

# IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA C4.5 UNTUK PREDIKSI KEPUASAN PELANGGAN TAKSI KOSTI

Mochamad Rizki Ilham S<sup>1</sup>, Purwanto<sup>2</sup>

Program Studi Teknik Informatika Fakultas Ilmu Komputer

Universitas Dian Nuswantoro Semarang

Jl.Nakula I No 5-11 Semarang 50131

Telp : (024) 3517361, Fax : (024) 3520165

Email : [Rizkiilham23@gmail.com](mailto:Rizkiilham23@gmail.com)

## Abstrak

Banyak cara yang dilakukan oleh masing-masing perusahaan penyedia jasa, khususnya transportasi untuk memenangkan persaingan, antara lain dengan meningkatkan kepuasan pelanggan transportasi. Salah satu perusahaan penyedia jasa transportasi yang berkomitmen dalam meningkatkan kepuasan pelanggannya adalah Taksi KOSTI. Penelitian ini bertujuan untuk menganalisis kepuasan pelanggan Taksi KOSTI dengan menggunakan teknik data mining dengan algoritma C4.5. Atribut masukan kepuasan pelanggan dalam penelitian ini mencakup harga, fasilitas, pelayanan dan loyalitas. Dalam penelitian ini, didapatkan bahwa hasil yang didapatkan berasal dari beberapa atribut masukan menghasilkan hubungan sebab-akibat dalam mengklasifikasikan konsumen puas dan tidak puas. Penelitian ini diharapkan dapat membantu pihak KOSTI dalam meningkatkan kepuasan konsumen untuk mempertahankan pelanggan dan meningkatkan laba perusahaan taksi KOSTI tersebut. Berdasarkan hasil klasifikasi menggunakan algoritma C4.5 menunjukkan bahwa diperoleh akurasi mencapai 88.01%, yang menunjukkan bahwa algoritma C4.5 cocok digunakan untuk mengukur tingkat kepuasan pelanggan taksi KOSTI.

Kata Kunci : Decision Tree, Data mining, Algoritma C4.5, Perusahaan Jasa Transportasi. Kepuasan Pelanggan.

## Abstract

Many methods are used by each service company, especially transport to win the competition, such as increasing customer satisfaction transport. One firm transportation service company are committed to improving their customer satisfaction is Kosti Taxi. This study aims to analyze customer satisfaction Kosti Taxi using data mining techniques with C4.5 algorithm. Atribut input customer satisfaction in this study covers price, facilities, service and loyalty. In this study, it was found that the results obtained from several attributes of the input generates cause-effect relationship in classifying consumers satisfied and dissatisfied. This research is expected to help the Kosti in increasing customer satisfaction to retain customers and increase profits Kosti taxi company. Based on the results of the classification algorithm C4.5 shows that the obtained accuracy of 88.01%, which indicates that the algorithm C4.5 suitable for measuring the level of customer satisfaction Kosti Taxi.

Keywords : Decision Tree, Data mining, Algoritma C4.5, Transportation Service company, Customer Satisfaction.

## PENDAHULUAN

### 1.Latar Belakang

Tingginya aktifitas manusia untuk memenuhi keperluan dan kebutuhan hidup sehari-hari, baik bekerja, belanja, ataupun hanya untuk jalan-jalan melepas rasa penat. Maka, perlu dididukungnya hal ini dengan sarana transportasi yang memadai, baik transportasi umum ataupun transportasi pribadi. Karena pada kenyataannya kendaraan pribadi tidak dimiliki oleh

semua masyarakat. Oleh karena itu sarana transportasi umum sangat dibutuhkan oleh masyarakat, dan hal ini mendorong berdirinya perusahaan jasa yang bergerak dibidang transportasi umum. Saat ini banyak jasa transportasi umum yang dapat digunakan, taksi adalah salah satunya [1].

Taksi merupakan transportasi umum yang memanfaatkan mobil sedan untuk digunakan mengangkut penumpangnya yang dilengkapi dengan radio komunikasi. Perbedaan antara taksi dengan

transportasi umum darat lainnya terletak pada jumlah penumpang yang di angkut. Taksi memiliki kapasitas penumpang sebanyak 4 orang dan 1 sopir. Sedangkan untuk transportasi lain seperti bus bisa mengangkut lebih banyak penumpang [2]. Oleh karena itu taksi juga disebut sebagai jasa transportasi eksklusif, karena taksi dapat memiliki sisi *privacy* yang lebih bagi penumpangnya dibanding dengan transportasi umum darat yang lain.

Seiring dengan kemajuan pendidikan dan perekonomian yang lebih sejahtera, serta perkembangan ilmu pengetahuan dan teknologi, maka hal ini meningkatkan kesadaran masyarakat terhadap unsur – unsur pelayanan yang diberikan oleh perusahaan. Pada saat ini kesadaran konsumen akan perlunya nilai kualitas pelayanan (*Service Quality*) yang diberikan oleh suatu perusahaan, baik dalam bentuk jasa maupun dalam bentuk barang semakin meningkat. Penilaian yang berbeda dari setiap konsumen di dapat dari suatu nilai kualitas pelayanan yang di berikan oleh perusahaan, karena suatu kualitas pelayanan yang baik harus sejalan dengan pelayanan yang diharapkan oleh konsumen [3].

Konsumen adalah orang yang memberikan penilaian/ opini terhadap kualitas pelayanan yang diberikan oleh suatu perusahaan. Konsumen menilai suatu kualitas pelayanan dengan membandingkan pelayanan yang mereka terima dengan pelayanan yang mereka harapkan, Konsumen akan menggunakan jasa perusahaan lagi apabila konsumen merasa puas dengan pelayanan yang diberikan oleh perusahaan. Dengan kata lain perusahaan yang memberikan pelayanan yang dapat memuaskan konsumen akan dapat memiliki konsumen yang loyal. Sehingga secara tidak langsung konsumen yang loyal tersebut akan merekomendasikan kepada konsumen lain dari mulut ke mulut tanda

dipaksa mengenai pelayanan yang pernah diterimanya [1].

Pada saat ini industri jasa transportasi taksi semakin marak di Indonesia. Perusahaan KOSTI (Koperasi Sopir Taksi Indonesia) Semarang adalah salah satu perusahaan yang menggeluti bidang usahanya di sektor industri transportasi yaitu taksi. Sebagai perusahaan swasta tentu KOSTI (Koperasi Sopir Taksi Indonesia) hadir untuk memberikan pelayanan yang terbaik pada konsumen dan sekaligus untuk mendapatkan keuntungan. Tetapi tidak hanya KOSTI (Koperasi Sopir Taksi Indonesia) yang menggeluti di bidang jasa transportasi taksi ini sehingga timbul adanya persaingan antar perusahaan jasa transportasi ini

Berikut adalah tabel daftar perusahaan – perusahaan jasa transportasi taksi di kota Semarang.

**Tabel :** Daftar Perusahaan taksi di kota Semarang

<b>Nama Perusahaan</b>	<b>Alamat Perusahaan</b>
Puri Kencana	Rogojembangan Raya No 68
New Atlas	Telogo Bodas No.1
Kosti Semarang	Pamularsih Raya No 15
Satria Express Perdana	Bumi Rejo 11 No 9
Centries Multi Persada	Majapahit No. 583
Astra Utama Taksi	Telogo Bodas 57 Gajahmungkur
Pandu Persada Sarana Mukti	Kaligawe Km 4/46 Semarang
Blue bird Pusaka	Brigjen Sufiarto No.492

Sumber : Dinas Perhubungan Kota Semarang [1].

Pada Tabel daftar perusahaan taksi diatas menunjukkan banyaknya perusahaan yang menggeluti dibidang jasa transportasi taksidan jumlahnya terus meningkat. Sekarang ini sampai tahun 2015 tercatat Kosti Semarang memiliki anggota 633 orang. Kosti berdiri sejak tahun 1998, dengan memberikan

pelayanan-pelayanan yang baik maka kosti dapat bertahan hingga sekarang [4].

Namun perkembangan yang pesat dalam dunia usaha ini mengharuskan perusahaan taksi menghadapi ketatnya persaingan. Pada umumnya perusahaan melakukan banyak cara untuk mempertahankan pelanggannya selama-lamanya, didalam persaingan usaha yang sangat ketat ini sangat sulit Untuk mewujudkan hal itu mengingat banyaknya perubahan yang dapat terjadi setiap saat. Seperti perubahan pada diri pelanggan. pesaing maupun perubahan kondisi secara luas yang selalu berubah secara dinamis. Hal ini membuat para pengambil kebijakan untuk terus mengembangkan suatu strategi yang dapat mencapai sasaran pertumbuhan penjualan, peningkata porsi pasar perusahaan, serta pencapaian kemampuan sebagai dasar pertumbuhan berkelanjutan [3].

Sementara itu, Data mining adalah proses pengekstrakan informasi dari kumpulan data yang besar [5]. Pengekstrakan informasi dilakukan berdasarkan metode data mining yang akan digunakan. Ada beberapa metode serta algoritma data mining yang digunakan untuk mengekstrak informasi antara lain: metode klasifikasi, metode asosiasi, metode *clustering*, metode prediksi, dan metode estimasi.

Algoritma C4.5 adalah salah satu algoritma dari metode klasifikasi Algoritma C4.5 adalah suatu algoritma yang dapat menghasilkan sebutah pohon keputusan. Pohon keputusan adalah suatu bagian dari metode klasifikasi dan prediksi yang sangat kuat dan terkenal, Metode pohon keputusan melakukan pengolahan/perubahan terhadap suatu data yang besar menjadi suatu pohon keputusan yang didalamnya terdapt aturan – aturan yang dapat dipahami oleh manusia.

Ibnu Fatchur Rohman [6] dalam penelitiannya menggunakan algoritma C4.5 pada daa kepuasan

konsumen DAMRI menghasilkan akurasi sebesar 93%. Selanjutnya Teguh budi santoso [7] melakukan penelitian untuk prediksi loyalitas pelanggan dengan menggunakan algoritma C4.5 dan menunjukkan bahwa diperoleh akurasi mencapai 97,5% yang menunjukkan bahwa algoritma C4.5 cocok digunakan untuk mengukur tingkat loyalitas pelanggan.

David Hartanto Kamagi [8] melaporkan bahwa penelitiannya menggunakan Algoritma C4.5 mendapatkan hasil yang bagus. Dari hasil uji coba terhadap 100 data, peneliti mendapatkan tingkat akurasi dari hasil prediksi kelulusan terhadap data testing sebesar 87.5%. Selanjutnya Anik Andriani [9] Penelitiannya terhadap Klasifikasi mahasiswa Dropout menggunakan Algoritma C4.5 memperoleh hasil yang baik. Hasil evaluasi dan validasi dengan confussion matrix menunjukkan tingkat akurasi pada algoritma C4.5 sebesar 97,75%. Dan Dyah satiti [10] memperoleh hasil yang baik pada penelitian Analisis Kepuasan konsumen pada suatu restoran cepat saji dengan data mining.

Untuk mengetahui bagaimana suatu perusahaan memberikan kualitas jasa yang sesuai dengan apa yang diharapkan oleh pelanggan, perlu dilakukanya evaluasi terhadap pelanggan. untuk mengetahui apakah pelanggan puas atau tidak puas dengan pelayanan yang di berikan. Dari permasalahan yang peneliti paparkan diatas maka peneliti melakukan penelutian dengan judul : “Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Kepuasan Pelanggan Taksi Kosti”.

## METODOLOGI PENELITIAN

### 1. Metode Pengumpulan Data

Penelitian ini dilakukan dengan membagi kuisioner pada bulan oktober s/d November 2015. KOSTI Semarang dipilih sebagai objek penelitian,

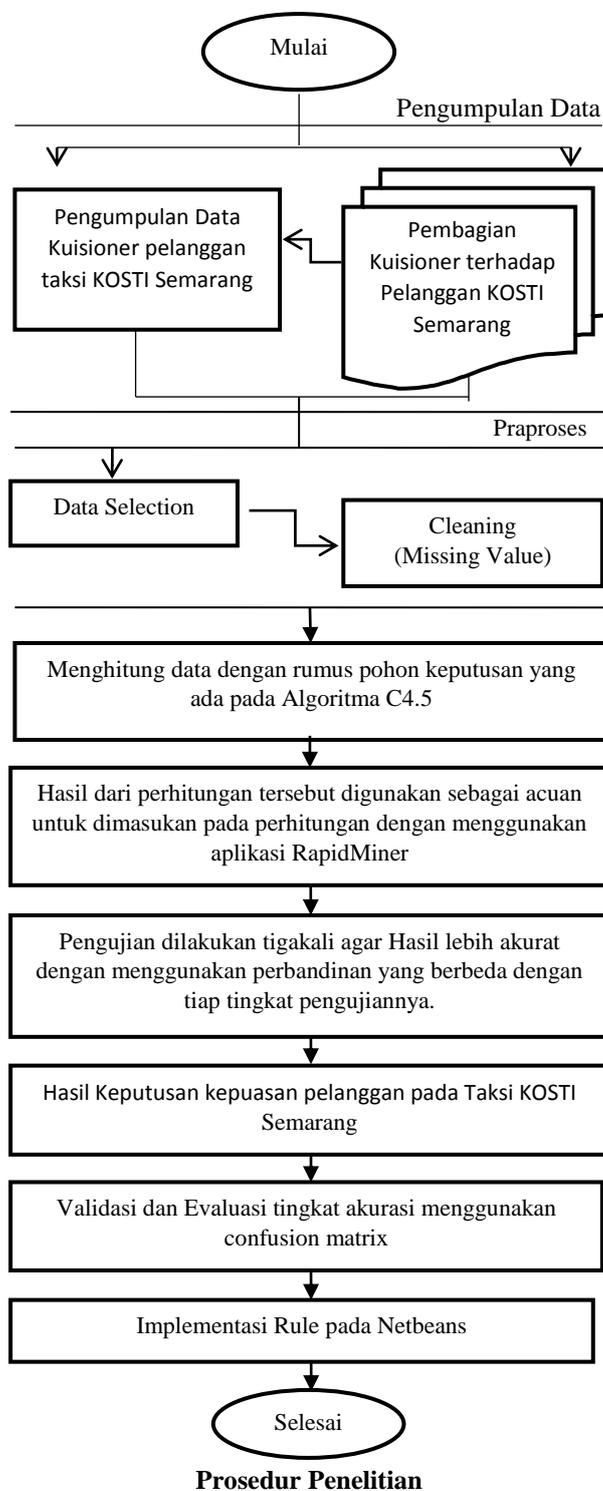
karena perusahaan taksi KOSTI hanya ada di Kota Semarang, dan sudah 17 tahun perusahaan KOSTI berdiri dan bertahan di tengah banyaknya perusahaan taksi yang beridiri di Kota Semarang. Dan hal ini menyebabkan peneliti ingin mengetahui tingkat kepuasan pelanggan taksi KOSTI. Metode yang peneliti gunakan pada pengumpulan data kepuasan pelanggan KOSTI Semarang adalah menggunakan metode kuisisioner.

Melalui metode kuisisioner di harapkan peneliti dapat memperoleh data yang akurat dan terpercaya. Kuisisioner adalah sebuah teknik pengumpulan data dengan cara memberikan seperangkat pertanyaan-pertanyaan tertulis kepada responden untuk di jawab [18]. Pada penelitian ini peneliti membagikan sebanyak 200 kuisisioner yang dibagikan kepada responden yaitu pelanggan KOSTI Semarang. Kuisisioner ini dibagikan menjadi 2 kelompok, Untuk kelompok pertama kuisisioner dibagikan melalui bantuan sopir taksi KOSTI Semarang yang dibagikan langsung kepada penumpang taksi KOSTI, kelompok dua kuisisioner dibagikan oleh peneliti kepada pelanggan - pelanggan yang sudah pernah naik taksi KOSTI Semarang. Kuisisioner terdiri dari 19 Pertanyaan dan 4 Kategori dan jawaban – jawaban yang sudah disediakan oleh peneliti untuk di pilih oleh responden. Untuk Atribut kuisisioner terdiri dari tarif pada argo taksi KOSTI, fasilitas yang di berikan di dalam taksi kepada pelanggan, Pelayanan dari Sopir taksi KOSTI dalam mengantarkan / menjawab pertanyaan pelanggan, Loyalitas dari pelanggan terhadap taksi KOSTI, dan hasil Puas tidak puasnya pelanggan terhadap jasa transportasi taksi KOSTI.

## 2. Analisi Data

Data yang digunakan dalam penelitian ini merupakan data berdasarkan survey yang dilakukan di lapangan dengan kuisisioner yang diberikan melalui sopir taksi kepada pelanggan taksi KOSTI Semarang.

Metode yang digunakan dalam penelitian ini yaitu Algoritma C4.5, prosedur penelitian ini sebagai berikut :



### **Keterangan**

Keterangan dari perosedur penelitian diatas adalah:

1. Penulis membuat kuisisioner dengan jumlah yang dibutuhkan pada penelitian ini, yaitu sebanyak 200 kuisisioner dan selanjutnya data kuisisioner tersebut digunakan sebagai data awal untuk mengetahui jumlah puas dan tidak puas pada data kepuasan pelanggan taksi KOSTI Semarang.
2. Menghitung Hasil dari Kuisisioner yang dibagikan di lapangan untuk mengetahui jumlah puas dan tidak puas pada data kuisisioner pelanggan taksi KOSTI Semarang.
3. Data Selection, memilih himpunan data yang akan digunakan pada penelitian ini yaitu data kuisisioner pelanggan taksi KOSTI.
4. Cleaning, yaitu membersihkan data kuisisioner pelanggan taksi KOSTI dengan menghapus data yang tidak digunakan atau yang tidak memiliki nilai attribute lengkap (missing value).
5. Hasil jumlah dari perhitungan jumlah puas dan tidak puas pada data kuisisioner di masukan atau di olah menggunakan rumus Algoritma C4.5
6. Hasil dari pengolahan atau perhitungan menggunakan Algoritma C4.5
7. Peneliti menggunakan Rapidminer untuk menghitung data kuisisioner menggunakan Algoritma C4.5.
8. Untuk mengetahui tingkat akurasi terhadap hasil perhitungan/ pengolahan menggunakan RapidMiner. Pengujian pada data dilakukan sebanyak 3 kali dengan pembagian data training dan data testing yang berbeda.

9. Hasil dari penelitian ini menghasilkan kputusan kepuasan pelanggan pada taksi KOSTI Semarang.

### 10. Evaluasi dan Validasi

Dalam tahap ini dilakuan validasi dan evaluasi untuk mengukur tingkat akurasi data kepuasan pelanggan menggunakan beberapa teknik yang terdapat dalam *framework* RapidMiner.

### 11. Implementasi

Pembuatan program dengan mengimplementasikan rule yang telah didapat dari pengolahan data RapidMiner dengan menggunakan tools Netbeans.

## **HASIL DAN PEMBAHASAN**

### **1. Proses Pengolahan Data**

Data yang peneliti gunakan dalam penelitian ini adalah data kuisisioner. Data diperoleh dari pembagian kuisisioner sebanyak 200 kuisisioner. Data yang dikumpulkan yaitu data kuisisioner yang dibagikan kepada pelanggan setia taksi KOSTI di kota Semarang pada bulan Oktober – November 2015 dengan attribute nama, jenis kelamin, umur, harga, fasilitas, pelayanan, loyalitas, kepuasan. Kuisisioner tersebut dapat dilihat pada gambar dibawah ini:

**KUISIONER PENELITIAN**  
**KEPUASAN PELANGGAN TERHADAP PELAYANAN**  
**JASA TRANSPORTASI TAKSI KOSTI SEMARANG**

Sehubungan dengan dilakukannya penelitian untuk tugas akhir program sarjana strata I (S1), saya mengharapakan bantuan anda untuk memberikan masukan anda dengan mengisi kuisisioner ini dengan keadaan sebenarnya. Bantuan anda sangat berharga sekali dalam penelitian ini. Terima kasih atas kebaikannya anda yang telah memberikan waktu dan kesediaan menanggapi hal ini

Nama : \_\_\_\_\_

Jenis Kelamin : Laki-Laki  Perempuan

Umur : ..... Tahun

Silahkan kuisisioner ini sesuai dengan penilaian anda, dengan tanda (X)

STS = Sangat Tidak Setuju = 1  
TS = Tidak Setuju = 2  
N = Netral = 3  
S = Setuju = 4  
SS = Sangat Setuju = 5

**1. HARGA**

NO	PERTANYAAN	STS (1)	TS (2)	N (3)	S (4)	SS (5)
1.	Taksi kosti tarif harganya relatif terjangkau					

**2. FASILITAS**

NO	PERTANYAAN	STS (1)	TS (2)	N (3)	S (4)	SS (5)
1.	Fasilitas di dalam mobil seperti (Argo, AC, pembuka jendela dll) dengan lengkap serta dapat berfungsi dengan baik.					
2.	Tempat duduk sangat bersih, rapih, dan bagus.					
3.	Taksi yang digunakan memiliki mesin yang selalu baik sehingga jarang mengalami kerusakan di jalan.					
4.	Taksi Dilengkapi Dengan Asuransi Untuk Tiap Penumpang					

**3. PELAYANAN**

NO	PERTANYAAN	STS (1)	TS (2)	N (3)	S (4)	SS (5)
1.	Sopir atau karyawan yang sedang bekerja menunjukkan penampilan yang menyenangkan					
2.	Kemampuan sopir dalam menjawab tempat yang dituju meyakinkan saya bahwa saya merasa aman					
3.	Taksi melaju dengan kecepatan yang sesuai di jalan					
4.	Taksi tepat waktu dengan tempat yang dituju					
5.	Taksi kosti menjamin keselamatan para penumpang					
6.	Sopir atau karyawan memberikan respon terhadap penumpang dengan baik					
7.	Sopir selalu konsisten melayani penumpang dengan baik dalam memberikan tarif argo					
8.	Taksi Kosti selalu mudah untuk di dapatkan di mana saja dan kapan saja.					

**4. LOYALITAS**

NO	PERTANYAAN	STS (1)	TS (2)	N (3)	S (4)	SS (5)
1.	Secara keseluruhan anda lebih sering menggunakan taksi Kosti dibandingkan taksi merk lain.					
2.	Jika ada seseorang yang meminta saran anda taksi apa yang sebaiknya digunakan, anda akan menyarankan taksi Kosti pada orang tersebut.					
3.	Anda akan mengatakan hal-hal yang baik / positif mengenai taksi Kosti kepada orang lain.					
4.	Harga bukanlah masalah yang penting bagi anda untuk menggunakan taksi Kosti.					
5.	Jika anda naik taksi, anda akan menempatkan taksi Kosti sebagai pilihan utama anda.					

5. Bagaimana kepuasan anda tentang tarif, pelayanan, dan fasilitas yang diberikan taksi kosti?

Puas

Tidak puas

**Kuisisioner**

Dari Pembagian 200 data kuisisioner kepada pelanggan taksi KOSTI di peroleh 135 kuisisioner yang

menyatakan PUAS dan 65 Kuisisioner yang menyatakan TIDAK PUAS.

Data kuisisioner diolah menggunakan Microsoft Office Excel seperti dibawah ini:

No	Nama	Jenis Kelamin (P/L)	x1-Harga Rata-rata	x2-Fasilitas	Rata-rata	x3-Pelayanan	Rata-rata	x4-Loyalitas	Rata-rata	Hasil															
1	ROSID	L	4	4	4	4	4	3	2	3	3	4	3	PUAS											
2	PANDI	P	5	5	5	5	5	3	3	4	2	3	4	5	3,375	4	3	3	4	2	3,2	PUAS			
3	YULI	P	4	4	4	4	2	5	3,75	2	4	1	3	2	3	2	4	2,625	3	3	4	2	3	3	PUAS
4	RISA	P	2	2	4	4	2	5	3,75	2	3	1	2	1	2	3	1	1,875	2	3	2	3	2	2,4	TIDAK PUAS
5	KUSWIN	P	4	4	4	4	4	5	4,25	3	2	3	3	4	4	2	2	2,875	3	4	2	3	3	3	PUAS
6	TATA	P	4	4	3	3	2	4	3	3	4	2	3	3	3	4	4	3,25	3	4	3	2	3	3	PUAS
7	NIA	P	5	5	5	4	4	5	4,5	4	4	3	2	4	3	4	3,75	3	4	3	4	4	3,6	PUAS	
8	BUDI	L	4	4	4	4	4	5	4,25	3	3	4	2	5	4	4	5	3,75	4	3	4	3	3	3,4	PUAS
9	SANTO	L	5	5	5	5	4	5	4,75	4	3	5	2	5	5	3	4	3,875	2	5	4	5	4	4	PUAS
10	SAPTO	L	5	5	5	5	4	5	4,75	4	4	5	5	5	4	4	4,25	4	3	3	2	3	3	3	PUAS
11	YAMIMA	P	2	2	4	4	2	5	3,75	4	3	4	2	5	4	5	4	2	1	2	3	3	2,2	TIDAK PUAS	
12	AGUNG	L	4	4	4	4	2	5	3,75	4	3	3	3	4	3	4	2	3,25	3	3	4	3	2	3	PUAS
13	HARSONO	L	2	2	4	4	2	5	3,75	3	2	1	2	3	2	2	1	2	3	2	3	1	2,4	TIDAK PUAS	
14	ALIM	L	2	2	5	5	5	5	5	3	2	2	1	2	2	3	2	1,125	3	2	3	1	2	2,2	TIDAK PUAS
15	SUCI	P	5	5	5	5	5	5	4,3	5	5	5	5	4	4	5	4,375	3	4	4	3	4	3,6	PUAS	
16	VERA	P	4	4	4	4	3	4	3,75	4	4	4	2	5	4	4	1	3,5	4	3	4	4	4	3,8	PUAS
17	SUKIYAH	P	4	4	4	4	3	5	4	5	5	4	2	5	4	5	5	4,375	4	3	4	4	3	3,6	PUAS
18	TIKI	P	4	4	4	4	2	5	3,75	4	4	4	2	5	3	3	5	3,75	4	3	4	3	4	3,6	PUAS
19	PUTRI	P	3	3	4	4	2	5	3,75	5	3	4	2	5	4	4	4	3,875	3	4	3	3	3	3,2	PUAS
20	FAIS	L	4	4	4	4	4	4	4	5	4	4	3	4	4	4	4	4	4	3	4	4	4	3,8	PUAS
21	SANTI	P	4	4	4	4	4	4	4	4	2	4	2	5	4	4	5	3,75	2	3	4	3	3	3	PUAS
22	DIDI	L	4	4	4	4	4	4	4	5	4	4	4	4	4	4	4	4,125	3	3	4	4	4	3,6	PUAS
23	ERIK	L	3	3	3	4	4	3	3,5	5	3	4	4	5	5	5	5	4,5	4	4	3	4	4	3,8	PUAS
24	SUWARDI	L	5	5	4	4	3	5	4	4	4	4	4	5	5	5	5	4,5	4	3	4	4	4	3,8	PUAS
25	SOFI	P	5	5	5	4	5	4,75	4	4	2	3	5	3	4	3	3,5	4	4	4	4	3	3,8	PUAS	
26	MIMI	P	1	1	2	1	2	2	1,5	1	1	1	1	1	1	1	1	1,125	1	1	1	1	1	1	TIDAK PUAS

**Data Excel Kuisisioner**

setelah data dimasukan atau diolah menggunakan Microsoft Office Excel, maka dilakukan pengolahan data kuisisioner, dengan cara mengambil data rata-rata x1 (Harga), rata-rata x2 (Fasilitas), rata-rata x3 (Pelayanan), rata-rata x4 (Loyalitas) dan Y (Hasil) dan dilakukan pemilihan atribut pada proses pengolahan data maka hasil tabel seperti tabel dibawah ini :

**Data Kuisisioner yang telah diolah**

1	NAMA	HARGA (X1)	FASILITAS (X2)	PELAYANAN (X3)	LOYALITAS (X4)	Hasil (Y)
2	ROSID	4	4	3	3	PUAS
3	PANDI	5	5	3,4	3,2	PUAS
4	YULI	4	3,8	2,6	3	PUAS
5	RISA	2	3,8	1,9	2,4	TIDAK PUAS
6	KUSWIN	4	4,3	2,9	3	PUAS
7	TATA	4	3	3,3	3	PUAS
8	NIA	5	4,5	3,4	3,6	PUAS
9	BUDI	4	4,3	3,8	3,4	PUAS
10	SANTO	5	4,8	3,9	4	PUAS
11	SAPTO	5	4,8	4,3	3	PUAS
12	YAMIMA	2	3,8	4	2,2	TIDAK PUAS
13	AGUNG	4	3,8	3,3	3	PUAS
14	HARSONO	2	3,8	2	2,4	TIDAK PUAS
15	ALIM	2	5	2,1	2,2	TIDAK PUAS
16	SUCI	5	5	4,4	3,6	PUAS
17	VERA	4	3,8	3,5	3,8	PUAS
18	SUKIYAH	4	4	4,4	3,6	PUAS

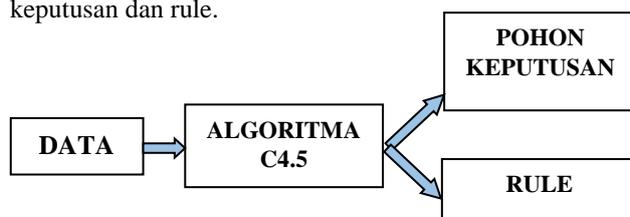
## 2. Proses Data Mining

Pada tahap selanjutnya data akan di proses menggunakan bantuan Tools Rapid Miner dan untuk pengujianya menggunakan cross validation yang ada di dalam Rapid Miner, maka data akan terlihat nilai akurasi, pohon keputusan, dan rulenya. Berikut adalah Data yang akan di import ke dalam Rapid Miner:

Nama	Harga	Fasilitas	Pelayanan	Loyalitas	Hasil
Polynomial	Numeric	numeric	Numeric	numeric	Binomial
Id	Attribute	attribute	Attributr	attribute	Label
Rosid	4	4	3	3	PUAS
Pandi	5	5	3.400	3.200	PUAS
Yuli	4	3.800	2.600	3	PUAS
Risa	2	3.800	1.900	2.400	TIDAK PUAS
....	....	....	....	....	....
Alim	2	5	2.100	2.200	TIDAK PUAS
Suci	5	5	4.400	3.600	PUAS

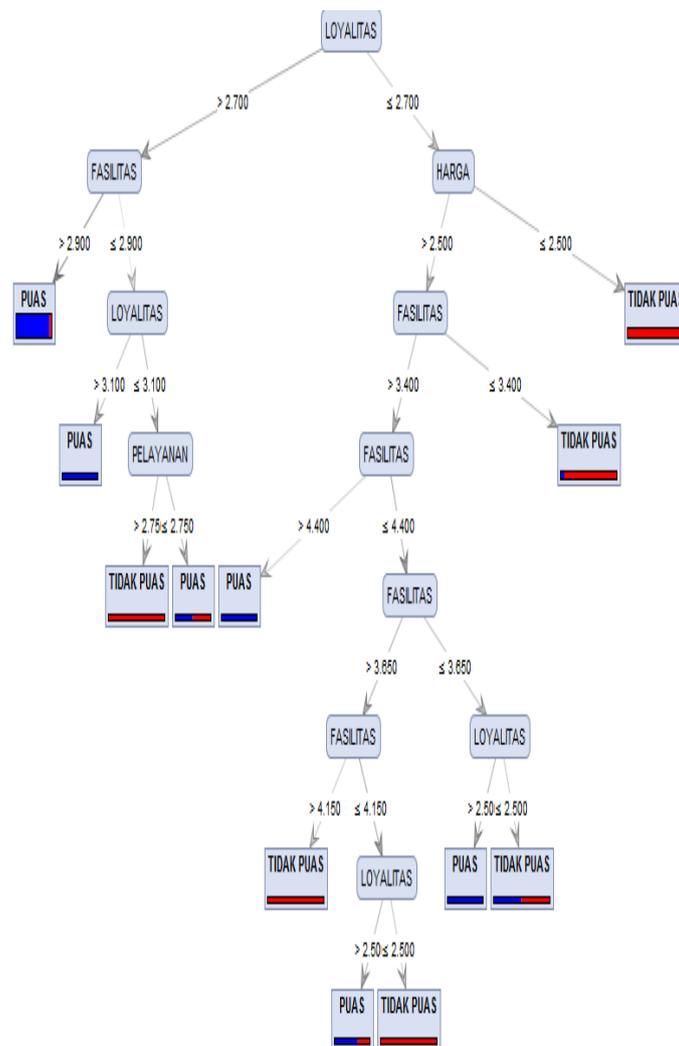
Keterangan : ■ : Nama Variabel  
■ : Jenis atau Tipe data  
■ : attribute / label

Dan berikutnya proses pengolahan data menggunakan algoritma C4.5 untuk melihat pohon keputusan dan rule.



### Alur Pengolahan Data

Setelah dilakukan perhitungan metode algoritma C4.5 pada Rapid Miner, maka terbentuk pohon keputusan seperti gambar di bawah ini :



Gambar 5. Pohon Keputusan

Pohon Keputusan diatas menghasilkan rule yang akan diimplementasikan pada program. Rule yang dihasilkan sebagai berikut :

#### Rules untuk Hasil = PUAS

- IF nilai Loyalitas > 2.700 and nilai Fasilitas > 2.900 THEN hasil = PUAS
- IF nilai Loyalitas > 2.700 and nilai Fasilitas ≤ 2.900 and nilai Loyalitas > 3.100 THEN Hasil = PUAS
- IF nilai Loyalitas > 2.700 and nilai Fasilitas ≤ 2.900 and nilai Loyalitas ≤ 3.100 and nilai Pelayanan ≤ 2.750 THEN Hasil = PUAS

4. IF nilai Loyalitas $\leq$ 2.700 and nilai Harga $>$ 2.500 and nilai Fasilitas $>$ 3.400 and fasilitas $>$ 4.400 THEN Hasil = PUAS
5. IF nilai Loyalitas $\leq$ 2.700 and nilai Harga $>$ 2.500 and nilai Fasilitas $>$ 3.400 and Fasilitas $\leq$ 4.400 and nilai Fasilitas $>$ 3.650 and nilai Fasilitas $\leq$ 4.150 and nilai Loyalitas $>$ 2.500 THEN Hasil = PUAS
6. IF nilai Loyalitas $\leq$ 2.700 and nilai Harga $>$ 2.500 and nilai Fasilitas $>$ 3.400 and Fasilitas $\leq$ 4.400 and nilai Fasilitas $\leq$ 3.650 and nilai Loyalitas $>$ 2.500 THEN Hasil = PUAS
<b>Rules untuk Hasil = TIDAK PUAS</b>
1. IF nilai Loyalitas $>$ 2.700 and nilai Fasilitas $\leq$ 2.900 and nilai Loyalitas $\leq$ 3.100 and nilai Pelayanan $>$ 2.75 THEN Hasil =TIDAK PUAS
2. IF nilai Loyalitas $\leq$ 2.700 and nilai Harga $>$ 2.500 and nilai Fasilitas $>$ 3.400 and Fasilitas $\leq$ 4.400 and nilai Fasilitas $>$ 3.650 and nilai Fasilitas $>$ 4.150 THEN Hasil = TIDAK PUAS
3. IF nilai Loyalitas $\leq$ 2.700 and nilai Harga $>$ 2.500 and nilai Fasilitas $>$ 3.400 and Fasilitas $\leq$ 4.400 and nilai Fasilitas $>$ 3.650 and nilai Fasilitas $\leq$ 4.150 and Loyalitas $\leq$ 2.500 THEN Hasil = TIDAK PUAS
4. IF nilai Loyalitas $\leq$ 2.700 and nilai Harga $>$ 2.500 and nilai Fasilitas $>$ 3.400 and Fasilitas $\leq$ 4.400 and nilai Fasilitas $\leq$ 3.650 and nilai Loyalitas $\leq$ 2.500 THEN Hasil = TIDAK PUAS
5. IF nilai Loyalitas $\leq$ 2.700 and nilai Harga $>$ 2.500 and nilai Fasilitas $\leq$ 3.400 THEN Hasil = TIDAK PUAS
6. IF nilai Loyalitas $\leq$ 2.700 and nilai Harga $\leq$ 2.500 THEN Hasil = TIDAK PUAS

Rule diatas yang nantinya akan digunakan pada implementasi program. Terdapat *Leaf Node* sejumlah 12, dengan pembagian ke kiri sejumlah 4 *Leaf Node* dan ke kanan sejumlah 8 *Leaf Node*.

Dari hasil pohon keputusan dan rule diatas terdapat beberapa data yang mengalami kesalahan atau error dalam proses klasifikasi, sebagai berikut :

**Tabel 2:** Data Error

No	X1	X2	X3	X4	Hasil	Prediksi
75	2	3.8	3.8	3.4	TIDAK PUAS	PUAS
80	3	3.8	2.8	2.6	TIDAK PUAS	PUAS
83	2	3.5	2.8	2.8	TIDAK PUAS	PUAS
98	2	2.8	2.3	2.8	TIDAK PUAS	PUAS
101	4	4	3	3.2	TIDAK PUAS	PUAS
122	4	2.8	2	2.6	PUAS	TIDAK PUAS
125	4	3.5	3.1	3	TIDAK PUAS	PUAS
126	4	4.3	3.6	4	TIDAK PUAS	PUAS
128	4	4.3	4.1	4	TIDAK PUAS	PUAS
133	2	3.8	3.3	3.2	TIDAK PUAS	PUAS
170	4	3.5	3.4	2.4	PUAS	TIDAK PUAS

### 3. Pengujian Rules Terhadap Data Kepuasan Pelanggan Taksi KOSTI dengan Variabel Data Kuisisioner

Pengujian ini dilakukan untuk mengetahui kinerja dari metode algoritma C4.5 dalam melakukan klasifikasi terhadap kelas yang telah ditentukan dalam uji coba ini. Pengujian ini dilakukan sebanyak 3 kali dengan *K-Fold Validation* yang berbeda pada operator cross validation. *K-Fold Validation* berfungsi untuk membagi jumlah data training dan data testing pada data yang di uji. Pengujian terdiri dari :

1. *K-Fold Validation* 10
2. *K-Fold Validation* 5
3. *K-Fold Validation* 3

Algoritma C4.5 melakukan training terhadap data – data yang telah dibagi oleh cross validation menjadi dua kotak yaitu training dan testing. Training terdiri dari decision tree, sedangkan testing terdiri dari apply model dan performance.

Berikut dijelaskan parameter dan operator yang digunakan pada model algoritma C4.5 :

1. Retrieving Data adalah operator yang digunakan untuk mengimport dataset yang akan digunakan, pada penelitian ini data diimport dari file excel.
2. Validation operator yang digunakan dalam penelitian ini adalah split validation, validation ini hanya membagi total dari keseluruhan dataset menjadi data training dan data testing.
3. Decision tree adalah metode klasifikasi yang digunakan dalam penelitian ini.
4. Apply model adalah operator yang digunakan dalam penelitian ini yaitu algoritma C4.5.
5. Performance adalah operator yang digunakan untuk mengukur performa akurasi dari model.

#### A. Pengujian 1

Pengujian ini dilakukan dengan membagi sebanyak 10 bagian pada 200 data kuisioner yang akan di uji. 10 bagian tersebut terdiri dari 9 bagian *data training* dan 1 bagian *data testing*. Dan dalam pengujian ini akan menghasilkan nilai akurasi, *precision*, *recall* dan *AUC(Area Under Curve)*.

Hasil akurasi *confusion matrix* pada pengujian 1 sebagai berikut :

*Confusion Matrix* Pengujian 1

	Actual	
	Puas	Tidak Puas
Pred. Puas	124	16
Pred. Tidak Puas	11	49

Jumlah benar positif (tp) sebanyak 124 data, salah positif (fp) sebanyak 11 data, jumlah benar negatif (tn) sebanyak 49 data dan jumlah salah negative (fn) sebanyak 16 data

Dari pengujian 1 pada Rapid Miner menghasilkan Akurasi sebesar 86,50%.

Berikut adalah perhitungan akurasi :

$$\begin{aligned}
 \text{Akurasi} &= \frac{tp + tn}{tp + tn + fp + fn} \times 100\% \\
 &= \frac{124 + 49}{124 + 49 + 11 + 16} \times 100\% \\
 &= (173 / 200) \times 100\% \\
 &= 0,865 \times 100\% \\
 &= 86,50\%
 \end{aligned}$$

Nilai *precision* dihitung dengan cara membagi jumlah data benar positif (*TP*) dibagi dengan jumlah data benar positif (*TP*) dan data salah negatif (*FP*)

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP + FN} \times 100\% \\
 &= \frac{124}{124 + 16} \times 100\% \\
 &= \frac{124}{140} \times 100\% \\
 &= 88,57\%
 \end{aligned}$$

Nilai *recall* dihitung dengan cara membagi jumlah data benar positif (*TP*) dibagi dengan jumlah data benar positif (*TP*) dan data salah positif (*FP*)

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FP} \times 100\% \\
 &= \frac{124}{124 + 11} \times 100\% \\
 &= \frac{124}{135} \times 100\% \\
 &= 91,85\%
 \end{aligned}$$

Hasil AUC (*Area Under Curve*) pada Pengujian 1 sebagai berikut



**Gambar 6.** AUC (*Area Under Curve*) Pengujian 1

Hasil AUC adalah 0,843 yang berarti pengujian 1 termasuk *good classification*

Evaluasi dan Validasi 1

K-Fold Validasi	Akurasi	Precision	Recall	AUC
10	86,50%	88,57%	91,85%	0.843

**B. Pengujian 2**

Pengujian ini dilakukan dengan membagi sebanyak 5 bagian pada 200 data kuisioner yang akan di uji. 5 bagian tersebut terdiri dari 4 bagian *data training* dan 1 bagian *data testing*. Dan dalam pengujian ini akan menghasilkan nilai akurasi, *precision*, *recall* dan *AUC*(*Area Under Curve*) Hasil akurasi *confusion matrix* pada pengujian 2 sebagai berikut :

*confusion matrix* Pengujian 2

	Actual	
	Puas	Tidak Puas
Pred. Puas	125	15
Pred. Tidak Puas	10	50

Jumlah benar positif (*tp*) sebanyak 125 data, salah positif (*fp*) sebanyak 10 data, jumlah benar negatif (*tn*) sebanyak 50 data dan jumlah salah negative (*fn*) sebanyak 15 data

Dari pengujian 1 pada Rapid Miner menghasilkan Akurasi sebesar 87,50%.

Berikut adalah perhitungan akurasi :

$$\begin{aligned}
 \text{Akurasi} &= \frac{tp + tn}{tp + tn + fp + fn} \times 100\% \\
 &= \frac{125 + 50}{125 + 50 + 10 + 15} \times 100\% \\
 &= (175 / 200) \times 100\% \\
 &= 0,875 \times 100\% \\
 &= 87,50\%
 \end{aligned}$$

Nilai *precision* dihitung dengan cara membagi jumlah data benar positif (*TP*) dibagi dengan jumlah data benar positif (*TP*) dan data salah negatif (*FP*)

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP + FN} \times 100\% \\
 &= \frac{125}{125 + 15} \times 100\% \\
 &= \frac{125}{140} \times 100\% \\
 &= 89,29\%
 \end{aligned}$$

Nilai *recall* dihitung dengan cara membagi jumlah data benar positif (*TP*) dibagi dengan jumlah data benar positif (*TP*) dan data salah positif (*FP*)

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FP} \times 100\% \\
 &= \frac{125}{125 + 10} \times 100\%
 \end{aligned}$$

$$= \frac{125}{135} \times 100\%$$

$$= 92,59\%$$

Hasil AUC (*Area Under Curve*) pada Pengujian 2 sebagai berikut



**Gambar 7.** AUC (*Area Under Curve*) Pengujian 2

Hasil AUC adalah 0,774 yang berarti pengujian 2 termasuk *fair classification*

Evaluasi dan Validasi 2

K-Fold Validasi	Akurasi	Precision	Recall	AUC
5	87,50%	89,29%	92,59%	0.774

**C. Pengujian 3**

Pengujian ini dilakukan dengan membagi sebanyak 3 bagian pada 200 data kuisioner yang akan di uji. 3 bagian tersebut terdiri dari 2 bagian *data training* dan 1 bagian *data testing*. Dan dalam pengujian ini akan menghasilkan nilai akurasi, *precision*, *recall* dan *AUC*(*Area Under Curve*).

Hasil akurasi *confusion matrix* pada pengujian 3 sebagai berikut :

*confusion matrix* Pengujian 3

	Actual	
	Puas	Tidak Puas
Pred. Puas	128	17
Pred. Tidak Puas	7	48

Jumlah benar positif (*tp*) sebanyak 128 data, salah positif (*fp*) sebanyak 7 data, jumlah benar negatif (*tn*) sebanyak 48 data dan jumlah salah negative (*fn*) sebanyak 17 data

Dari pengujian 1 pada Rapid Miner menghasilkan Akurasi sebesar sebesar 88,00%.

Berikut adalah perhitungan akurasi :

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

$$= \frac{128 + 48}{125 + 50 + 7 + 17} \times 100\%$$

$$= (176 / 200) \times 100\%$$

$$= 0,88 \times 100\%$$

$$= 88,00\%$$

Nilai *precision* dihitung dengan cara membagi jumlah data benar positif (*TP*) dibagi dengan jumlah data benar positif (*TP*) dan data salah negatif (*FP*)

$$\text{Precision} = \frac{TP}{TP + FN} \times 100\%$$

$$= \frac{128}{128+17} \times 100\%$$

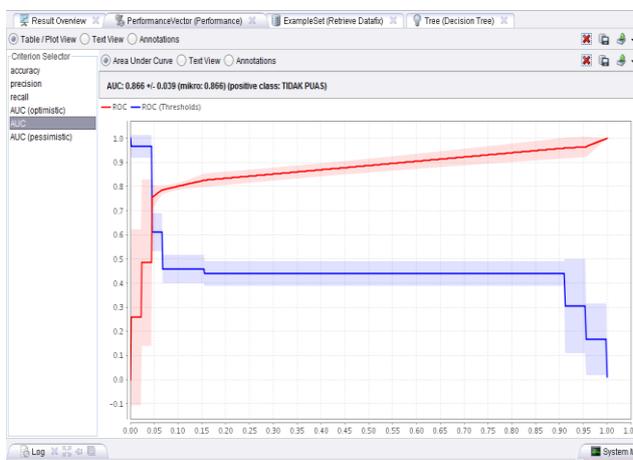
$$= \frac{128}{145} \times 100\%$$

$$= 88,28\%$$

Nilai *recall* dihitung dengan cara membagi jumlah data benar positif (*TP*) dibagi dengan jumlah data benar positif (*TP*) dan data salah positif (*FP*)

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FP} \times 100\% \\
 &= \frac{128}{128 + 7} \times 100\% \\
 &= \frac{128}{135} \times 100\% \\
 &= 94,81\%
 \end{aligned}$$

Hasil AUC (*Area Under Curve*) pada Pengujian 3 sebagai berikut



**Gambar 8.** AUC (*Area Under Curve*) Pengujian 3

Hasil AUC adalah 0,866 yang berarti pengujian 3 termasuk *good classification*

Evaluasi dan Validasi 3

K-Fold Validasi	Akurasi	Precision	Recall	AUC
3	88,01%	88,28%	94,81%	0,866

#### 4. Analisa Hasil Pengujian

Pengujian 1, 2, dan 3 menghasilkan akurasi, *precision*, *recall*, dan AUC (*Area Under Curve*) sebagai berikut :

Hasil Pengujian 1, 2, dan 3

K-Fold Validasi	Akurasi	Precision	Recall	AUC
3	88,01%	88,28%	94,81%	0,866
5	87,50%	89,29%	92,59%	0.774
10	86,50%	88,57%	91,85%	0.843

Berdasarkan table yang telah dijelaskan di atas, maka dapat diketahui bahwa pada K-Fold Validasi 10 memiliki nilai akurasi 86,50%, Precision 88,28%, dan recall 94,81%. sedangkan K-Fold Validasi 5 memiliki nilai akurasi 87,50%, Precision 89,29%, dan recall 92,59%. Dan untuk K-Fold Validasi 3 memiliki nilai akurasi tertinggi yaitu 88,00%, Precision 88,28%, dan recall 94,81%. Maka semakin kecil nilai K-Fold Validation semakin tinggi nilai akurasi yang di dapat.

#### 5. Implementasi

Berdasarkan proses pengolahan data menggunakan Algoritma C4.5 yang telah menghasilkan pohon keputusan dan rule yang telah terbentuk, selanjutnya akan diimplementasikan rule tersebut untuk membuat program prediksi data kuisisioner menggunakan tools NetBeans. Berikut tampilan programnya :



## PENUTUP

### 1. Kesimpulan

Dari hasil penelitian yang telah dilakukan oleh peneliti, maka dapat disimpulkan bahwa kepuasan pelanggan taksi KOSTI dapat diprediksi dan dievaluasi dengan memanfaatkan teknik data mining menggunakan algoritma *decision tree* C4.5 untuk memprediksi (menentukan kelas) dari kepuasan pelanggan dengan data training yang telah diperoleh.

Dari 3 kali percobaan dan pengujian prediksi kepuasan pelanggan taksi KOSTI dengan *rapidMiner* menggunakan metode *decision tree* C4.5, diperoleh akurasi sebesar 88,01% dengan kriteria akurasi *Good classification* menggunakan *confusion matrix*.

### 2. Saran

Untuk meningkatkan kinerja dan menyempurnakan penelitian yang telah dibuat, peneliti memberikan saran sebagai berikut :

1. Penelitian ini dapat dikembangkan dengan menggabungkan atau membandingkan dengan algoritma klasifikasi lain untuk mendapatkan hasil prediksi yang lebih baik.
2. Sebaiknya jumlah data kuisioner ditambah, sehingga dapat diperoleh hasil akurasi fungsi algoritma yang lebih baik.
3. Penyempurnaan aplikasi java agar lebih baik dan dapat digunakan oleh pihak yang membutuhkan.

## DAFTAR PUSTAKA

- [1] Deddy Setyawan, "Analisis Kepuasan Pengguna Jasa Transportasi Taksi Untuk Meningkatkan Loyalitas," Universitas Diponegoro, 2010.
- [2] Syukria Darman, " Pengaruh Kualitas Pelayanan Jasa Transportasi Taksi Terhadap Loyalitas Konsumen Pada PT Putra Transpor Nusantara Bandung," Bandung, Universitas Widyatama, 2012.
- [3] Resty Mardiana, "Faktor – Faktor Yang Mempengaruhi Kepuasan Pengguna Jasa Taksi Blue Bird," Jakarta, Universitas Gunadarma, 2010.
- [4] Administrator, "SHU Kosti Semarang Naik," 30 Agustus 2015. [Online]. Available : <http://kostisemarang.com/index.php/news>. [Accessed 10 September 2015].
- [5] Han, Jiawei dan Kamber, Micheline, "Data Mining Concepts and Techniques Second Edition," San Francisco: Morgan Kauffman, 2001.
- [6] Ibnu Fathur Rochman, "Penerapan Algoritma C4.5 Pada Kepuasan Pelanggan Perum DAMRI," Universitas Dian Nuswantoro, 2015.
- [7] Teguh Budi Santoso, "Analisa dan penerapan metode C4.5 Untuk Prediksi Loyalitas Pelanggan," Universitas Satya Negara Indonesia, 2014.
- [8] David Hartanto Kamagi, Seng Hansun, " Implementasi Data Mining dengan Algoritma C4.5 untuk memprediksi Tingkat Kelulusan Mahasiswa," Universitas Multimedia Nusantara, Juni 2014.
- [9] Anik Andriani, " Penerapan Algoritma C4.5 Pada Progam Klasifikasi Mahasiswa Dropout," AMIK BSI Jakarta, 2012.
- [10] Dyah Satiti, Sucipto, Shyntia Atica, " Analisis Preferensi Konsumen Waralaba Makanan Cepat Saji Dengan Menggunakan Pendekatan Data Mining Di Restoran X Surabaya," Universitas Brawijaya, 2014.
- [11] Wirasmita, Rivai , Sitorus, Roland, Manurung, Brenggan, " Kamus Lengkap Ekonomi," Pionir Jaya, Bandung, 1999.

- [12] Kotler, Philip, “ Marketing Management: Analysis. Planning. Implementation. And Control, Ninth Edition,” Prentice-Hall .Inc, New Jersey, 1997.
- [13] Santosa, B. “ *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*,” Graha Ilmu, Yogyakarta. 2007.
- [14] Undavia J N, Patel dan Dolia, “Comparison of Classification Algorithms to Predict Students’ Post Graduation Course in Weka Environment,” *International Journal of Advanced Research in Computer Science and Software Engineering*. vol. 3. no. 9. 2013.
- [15] Gorunescu F, “ Data Mining Concept Model Technique,” Craiova, Romania: Springer, 2011.
- [16] Chapman P, “ CRISP-DM 1.0: Step-by-step Data Mining Guide, SPSS,” 2000.
- [17] Kusriani dan Emha Taufiq Luthfi, “Alogaritma Data Mining.” Yogyakarta: Andi Publisher, 2009.
- [18] Basuki dan Syarif, "Modul Ajar Decision Tree," Surabaya, 2003
- [19] Burns, Alvin C. dan Bush, Ronald F, “Marketing Research”, Prentice Hall, New Jersey, 2000
- [20] Aritonang, Lerbin, “Penelitian Pemasaran”, UPT Penerbitan Universitas Tarumanagara, Jakarta, 1998.
- [21] Swasta, Basu. “Manajemen penjualan” , cetakan kedua belas, Yogyakarta: BFSE, 2005.