

**IMPLEMENTASI DATA MINING MENGGUNAKAN
ALGORITMA C4.5 UNTUK MEMPREDIKSI
KEPUASAN PELANGGAN GOJEK**

SKRIPSI



ADHITYO NUGRAHA

H13114325

**PROGRAM STUDI ILMU KOMPUTER
DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
AGUSTUS 2020**

**IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA C4.5
UNTUK MEMPREDIKSI KEPUASAN PELANGGAN GOJEK**

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
pada Program Studi Ilmu Komputer Departemen Matematika Fakultas
Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin Makassar

UNIVERSITAS HASANUDDIN

ADHITYO NUGRAHA

H13114325

**PROGRAM STUDI ILMU KOMPUTER DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

UNIVERSITAS HASANUDDIN

AGUSTUS 2020

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa
Skripsi yang saya buat dengan judul:

**Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk
Memprediksi Kepuasan Pelanggan Gojek**

UNIVERSITAS HASANUDDIN

adalah benar hasil karya saya sendiri bukan hasil plagiat dan belum pernah
dipublikasikan dalam bentuk apapun

Makassar, 14 Agustus 2020



[Handwritten signature]

Antyo Nugraha

NIM H 131 14 325

**IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA C4.5
UNTUK MEMPREDIKSI KEPUASAN PELANGGAN GOJEK**

Disetujui Oleh:

Pembimbing Utama,



Dr. Muhammad Hasbi, M.Sc

NIP. 196307201989031003

Pembimbing Pertama,



Supri Bin Hj Amir, S.Si., M.Eng

NIP. 198805042019031012

Ketua Program Studi

Ilmu Komputer



Dr. Muhammad Hasbi, M.Sc

NIP. 196307201989031003





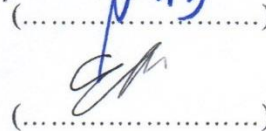

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Adhityo Nugraha
NIM : H13114325
Program Studi : Ilmu Komputer
Judul Skripsi : “Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Kepuasan Pelanggan Gojek”

Telah berhasil dipertahankan dihadapan dewan penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

1. Ketua : Dr. Drs. H. Muhammad Hasbi, M.Sc. 
2. Sekertaris : Supri bin Hj. Amir, S.Si, M.Eng 
3. Anggota : Dr. Diaraya, M.AK 
4. Anggota : Edy Saputra, S.Si., M.Si 

Ditetapkan di : Makassar
Tanggal : 14 Agustus 2020



KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas penyertaan dan kasih-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “**Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Memprediksi Kepuasan Pelanggan Gojek**”. Penulis mengungkapkan terima kasih yang tak terhingga kepada semua pihak tanpa terkecuali yang telah membantu, serta mendukung penulis hingga penelitian ini bisa terselesaikan.

Oleh karena itu, penulis menyampaikan ucapan terima kasih yang tulus dan tak terhingga kepada Ayahanda dan Ibunda tercinta, **Alm. Albert Polean** dan **Albertina**, yang telah mendidik dan membesarkan penulis dengan penuh kesabaran bertabur cinta dan kasih sayang serta dengan penuh ketulusan hati dan kesungguhan dalam memberikan dukungan moril serta doanya yang tak ternilai harganya demi keberhasilan penulis selama menjalani proses pendidikan. Serta kepada saudara **Arthur S. Polean, SE., Andreas J. Polean, S.ST.Pel., Anggimoto S. Polean, A.Md** dengan kesungguhan dan ketulusannya telah memberikan dukungan serta doa yang tak ternilai. Tidak lupa pula penulis sampaikan terima kasih kepada :

1. Ibu **Prof. Dr. Dwia Aries Tina Palubuhu, MA** selaku Rektor Universitas Hasanuddin, Bapak **Dr. Eng. Amiruddin** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Bapak **Dr. Nurdin, S.Si., M.Si** selaku Ketua Departemen Matematika, Bapak **Dr. Diaraya, M.Ak** selaku Ketua Program Studi Ilmu Komputer, dan **segenap dosen pengajar** serta **Staf Departemen Matematika**, yang telah membekali ilmu dan kemudahan-kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Matematika. Terimakasih atas kerjasamanya, baik dibidang akademik maupun dibidang kemahasiswaan.
2. Bapak **Dr. H. Muhammad Hasbi, M.Sc** selaku pembimbing utama sekaligus ketua penguji dan Bapak **Supri bin Hj. Amir, S.Si, M.Eng** selaku pembimbing pertama sekaligus sekertaris penguji penulis, terimakasih atas kesediaan, kesabaran, dan kesetiiaannya untuk membimbing dan menuntun penulis hingga menyelesaikan tugas akhir ini.

3. Bapak **Dr. Diaraya, M.Ak** selaku penguji, Bapak **Edy Saputra, S.Si., M.Si** selaku penguji, dan Bapak **Alm. Dr. Loeky Harianto, MS, M. A. T., M.Sc.** selaku penasehat akademik yang telah meluangkan waktunya untuk memberikan semangat, saran dan arahan kepada penulis dalam penyusunan skripsi ini.
4. Teman-teman **PPGT Labuang Baji**, yang selalu mensupport dan mendoakan hingga bisa sampai pada saat ini.
5. Teman-teman **Pengurus PPGT Klasis Makassar periode 2019-2021** yang selalu menjadi penyemangat dan motivasi serta tidak lupa mendoakan.
6. Teman-teman program studi **ILMU KOMPUTER 2014**, terima kasih untuk dukungan dan waktunya saat berjuang sama-sama di dunia perkuliahan.

Serta kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih untuk semuanya. Semoga apa yang telah dituliskan oleh penulis pada skripsi ini dapat bermanfaat bagi sesama.

Makassar, 11 Agustus 2020



Adhityo Nugraha

**PERYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMIK**

Sebagai civitas akademik Universitas Hasanuddin saya bertanda tangan di bawah ini :

Nama : Adhityo Nugraha
NIM : H 131 14 325
Program Studi : Ilmu Komputer
Departemen : Matematika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Prediktor Royalti Non-eksklusif (*Non-exclusive Royalty-Free Right*)** atas tugas akhir saya yang berjudul :

**“Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi
Kepuasan Pelanggan Gojek”**

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada tanggal, 14 Agustus 2020

Yang Menyatakan,



Adhityo Nugraha

ABSTRAK

Algoritma C4.5 adalah algoritma yang digunakan untuk membuat 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. Di era modern ini, semua serba cepat dan semua serba praktis dan masyarakat mempunyai aktivitas yang beragam dan cara untuk memenuhi aktivitas tersebut, masyarakat memerlukan transportasi sebagai alat penunjang/alat bantu dalam melakukan aktivitasnya. Oleh karena itu pengusaha jasa transportasi berlomba-lomba dalam menggaet konsumennya, sehingga muncul aplikasi-aplikasi penyedia jasa transportasi online yang menawarkan kemudahan pemesanan, kenyamanan armada, ketepatan waktu dan lain sebagainya. Penelitian ini bertujuan untuk mengetahui akurasi algoritma C4.5 untuk memprediksi kepuasan pelanggan pada Gojek dan Menerapkan algoritma C4.5 pada kepuasan pelanggan Gojek. Berdasarkan hasil analisis cara menerapkan algoritma C4.5 pada data kepuasan pelanggan Gojek adalah dengan menggunakan Aplikasi Rapid Minner sehingga mendapatkan pohon keputusan dari data pelanggan Gojek. Hasil dari percobaan penerapan Algoritman C4.5 pada data kepuasan pelanggan Gojek yang telah dilakukan penulis sebanyak 3 kali, maka dapat di ketahui bahwa percobaan 1, 2, dan 3 ini dapat dikatakan baik dan berhasil, karena sudah terlihat jelas bahwa nilai akurasi yang tinggi dan semakin meningkat sehingga dapat dikatakan akurat. Algoritma C4.5 pada kepuasan pelanggan Gojek dapat diterapkan dengan baik.

Kata Kunci :Kepuasan Pelanggan, Algoritma C4.5, RapidMinner

ABSTRACT

The C4.5 algorithm is an algorithm used to create a decision tree. The decision tree is a very powerful and well-known method of classification and prediction. The decision tree method converts very large facts into decision trees that represent rules. The rules can easily be understood in natural language. In this modern era, everything is fast paced and everything is all practical and people have various activities and ways to fulfill these activities, people need transportation as a means of support / means of assisting in carrying out their activities. Therefore, transportation service entrepreneurs are competing in attracting their customers, so that online transportation service provider applications have emerged that offer easy ordering, fleet convenience, punctuality and so on. This study aims to determine the accuracy of the C4.5 algorithm for predicting customer satisfaction on Gojek and applying the C4.5 algorithm to Gojek customer satisfaction. Based on the results of the analysis of how to apply the C4.5 algorithm to Gojek customer satisfaction data, it is to use the Rapid Minner application so that you get a decree tree from Gojek customer data. The results of the experiment using C4.5 algorithm on Gojek customer satisfaction data that the author has done 3 times, it can be obeyed that experiments 1, 2, and 3 can be said to be good and successful, because it is clear that the high accuracy value and increasing so that it can be said to be accurate. The C4.5 algorithm on Gojek customer satisfaction can be applied well.

Keywords: Customer Satisfaction, C4.5 Algorithm, Rapid Minner

DAFTAR ISI

KATA PENGANTAR	ii
ABSTRAK	vii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
1.5 Batasan Masalah	3
1.6 Penelitian Terkait	4
BAB II TINJAUAN PUSTAKA	7
2.1 Landasan Teori	7
2.1.1 <i>Data Mining</i>	7
2.1.2 Kualitas Jasa dan Pelayanan	8
2.1.3 Kepuasan Konsumen	8
2.1.4 Loyalitas Pelanggan	9
2.1.5 Variabel Koisuner	9
2.1.6 Klasifikasi	11
2.1.7 Decision Tree Algoritma (C.45)	13
2.1.8 <i>Rapid Miner</i>	17
BAB III METODE PENELITIAN	18
3.1 Tahapan Penelitian	18
3.2 Waktu dan Lokasi Penelitian	19
3.3 Rancangan Sistem	19
3.4 Sumber Data	19
3.5 Instrumen Penelitian	19

BAB IV HASIL DAN PEMBAHASAN	20
4.1 Arsitektur Sistem.....	20
4.2 Pengujian.....	20
4.2.1 Rancangan Pengujian	20
4.3 Implementasi Pengujian	20
4.3.1 Data Kuesioner di import ke Aplikasi Rapidminer.....	20
4.3.2 Pengujian Rules terhadap data pelanggan Gojek dengan Variable data kuesioner	24
4.4 Analisa Hasil Pengujian	42
BAB V KESIMPULAN DAN SARAN.....	43
5.1 Kesimpulan	43
5.2 Saran.....	43
DAFTAR PUSTAKA	44

DAFTAR GAMBAR

Gambar 2.1 Konsep Decision Tree	15
Gambar 2.2 Contoh Pohon Keputusan.....	15
Gambar 3.1 Tahapan Penelitian	18
Gambar 3.2 Rancangan Sistem	19
Gambar 4.1 Menghubungkan data dengan algoritma C4.5 (Decision Tree)	21
Gambar 4.2 Pohon Keputusan.....	21
Gambar 4.3 Deskripsi tentang pohon keputusan	22
Gambar 4.4 Tingkat Akurasi Prediksi.....	23
Gambar 4. 5 Pohon keputusan data training 40%	25
Gambar 4.6 Rules Keputusan data training 40%	25
Gambar 4.7 Tingkat akurasi prediksi pengujian 1	29
Gambar 4.8 Kurva AUC (Area Under Curva) Pengujian 1	30
Gambar 4.9 Pohon keputusan data training 60%	32
Gambar 4.10 Rules Keputusan data training 60%	33
Gambar 4.11 Tingkat akurasi prediksi pengujian 2	35
Gambar 4.12 Kurva AUC (Area Under Curva) Pengujian 2	36
Gambar 4.13 Pohon keputusan data training 80%	38
Gambar 4.14 Rules Keputusan data training 80%	39
Gambar 4.15 Tingkat akurasi prediksi pengujian 3	40
Gambar 4.16 Kurva AUC (Area Under Curva) Pengujian 3	41

DAFTAR TABEL

Tabel 4.1 Data yang akan di import ke dalam Rapidminer.....	21
Tabel 4.2 Data Error.....	23
Tabel 4.3 Data Training 30%	24
Tabel 4.4 Data Testing 70%	26
Tabel 4.5 Data Error Pengujian 1	26
Tabel 4.6 Evaluasi dan Validasi.....	30
Tabel 4.7 Data Training 60%	31
Tabel 4.8 Data Testing 40%	33
Tabel 4.9 Data Error Pengujian 2.....	34
Tabel 4.10 Evaluasi dan Validasi 2.....	36
Tabel 4.11 Data Training 80%	37
Tabel 4.12 Data Testing 20%	39
Tabel 4.13 Data Error Pengujian 3.....	40
Tabel 4.14 Evaluasi dan Validasi 3.....	42
Tabel 4.15 Hasil Pengujian 1,2 dan 3	42

BAB I

PENDAHULUAN

1.1 Latar Belakang

Transportasi menjadi salah satu penunjang penting dalam kegiatan sehari-hari terutama di daerah perkotaan. Transportasi yang baik dapat mencerminkan keteraturan kota yang baik pula, hal ini disebabkan karena transportasi merupakan suatu alat bantu dalam mengarahkan pembangunan di daerah perkotaan. Selain itu, transportasi juga merupakan prasarana bagi pergerakan manusia atau barang yang timbul akibat adanya kegiatan di daerah perkotaan. Transportasi merupakan sarana perkembangan yang penting dan strategis dalam memperlancar roda perekonomian, memperkuat persatuan dan kesatuan serta mempengaruhi semua aspek kehidupan. Transportasi dapat membantu perekonomian yang baik disuatu daerah bahkan disuatu Negara.

Di era modern ini, semua serba cepat dan semua serba praktis dan masyarakat mempunyai aktivitas yang beragam dan cara untuk memenuhi aktivitas tersebut, masyarakat memerlukan transportasi sebagai alat penunjang/alat bantu dalam melakukan aktivitasnya. Oleh karena itu pengusaha jasa transportasi berlomba-lomba dalam menggaet konsumennya, sehingga muncul aplikasi-aplikasi penyedia jasa transportasi *online* yang menawarkan kemudahan pemesanan, kenyamanan armada, ketepatan waktu dan lain sebagainya. Sebagai salah satu sarana transportasi yang banyak digunakan masyarakat Indonesia, yaitu jasa Gojek *Online* Indonesia yang menawarkan berbagai alternatif penyediaan jasa yang telah tersedia diberbagai kota besar pada khususnya, guna membantu masyarakat dalam melakukan kegiatan kesehariannya, seperti aktivitas pekerjaan, kuliah, *travelling*, penelitian, perdagangan maupun perjanjian lainnya.

Banyaknya perguruan tinggi di Kota Makassar baik PTN maupun PTS dan membludaknya mahasiswa yang menempuh studi di perguruan tinggi, usaha jasa transportasi ini di Kota Makassar sangat dibutuhkan karena notabennya mahasiswa yang berdiam di Kota Makassar berasal dari luar wilayah Kota Makassar, dan tidak

menutup kemungkinan mahasiswa yang asalnya dari area Makassar juga membutuhkan jasa transportasi ini karena jasa transportasi ini memberi kemudahan kepada konsumennya yang memakai jasa tersebut, bahkan dari mahasiswa tersebut ada yang menjadi pengemudi kendaraan bermotor Gojek *Online* Indonesia.

Alasan menggunakan jasa layanan jemput-antar kendaraan bermotor Gojek karena harga yang terbilang murah berkisar Rp 10.000 sampai Rp 15.000 yang dapat menjangkau kantong kalangan mahasiswa pada umumnya. Selain itu, keberadaan pengemudi Gojek ini berada dimana mana, pengemudi yang ramah kepada sesama pengendaranya selain itu ramah pula kepada konsumennya, pelayanan yang begitu cepat dan kepandaian dalam mengendarai kendaraannya tidak diragukan, selain itu banyak menyediakan jasa pelayanan yang dapat ditemui dengan sekali sentuh saja pada media *smartphone* yang dimiliki, dimana sekarang ini penggunaan *smartphone* disemua kalangan sudah banyak dan tak dapat dipungkiri dari pengguna tersebut terkadang memiliki lebih dari satu *smartphone* yang dipakainya. Namun tak semua warga yang berdiam dalam wilayah Indonesia dapat menggunakan jasa tersebut, hal ini dikarenakan fasilitas dalam aplikasi *smartphone* yang dimiliki kurang mendukung dengan sistem operasi yang ada. Sebagai gantinya bila ingin menikmati jasa pelayanan Gojek *Online* terkadang harus melakukan pemesanan melalui *smartphone* rekanannya.

Dengan adanya fasilitas tersebut diharapkan penumpang akan merasa nyaman dalam perjalanan sehingga dapat memperoleh rasa puas kepada penggunanya atas berbagai fasilitas yang diperolehnya. Dan diharapkan dengan keadaan tersebut membuat loyalitas penumpang akan terbentuk karena telah merasakan kepuasan sehingga akan melakukan pemakaian ulang pada waktu yang akan datang.

Algoritma C4.5 adalah algoritma yang digunakan untuk membuat 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.

Untuk mengetahui apakah perusahaan telah memberikan kualitas jasa yang sesuai dengan harapan pelanggan, maka perlu dilakukan evaluasi dari sisi

pelanggannya untuk mengetahui apakah pelanggan puas atau tidak puas dengan pelayanan yang di berikan. Oleh karena itu melihat permasalahan di atas maka dilakukan dengan penelitian dengan mengambil judul: “Implementasi *Data Mining* Menggunakan Algoritma C4.5 Untuk Prediksi Kepuasan Pelanggan Gojek”

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas maka rumusan masalah pada penelitian ini adalah :

1. Bagaimana cara menerapkan algoritma C4.5 pada kepuasan pelanggan Gojek?
2. Menentukan akurasi algoritma C4.5 untuk memprediksi kepuasan pelanggan pada Gojek?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah diatas maka tujuan pada penelitian ini adalah sebgai berikut :

1. Mengetahui akurasi algoritma C4.5 untuk memprediksi kepuasan pelanggan pada Gojek.
2. Menerapkan algoritma C4.5 pada kepuasan pelanggan Gojek.

1.4 Manfaat Penelitian

Manfaat yang diharapkan dari hasil penelitian ini adalah :

1. Penelitian ini dapat memberikan informasi bagi peneliti atau calon peneliti lain untuk menerapkannya kedalam sistem yang lebih luas dan lebih kompleks atau sebagai bahan acuan yang dapat dikembangkan bagi kemungkinan pengembang konsep dan materi lebih lanjut serta dapat melengkapi referensi pustaka akademik.
2. Untuk mendapatkan pendukung keputusan bagi Gojek untuk memperbaiki pelayanannya dalam mempertahankan pelanggan.

1.5 Batasan Masalah

Agar pembahasan pada penelitian ini lebih terarah, maka masalah yang akan dibahas dibatasi pada :

1. Dataset yang penulis analisa ini difokuskan pada data kuisioner yang dibagikan kepada 1000 responden pelanggan Gojek di Makassar pada bulan Oktober 2019 – Januari 2020.
2. Dataset yang penulis analisa ini difokuskan untuk prediksi kepuasan pelanggan dengan parameter yang digunakan meliputi harga, fasilitas, pelayan dan loyalitas.
3. Pengembangan analisa ini akan dititik beratkan pada implementasi metode klasifikasi *data mining* dengan *algoritma decision tree C4.5*.

1.6 Penelitian Terkait

1. Penerapan Algoritma C4.5 pada Kepuasan Pelanggan Perum DAMRI (Ibnu Fatchur Rohman, Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universtas Dian Nuswantoro Semarang).

Pada penelitian ini dapat disimpulkan sebagai berikut :

Dari percobaan yang telah dilakukan penulis sebanyak 3 kali menggunakan algoritma C.45 dapat diketahui bahwa percobaan 1, 2, dan 3 ini dapat dikatakan baik dan berhasil, karena sudah terlihat jelas bahwa nilai akurasi yang terus bertambah dan semakin akurat.

2. Pengaruh Kualitas Pelayanan Transportasi *Online* Gojek terhadap Kepuasan Pelanggan pada Mahasiswa/I Administrasi Niaga Politeknik Negeri Jakarta.

(Rifaldi, Kadunci dan Sulistyowati, Administrasi Bisnis Terapan, Politeknik Negeri Jakarta)

Pada penelitian ini dapat dideskripsikan menurut jenis kelamin, program studi, semester, dan umur. Penelitian ini menggunakan dua variabel, yaitu variabel kualitas pelayanan sebagai variabel *independent* (bebas) dengan menggunakan lima indikator penelitian, yaitu Bukti tampilan (*Tangibles*), Keandalan (*Reability*), Daya tanggap (*Responsiveness*), Jaminan (*Assurance*), dan Empati (*Emphaty*). Dan variabel kepuasan pelanggan sebagai variabel *dependent* (terikat) dengan menggunakan lima indikator penelitian, yaitu kualitas produk, kualitas pelayanan, emosional, harga, dan biaya.

3. Analisa dan Penerapan Metode C4.5 untuk Prediksi Loyalitas Pelanggan (Teguh Budi Santoso, Jurusan Teknik Informatika, Fakultas Teknik, Universitas Satya Negara Indonesia).

Pembentukan model prediksi menggunakan metode C4.5. pada algoritma C4.5 dilakukan perhitungan *entropy* dan *information gain* dimana atribut loyalitas pelanggan sebagai atribut tujuan, sedangkan harga, pelayanan, promosi, citra perusahaan, dan kepercayaan sebagai atribut sumber untuk memperoleh node akar dan node lainnya. Berdasarkan hasil klasifikasi menggunakan algoritma C4.5 menunjukkan bahwa diperoleh akurasi mencapai 97.5%, yang menunjukkan bahwa algoritma C4.5 cocok digunakan untuk mengukur tingkat loyalitas pelanggan data seluler.

4. Analisis Kepuasan Konsumen Gojek di Wilayah Kota Bandung (Astri Fuji RS, Desta Fransiska, Meitry Ayu P, Niky Juliani, Politeknik Negeri Bandung)

Hasil yang diperoleh dari penelitian ini menerangkan bahwa (1) kualitas layanan Gojek sangat berpengaruh terhadap kepuasan konsumen, (2) untuk mencapai kualitas pelayanan yang dapat lebih memuaskan konsumen, faktor *responsiveness* dan *assurance* yang dimiliki oleh *driver* Gojek harus dirubah ke arah yang lebih baik lagi, (3) faktor harga sangat memiliki pengaruh yang signifikan terhadap pemenuhan kepuasan konsumen.

5. Implementasi *Data Mining* Menggunakan Algoritma C4.5 untuk Prediksi Kepuasan Pelanggan Taksi Kosti

(Moch Rizky Ilham S, Teknik Informatika, Universitas Dian Nuswantoro).

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*.

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Data Mining

Data mining (Santosa, 2007) adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Dalam *data mining* terdapat dua pendekatan metode pelatihan, yaitu (Undavia, 2013):

- a. *Unsupervised learning*, metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*). Guru di sini adalah label dari data.
- b. *Supervised learning*, yaitu metode belajar dengan adanya latihan dan pelatih. Dalam pendekatan ini, untuk menemukan fungsi keputusan, fungsi pemisah atau fungsi regresi, digunakan beberapa contoh data yang mempunyai *output* atau label selama proses *training*.

Ada beberapa teknik yang dimiliki *data mining* berdasarkan tugas yang bisa dilakukan, setiap teknik memiliki algoritma masing-masing. Teknik dalam *data mining* terbagi menjadi enam kategori, yaitu (Chapman, 2000) :

- c. Deskripsi
Para peneliti biasanya mencoba menemukan cara untuk mendeskripsikan pola dan *trend* yang tersembunyi dalam data.
- d. Estimasi
Estimasi mirip dengan klasifikasi, kecuali variabel tujuan yang lebih kearah numerik dari pada kategori.
- e. Prediksi
Prediksi memiliki kemiripan dengan estimasi dan klasifikasi. Hanya saja, prediksi hasilnya menunjukkan sesuatu yang belum terjadi (mungkin terjadi dimasa depan).

f. **Klasifikasi**

Dalam klasifikasi variabel, tujuan bersifat kategorik. Misalnya, kita akan mengklasifikasikan pendapatan dalam tiga kelas, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

g. **Klastering**

Clustering lebih ke arah pengelompokan *record*, pengamatan, atau kasus dalam kelas yang memiliki kemiripan.

h. **Asosiasi**

Mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada satu waktu.

2.1.2 Kualitas Jasa dan Pelayanan

Pelayanan merupakan salah satu unsur yang sangat penting dalam menciptakan kepuasan konsumen. Agar harapan konsumen terpenuhi, perusahaan harus memberikan pelayanan yang berkualitas. Kualitas dapat diartikan sebagai pengukuran seberapa baik tingkat pelayanan yang diberikan dan sesuai dengan harapan konsumen, jadi dengan kata lain memberikan pelayanan berkualitas berarti menyesuaikan diri dengan harapan konsumen. Ini merupakan salah satu faktor keberhasilan dalam persaingan yang makin ketat. Pelayanan yang berkualitas adalah orientasi semua sumber daya manusia dalam suatu perusahaan terhadap kepuasan pelanggan (Wirasmita, 1999).

Definisi kualitas jasa ada beberapa macam antara lain:

Menurut Wirasmita, Sitorus dan Manurung, definisi kualitas jasa adalah:

“Suatu sifat atau ciri yang membedakan nilai dari suatu barang atau jasa dengan nilai dari barang atau jasa yang lain yang sejenis”.

2.1.3 Kepuasan Konsumen

Kepuasan konsumen merupakan hal yang sangat penting dalam industri jasa. Karena dalam industri jasa, pelayanan yang dapat memuaskan konsumen akan memberikan imbalan yang menguntungkan, serta meningkatkan daya saing perusahaan.

Philip Kotler mendefinisikan kepuasan pelanggan adalah:

“Satisfaction is a person’s feelings of pleasure or disappointment resulting from comparing a product’s perceived performance (or outcome) in relation to his or her expectations. “ (Kolter, 1997).

Secara umum kepuasan konsumen dan ketidakpuasan konsumen merupakan hasil dari perbedaan antara harapan dengan kinerja yang dirasakan oleh konsumen, Atau dengan kata lain ada dua kemungkinan yang akan terjadi, yaitu:

1. Kinerja yang dirasakan konsumen lebih besar dari yang diharapkan, artinya konsumen merasa puas dengan kualitas pelayanan yang diberikan oleh perusahaan.
2. Kinerja yang dirasakan konsumen lebih kecil dari yang diharapkan, artinya konsumen tidak puas dengan kualitas pelayanan yang diberikan perusahaan.

2.1.4 Loyalitas Pelanggan

Loyalitas pelanggan secara umum dapat diartikan kesetiaan seseorang atas suatu produk, baik barang maupun jasa tertentu. Istilah loyalitas pelanggan menurut Swastha (Swastha, 2005) sebetulnya berasal dari loyalitas merek yang mencerminkan loyalitas pelanggan pada merek tertentu. Pelanggan yang setia pada merek tertentu cenderung terikat pada merek tersebut dan akan membeli produk yang sama lagi sekalipun tersedia alternatif lainnya.

2.1.5 Variabel Koisuner

1. Harga

Menurut Basu Swastha definisi dari harga adalah “Sejumlah Uang yang dibutuhkan untuk mendapat sejumlah kombinasi dari barang beserta pelayanan” (Swastha, 2005).

2. Kualitas Pelayanan

Menurut K. Daughlas Hoffman & Bateson (2006) *“service quality is an formed by long term, overall evaluation of a firm performace”* sedangkan menurut zeithaml (2006) kualitas layanan adalah ketidak sesuaian antara harapan atau keinginan konsumen dengan persepsi konsumen dan dimensi kualitas jasa menurut Zeithaml terdapat lima dimensi dalam kualitas pelayanan yaitu *tangibles, reliability, responsiveness, assurance, dan empathy* yaitu:

a. Berwujud (*tangible*)

Yaitu kemampuan suatu perusahaan dalam menunjukkan eksistensinya kepada pihak eksternal. Penampilan dan kemampuan sarana dan prasarana fisik perusahaan yang dapat diandalkan keadaan lingkungan sekitarnya merupakan bukti nyata dari layanan yang diberikan oleh para pemberi jasa. Hal ini meliputi fasilitas fisik (contoh: gedung, gudang dan lain-lain), perlengkapan dan peralatan yang digunakan (teknologi) serta penampilan pegawainya.

b. Keandalan (*reliability*)

Yaitu kemampuan perusahaan untuk memberikan layanan sesuai dengan dijanjikan secara akurat dan terpercaya. Kinerja harus sesuai dengan harapan pelanggan yang berarti ketepatan waktu, layanan yang sama untuk semua pelanggan tanpa kesalahan, sikap yang simpatik dan dengan akurasi yang tinggi.

c. Ketanggapan (*responsiveness*)

Yaitu suatu kebijakan untuk membantu dan memberikan layanan yang cepat (*responsive*) dan tepat kepada pelanggan dengan penyampaian informasi yang jelas. Membiarkan konsumen menunggu, persepsi yang negatif dalam kualitas layanan.

d. Jaminan dan kepastian (*assurance*)

Yaitu pengetahuan, kesopansantunan dan kemampuan para pegawai perusahaan untuk menumbuhkan rasa percaya para pelanggan kepada perusahaan. Hal ini meliputi beberapa komponen antara lain komunikasi (*communication*), kredibilitas (*credibility*), keamanan (*security*), kompetensi (*competence*) dan sopan santun (*courtesy*).

e. Empati (*empathy*)

Yaitu memberikan perhatian yang tulus dan bersifat individual atau pribadi yang diberikan kepada para pelanggan dengan berupaya memahami keinginan konsumen. Dimana suatu perusahaan diharapkan memiliki pengertian dan pengetahuan tentang pelanggan, memahami kebutuhan pelanggan secara spesifik, serta memiliki waktu pengoperasian yang nyaman bagi pelanggan.

3. Fasilitas

Menurut Kotler (Kotler, 1997), mendefinisikan fasilitas yaitu segala sesuatu yang bersifat peralatan fisik dan disediakan oleh pihak penjual jasa untuk mendukung kenyamanan konsumen.

4. Loyalitas Konsumen

a. *Behaviour*

Keinginan konsumen untuk menggunakan Gojek di masa yang akan datang (*Repurchase Behaviour*)

Kecenderungan niat konsumen untuk selalu menggunakan Gojek disaat ingin menggunakan jasa transportasi *online* (*Repeat Purchase Intentions*)

b. *Attitude*

Niat konsumen untuk merekomendasikan Gojek kepada orang lain (*word of mouth*) Niat konsumen untuk mengatakan hal-hal positif tentang Gojek kepada orang lain Niat konsumen untuk mendorong orang lain agar menggunakan Gojek

c. *Cognitive*

Kerelaan konsumen untuk tetap menggunakan Gojek walaupun harga untuk menggunakan Gojek lebih mahal (*Willingness to pay more*) komitmen dari konsumen bahwa harga bukanlah masalah yang penting, dan akan tetap lebih memilih Gojek (*Preference*) Kecenderungan niat konsumen untuk selalu menggunakan Gojek dan tidak mau menggunakan Transportasi *Online* yang lain (*choice reduction behavior*) Kecenderungan untuk menempatkan Gojek sebagai pilihan utama (*first choice in mind*)

2.1.6 Klasifikasi

Data Mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu: Deskripsi, Estimasi, Prediksi, Klasifikasi, Pengklusteran, dan Asosiasi.

Klasifikasi merupakan bagian dari algoritma *data mining*, klasifikasi ini adalah algoritma yang menggunakan data dengan target (*class/label*) yang berupa nilai kategorikal/nominal. Menurut Gorunescus (Gorunescus, 2011) proses klasifikasi didasarkan pada empat komponen mendasar, yaitu:

1. Kelas (*Class*)

Variabel dependen dari model, merupakan variabel kategorikal yang merepresentasikan “label” pada objek setelah klasifikasinya. Contoh kelas semacam ini adalah: adanya kelas penyakit jantung, loyalitas pelanggan, kelas bintang (galaksi), kelas gempa bumi (badai), dll.

2. Prediktor (*Predictor*)

Variabel independen dari model, direpresentasikan oleh karakteristik (atribut) dari data yang akan diklasifikasikan dan berdasarkan klasifikasi yang telah dibuat. Contoh prediktor tersebut adalah : merokok, konsumsi alkohol, tekanan darah, frekuensi pembelian, status perkawinan, karakteristik (satelit) gambar, catatan geologi yang spesifik, kecepatan dan arah angin, musim , lokasi terjadinya fenomena , dll.

3. Pelatihan dataset (*Training dataset*)

Kumpulan data yang berisi nilai-nilai dari kedua komponen sebelumnya dan digunakan untuk melatih model dalam mengenali kelas yang cocok/sesuai, berdasarkan prediktor yang tersedia. Contoh set tersebut adalah: kelompok pasien yang diuji pada serangan jantung, kelompok pelanggan supermarket (diselidiki oleh intern dengan jajak pendapat), database yang berisi gambar untuk monitoring teleskopik dan pelacakan objek astronomi, database badai, database penelitian gempa.

4. Dataset Pengujian (*Testing Dataset*)

Berisi data baru yang akan diklasifikasikan oleh (*classifier*) model yang telah dibangun di atas sehingga akurasi klasifikasi (*model performance*) dapat dievaluasi.

Berikut beberapa model (metode) klasifikasi yang paling populer (Gorunescu, 2011):

1. *Decision/classification trees;*
2. *Bayesian classifiers/Naive Bayes classifiers;*
3. *Neural networks;*
4. *Statistical analysis;*
5. *Genetic algorithms;*

6. *Rough sets*;
7. *k-nearest neighbor classifier*;
8. *Rule-based methods*;
9. *Memory based reasoning*;
10. *Support vector machines*.

2.1.7 Decision Tree Algoritma (C.45)

Algoritma C4.5 merupakan kelompok algoritma *Decision Tree*. Algoritma ini mempunyai input berupa *training samples* dan *samples*. *Training samples* berupa data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya. Sedangkan *samples* merupakan *field-field* data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data (Sunjana, 2010).

Algoritma C4.5 adalah salah satu metode untuk membuat *decision tree* berdasarkan training data yang telah disediakan. Algoritma C 4.5 dibuat oleh **Ross Quinlan** yang merupakan pengembangan dari ID3 yang juga dibuat oleh Quinlan (Quinlan, 1993). Beberapa pengembangan yang dilakukan pada C4.5 antara lain adalah: bisa mengatasi *missing value*, bisa mengatasi *continue data*, dan *pruning*.

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. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel *input* dengan sebuah variabel target.

Karena pohon keputusan memadukan antara eksplorasi data dan pemodelan, pohon keputusan sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain. Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan *record* yang lebih kecil dengan

menerapkan serangkaian aturan keputusan. Dengan masing-masing rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain (Berry dan Linoff, 2004).

Sebuah model pohon keputusan terdiri dari sekumpulan aturan untuk membagi sejumlah populasi yang heterogen menjadi lebih kecil, lebih homogen dengan memperhatikan pada variabel tujuannya. Sebuah pohon keputusan mungkin dibangun dengan seksama secara manual atau dapat tumbuh secara otomatis dengan menerapkan salah satu atau beberapa algoritma pohon keputusan untuk memodelkan himpunan data yang belum terklasifikasi.

Variabel tujuan biasanya dikelompokkan dengan pasti dan model pohon keputusan lebih mengarah pada perhitungan *probability* dari tiap-tiap *record* terhadap kategori-kategori tersebut atau untuk mengklasifikasi *record* dengan mengelompokkannya dalam satu kelas. Pohon keputusan juga dapat digunakan untuk mengestimasi nilai dari variabel *continue* meskipun ada beberapa teknik yang lebih sesuai untuk kasus ini.

Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C4.5 (Larose, 2006).

Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan *record*. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon. Misalkan untuk menentukan main tenis, kriteria yang diperhatikan adalah cuaca, angin, dan temperatur.

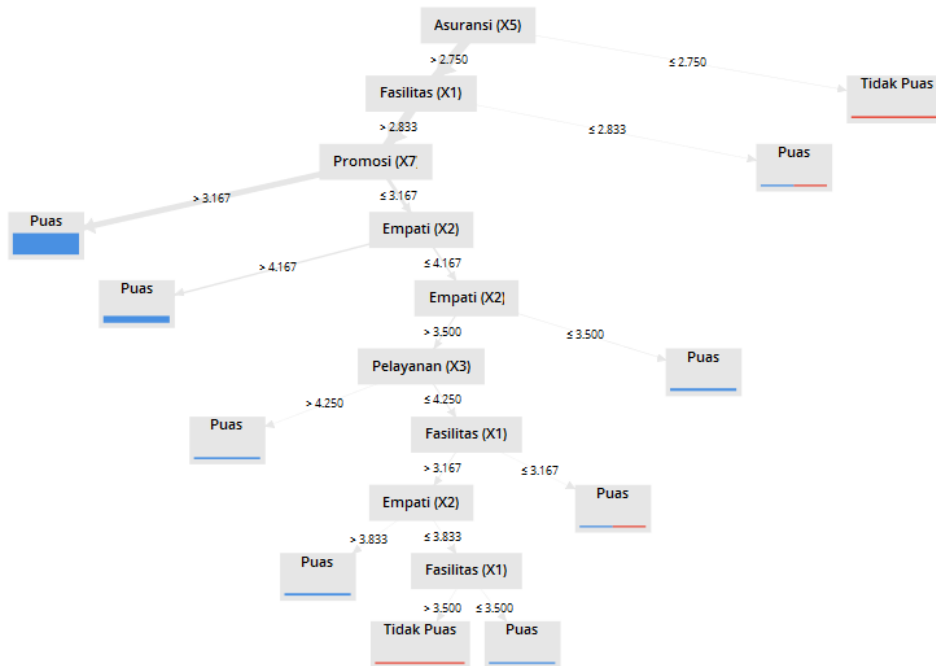
Salah satu atribut merupakan atribut yang menyatakan data solusi per item data yang disebut target atribut. Atribut memiliki nilai-nilai yang dinamakan dengan *instance*. Misalkan atribut cuaca mempunyai instance berupa cerah, berawan, dan hujan (Basuki dan Syarif, 2003)

Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan rule (Basuki dan Syarif, 2003).



Gambar 2.1 Konsep Decision Tree

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.



Gambar 2.2 Contoh Pohon Keputusan

Algoritma C4.5 merupakan salah satu algoritma *machine learning*. Dengan algoritma ini, mesin (komputer) akan diberikan sekelompok data untuk dipelajari yang disebut learning dataset. Kemudian hasil dari pembelajaran selanjutnya akan digunakan untuk mengolah data-data yang baru yang disebut test dataset. Karena algoritma C4.5 digunakan untuk melakukan klasifikasi, jadi hasil dari pengolahan

test dataset berupa pengelompokan data ke dalam kelas-kelasnya Umumnya, langkah-langkah algoritma C4.5 yang digunakan untuk membentuk pohon keputusan adalah.

- a. Pilih atribut sebagai *root*.
- b. Buat cabang untuk setiap nilai.
- c. Bagi tiap cabang kedalam kelas.
- d. Ulangi proses untuk setiap cabang sampai semua kasus pada tiap cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai *root*, didasarkan pada nilai *gain* tertinggi dari atribut yang tersedia. Sementara itu, untuk mendapat nilai *gain* tertinggi kita harus menghitung nilai *entropy* dari semua nilai didalam atribut. *Entropy* berperan sebagai parameter untuk mengukur varian dari data sampel. Setelah nilai *entropy* dalam data sampel diketahui, atribut yang paling berpengaruh akan menjadi pengukur dalam pengklasifikasian data, ukuran ini disebut sebagai *Information gain*.

Rumus menghitung entropy pada algoritma C4.5

$$Entropi (S) = \sum_{i=1}^k -p_i * \log_2 p_i$$

Keterangan :

- S adalah Himpunan (dataset) kasus
- k adalah banyaknya partisi S
- Pi adalah probabilitaas yang didapat dari Sum (Ya) atau Sum (Tidak) dibagi total kasus

Setelah mendapatkan entropi dari keseluruhan kasus, lakukan analisis pada setiap atribut dan nilai-nilainya dan hitung entropinya. Langkah berikutnya yaitu dengan menghitung *Gain*, rumus daripada *Gain* adalah sebagai berikut:

$$Gain (A) = Entropi (S) - \sum_{i=1}^k \frac{S_i}{S} * Entropi (S_i)$$

2.1.8 Rapid Miner

Rapid Miner merupakan perangkat lunak yang bersifat terbuka (*open source*). *Rapid Miner* adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi. *Rapid Miner* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. *Rapid Miner* memiliki kurang lebih 500 operator *data mining*, termasuk operator untuk *input*, *output*, data preprocessing dan visualisasi. *Rapid Miner* merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin *data mining* yang dapat diintegrasikan pada produknya sendiri. *Rapid Miner* ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi.

Rapid Miner sebelumnya bernama YALE (*Yet Another Learning Environment*), dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh RalfKlinkenberg, Ingo Mierswa, dan Simon Fischer di *Artificial Intelligence Unit* dari *University of Dortmund*. *Rapid Miner* didistribusikan di bawah lisensi AGPL (GNU Affero General Public License) versi 3. Hingga saat ini telah ribuan aplikasi yang dikembangkan menggunakan *Rapid Miner* di lebih dari 40 negara. *Rapid Miner* sebagai *software open source* untuk *data mining* tidak perlu diragukan lagi karena *software* ini sudah terkemuka di dunia. *Rapid Miner* menempati peringkat pertama sebagai *software data mining* pada *polling* oleh KDnuggets, sebuah portal data-mining pada 2010-2011