 and nilai fasilitas >3.300 THEN hasil = PUAS |
| 3. IF nilai pelayanan <=2.950 and nilai fasilitas >2.900 and nilai harga >3.500 THEN hasil = PUAS |
| 4. IF nilai pelayanan <=2.950 and nilai fasilitas >2.900 and nilai harga <=3.500 and nilai pelayanan >2.850 THEN hasil = PUAS |
| Rule untuk Hasil = TIDAK PUAS |
| 1. IF nilai pelayanan >2.950 and nilai pelayanan <=3.150 and nilai fasilitas <=3.300 THEN hasil = TIDAK PUAS |
| 2. IF nilai pelayanan <=2.950 and nilai fasilitas <=2.900 THEN hasil = TIDAK PUAS |

3. IF nilai pelayanan <=2.950 and nilai fasilitas >2.900 and nilai harga <=3.500 and nilai pelayanan <=2.850 THEN hasil = TIDAK PUAS

3.3. Pengujian

Untuk mengetahui apakah solusi yang diperoleh dari pohon keputusan tersebut valid atau tidak, maka akan dilakukan pengujian pada validitas sistem.

Data set dibagi menjadi 2 bagian yaitu data *training* dan data *testing*, pengujian akan menggunakan metode *confusion matrix* dengan perbandingan yang berbeda antara data *training* dan data *testing*. Berikut uraian pengujian yang dilakukan 3 kali :

1. Data *training* 60% dan data *testing* 40% (120 dan 80)
2. Data *training* 70% dan data *testing* 30% (140 dan 60)
3. Data *training* 80% dan data *testing* 20% (160 dan 40)

3.3.1. Pengujian 1

Pengujian pertama menggunakan perbandingan 60% data *training* dan 40% data *testing* dari keseluruhan data set yang berjumlah 200 sehingga diperoleh 120 untuk data *training* dan 80 untuk data *testing*. Dari pengujian ini nantinya akan diperoleh nilai dari akurasi, *recall*, dan *precision*.

Hasil akurasi dari *confusion matrix* untuk pengujian 1 pada Tabel 2 sebagai berikut:

Tabel 2. *Confusion matrix* pengujian 1.

| | Actual | |
|------------------|--------|------------|
| | Puas | Tidak Puas |
| Pred. Puas | 61 | 5 |
| Pred. Tidak Puas | 1 | 13 |

Dari tabel di atas diperoleh jumlah *true positive* (tp) sebanyak 61 *record*, *false positive* (fp) sebanyak 5 *record*, jumlah *true negative* (tn) sebanyak 13 *record*, dan *false negative* (fn) sebanyak 1 *record*, ditunjukkan pada Tabel 3.

$$\begin{aligned}
 - \text{ Akurasi} &= \frac{TP+TN}{(TP+TN+FP+FN)} \times 100\% \\
 &= \frac{61+13}{(61+13+1+5)} \times 100\% \\
 &= \frac{74}{80} \times 100\% \\
 &= 92,5\%
 \end{aligned}$$

$$\begin{aligned}
 - \text{ Precision} &= \frac{TP}{(TP+FP)} \times 100\% \\
 &= \frac{61}{(61+5)} \times 100\% \\
 &= \frac{61}{66} \times 100\% \\
 &= 92,42\%
 \end{aligned}$$

$$\begin{aligned}
 - \text{ Recall} &= \frac{TP}{(TP+FN)} \times 100\% \\
 &= \frac{61}{(61+1)} \times 100\% \\
 &= \frac{61}{62} \times 100\% \\
 &= 98,39\%
 \end{aligned}$$

Tabel 3. Evaluasi dan validasi 1.

| Persentase Data | Data Training | Data Testing | Akurasi | Precision | Recall |
|-----------------|---------------|--------------|---------|-----------|--------|
| 60% : 40% | 120 | 80 | 92,5% | 92,42% | 98,39% |

3.3.2. Pengujian 2

Pengujian kedua menggunakan perbandingan 70% data *training* dan 30% data *testing* dari keseluruhan data set yang berjumlah 200 sehingga diperoleh 140 untuk data *training* dan 60 untuk data *testing*. Dari pengujian ini nantinya akan diperoleh nilai dari akurasi, *recall*, dan *precision*.

Hasil akurasi dari *confusion matrix* untuk pengujian 2 pada Tabel 4 sebagai berikut:

Tabel 4. *Confusion matrix* pengujian 2.

| | Actual | |
|------------------|--------|------------|
| | Puas | Tidak Puas |
| Pred. Puas | 46 | 3 |
| Pred. Tidak Puas | 1 | 10 |

Dari tabel di atas diperoleh jumlah *true positive* (tp) sebanyak 46 *record*, *false positive* (fp) sebanyak 1 *record*, jumlah *true negative* (tn) sebanyak 10 *record*, dan *false negative* (fn) sebanyak 3 *record*, ditunjukkan pada Tabel 5.

- Akurasi = $\frac{TP+TN}{(TP+TN+FP+FN)} \times 100\%$
 $= \frac{46+10}{(46+10+1+3)} \times 100\%$
 $= \frac{56}{60} \times 100\%$
 $= 93,33\%$
- Precision = $\frac{TP}{(TP+FN)} \times 100\%$
 $= \frac{46}{(46+3)} \times 100\%$
 $= \frac{46}{49} \times 100\%$
 $= 93,88\%$
- Recall = $\frac{TP}{(TP+FP)} \times 100\%$
 $= \frac{46}{(46+1)} \times 100\%$
 $= \frac{46}{47} \times 100\%$
 $= 97,87\%$

Tabel 5. Evaluasi dan validasi 2.

| Persentase Data | Data Training | Data Testing | Akurasi | Precision | Recall |
|-----------------|---------------|--------------|---------|-----------|--------|
| 70% : 30% | 140 | 60 | 93,33% | 93,88% | 97,87% |

3.3.3. Pengujian 3

Pengujian ketiga menggunakan perbandingan 80% data *training* dan 20% data *testing* dari keseluruhan data set yang berjumlah 200 sehingga diperoleh 160 untuk data *training* dan 40 untuk data *testing*. Dari pengujian ini nantinya akan diperoleh nilai dari akurasi, *recall*, dan *precision*.

Hasil akurasi dari *confusion matrix* untuk pengujian 3 pada Tabel 6 sebagai berikut:

Tabel 6. *Confusion matrix* pengujian 3.

| | Actual | |
|------------------|--------|------------|
| | Puas | Tidak Puas |
| Pred. Puas | 29 | 1 |
| Pred. Tidak Puas | 1 | 19 |

Dari tabel di atas diperoleh jumlah *true positive* (tp) sebanyak 29 *record*, *false positive* (fp) sebanyak 1 *record*, jumlah *true negative* (tn) sebanyak 9 *record*, dan *false negative* (fn) sebanyak 1 *record*, ditunjukkan pada Tabel 7.

- Akurasi = $\frac{TP+TN}{(TP+TN+FP+FN)} \times 100\%$
 $= \frac{29+9}{(29+9+1+1)} \times 100\%$
 $= \frac{38}{40} \times 100\%$
 $= 95\%$
- Precision = $\frac{TP}{(TP+FN)} \times 100\%$

$$\begin{aligned}
 &= \frac{29}{(29+1)} \times 100\% \\
 &= \frac{29}{30} \times 100\% \\
 &= 96,67\% \\
 \\
 - \text{ Recall} &= \frac{TP}{(TP+FP)} \times 100\% \\
 &= \frac{29}{(29+1)} \times 100\% \\
 &= \frac{29}{30} \times 100\% \\
 &= 96,67\%
 \end{aligned}$$

Tabel 7. Evaluasi dan validasi 3.

| Persentase Data | Data Training | Data Testing | Akurasi | Precision | Recall |
|-----------------|---------------|--------------|---------|-----------|--------|
| 80% : 20% | 160 | 40 | 95% | 96,67% | 96,67% |

3.4. Analisa Hasil Pengujian

Pada pengujian 1, 2, dan 3 memperoleh akurasi, *precision*, dan *recall* pada Tabel 8 sebagai berikut:

Tabel 8. Hasil pengujian 1, 2, dan 3.

| Persentase Data | Data Training | Data Testing | Akurasi | Precision | Recall |
|-----------------|---------------|--------------|---------|-----------|--------|
| 60% : 40% | 120 | 80 | 92,5% | 92,42% | 98,39% |
| 70% : 30% | 140 | 60 | 93,33% | 93,88% | 97,87% |
| 80% : 20% | 160 | 40 | 95% | 96,67% | 96,67% |

Berdasarkan tabel di atas, maka dapat diketahui dengan perbandingan data *training* dan data *testing* 60% : 40% adalah nilai akurasi 92,5%, *precision* 92,42% dan *recall* 98,39%, dan perbandingan 70% : 30% adalah nilai akurasi 93,33%, *precision* 93,88%, dan *recall* 97,87%, sedangkan perbandingan 80% : 20% adalah nilai akurasi 95%, *precision* 96,67%, dan *recall* 96,67%.

4. Kesimpulan

Hasil dari penelitian dan peneliti melakukan pengujian sebanyak 3 kali maka dapat disimpulkan sebagai berikut:

1. Dengan pengujian 3 kali yang sudah dilakukan, maka dapat diketahui bahwa pengujian 1 dengan perbandingan data *training* dan data *testing* 60% : 40% adalah nilai akurasi 92,5%, *precision* 92,42% dan *recall* 98,39%, dan pengujian 2 dari perbandingan 70% : 30% adalah nilai akurasi 93,33%, *precision* 93,88%, dan *recall* 97,87%, sedangkan pengujian 3 dengan perbandingan 80% : 20% adalah nilai akurasi 95%, *precision* 96,67%, dan *recall* 96,67% dapat dikatakan berhasil dan baik, karena bisa terlihat dengan jelas bahwa nilai akurasi yang terus bertambah dan semakin akurat.
2. Penelitian ini berkontribusi pada penerapan *Smart City* di Semarang, dalam hal ini adalah sebagai salah satu usaha peningkatan fasilitas dan pelayanan publik, terutama dalam bidang transportasi pada Trans Semarang.

5. Saran

Selama melakukan penelitian data penumpang Bus Trans Semarang saran yang dapat disampaikan antara lain:

1. Penelitian ini bisa dikembangkan dengan menggabungkan atau membandingkan dengan algoritma klasifikasi lain untuk memperoleh hasil prediksi yang lebih baik.
2. Sebaiknya jumlah data kuesioner ditambah, supaya memperoleh hasil akurasi yang lebih baik.

Daftar Pustaka

- [1] Badan Pusat Statistik. (2017) Banyaknya Kendaraan Bermotor dirinci menurut Jenis Kendaraan tahun 2012-2014. [Online]. Available: <https://www.bps.go.id/>
- [2] A. Susanty, S. Nugroho, K. A. Khantari, "Penyusunan Skenario Kebijakan Untuk Pengembangan Bus Rapid Transit (BRT) Trans Semarang Dengan Pendekatan Sistem Dinamik," vol. 1, no. 18, pp. 17-26, 2014.

-
- [3] H. Dhika, F. Destiwati, A. Fitriansyah, "Implementasi Algoritma C4.5 Terhadap Kepuasan Pelanggan," *Prosiding SNaPP Sains dan Teknologi*, vol. 6, no. 1, 2016.
 - [4] M. K. Rizal, Suhartini, J. Sugiantara, "Penerapan Algoritma C4.5 Untuk Prediksi Tingkat Kepuasan Mahasiswa Terhadap Sistem Informasi Akademik Sekolah Tinggi Keguruan dan Ilmu Pendidikan Hamzanwadi Selong," *Jurnal Informatika Hamzanwadi*, vol. 2, no. 1, pp. 48-62, 2017.
 - [5] M. R. I. Saputra, "Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Kepuasan Pelanggan Taksi Kosti," pp. 1-14, 2016.
 - [6] A. Yuliana, D. B. Pratomo, "Algoritma C4.5 Untuk Prediksi Kepuasan Mahasiswa Terhadap Kinerja Dosen Politeknik TEDC Bandung," in *Seminar Nasional Inovasi Teknologi UN PGRI Kediri*, Kediri, 2017, pp. 377-384.
 - [7] E. Irfani, "Prediksi Keluhan Pelanggan Pada Apartemen Menggunakan Algoritma C4.5," *Jurnal Paradigma*, vol. XVI, no. 2, pp. 13-20. 2014.
 - [8] F. R. Wiratama, "Analisis Kepuasan Pelanggan Bank Danamon Kantor Cabang Utama Semarang," 2016.