

**PENERAPAN ALGORITMA KLASIFIKASI C4.5 BERBASIS
ADABOOST UNTUK PREDIKSI LOYALITAS PELANGGAN
DISTRIBUTOR PULSA ELEKTRONIK**



TESIS

Diajukan untuk memenuhi salah satu syarat kelulusan Program Strata 2 (S2)

SUPRIATININGSIH

14002068

**PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER**

NUSA MANDIRI

JAKARTA

2018

SURAT PERNYATAAN ORISINILITAS

Yang bertanda tangan di bawah ini :

Nama : Supriatiningsih
NIM : 14002068
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Software Engineering*

Dengan ini menyatakan bahwa tesis yang telah saya buat dengan judul: "Penerapan Algoritma Klasifikasi C4.5 Berbasis Adaboost Untuk Prediksi Loyalitas Pelanggan Distributor Pulsa Elektronik" adalah hasil karya sendiri, dan semua sumber baik yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar dan tesis belum pernah diterbitkan atau dipublikasikan dimanapun dan dalam bentuk apapun.

Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila dikemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa tesis yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri dicabut/dibatalkan.

Jakarta, 28 Agustus 2018

Yang menyatakan,



Supriatiningsih

PERSETUJUAN DAN PENGESAHAN TESIS

Tesis ini diajukan oleh:

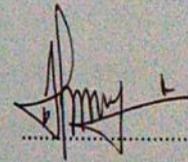
Nama : Supriatiningsih
 NIM : 14002068
 Program Studi : Ilmu Komputer
 Jenjang : Program STRATA-2
 Judul Tesis : Penerapan Algoritma Klasifikasi C4.5 Berbasis Adaboost Untuk Prediksi Loyalitas Pelanggan Distributor Pulsa Elektronik

Telah dipertahankan pada periode 2018-1 dihadapan penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh MAGISTER ILMU KOMPUTER (M.Kom) pada Program STRATA-2 Program Studi Ilmu Komputer di Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri Pascasa.

Jakarta, 28 Agustus 2018

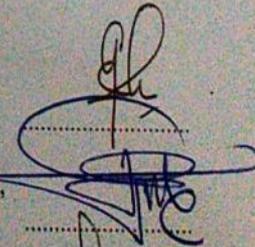
PEMBIMBING TESIS

Dosen Pembimbing : Dr. Didi Rosiyadi, M.Kom

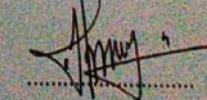


DEWAN PENGUJI

Penguji I : Dr. Sularso Budilaksono, M.Kom



Penguji II : Dr. Mochamad Wahyudi, MM, M.Kom,
: M.Pd



Penguji III / Dosen Pembimbing : Dr. Didi Rosiyadi, M.Kom

KATA PENGANTAR

Puji syukur alhamdulillah, penulis panjatkan kehadiran Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya, sehingga pada akhirnya penulis dapat menyelesaikan tesis ini tepat pada waktunya. Dimana tesis ini penulis sajikan dalam bentuk buku yang sederhana. Adapun judul tesis, yang penulis ambil sebagai berikut “Penerapan Algoritma Klasifikasi C4.5 Berbasis Adaboost Untuk Prediksi Loyalitas Pelanggan Distributor Pulsa Elektronik”.

Tujuan penulisan tesis ini dibuat sebagai salah satu syarat untuk mendapatkan gelar Magister Ilmu Komputer (M.Kom) pada Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (PPs MIK STMIK Nusa Mandiri).

Tesis ini diambil berdasarkan data transaksi yang ada pada distributor pulsa elektronik PT.Sembilan Reload Purwokerto. Penulis juga mencari dan menganalisa berbagai macam sumber referensi, baik dalam bentuk jurnal ilmiah, buku-buku literatur, internet, dll yang terkait dengan pembahasan pada tesis ini.

Penulis menyadari bahwa tanpa bimbingan dan dukungan dari semua pihak dalam pembuatan tesis ini, maka penulis tidak dapat menyelesaikan tesis ini tepat pada waktunya. Untuk itu ijinkanlah penulis pada kesempatan ini untuk mengucapkan terima kasih yang sebesar-besarnya kepada:

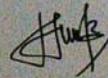
1. Bapak Prof. Dr. Ir. R. Eko Indrajit, M.Sc, MBA selaku Direktur Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri.
2. Bapak Dr. Mochamad Wahyudi, M.M, M.Kom, M.Pd selaku Ketua Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri.
3. Ibu Dr. Didi Rosiyadi M.Kom yang sudah dengan setulus hati membimbing, mengarahkan menyumbangkan ide, waktu, dan tenaganya dalam membimbing penulis untuk menyelesaikan tesis ini.
4. Bapak Ir. Naba Aji Notoseputro selaku Direktur AMIK BSI yang telah mengijinkan penulis untuk menyelesaikan pendidikan S2.

5. Orang tua tercinta serta seluruh keluarga yang telah memberikan dukungan material dan moral kepada penulis.
6. Seluruh staf pengajar (dosen) Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri yang telah memberikan pelajaran yang berarti bagi penulis selama menempuh studi.
7. Seluruh staf dan karyawan Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri yang telah melayani penulis dengan baik selama kuliah.
8. Teman-teman satu kelas, sahabat yang telah berjuang bersama-sama.

Serta semua pihak yang terlalu banyak untuk penulis sebutkan satu persatu sehingga terwujudnya penulisan tesis ini. Penulis menyadari bahwa penulisan tesis ini masih jauh sekali dari sempurna, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan karya ilmiah yang penulis hasilkan untuk yang akan datang.

Akhir kata semoga tesis ini dapat bermanfaat bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Jakarta, 28 Agustus 2018



Supriatiningsih

Penulis

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA
ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini, saya :

Nama : Supriatiningsih
NIM : 14002068
Program Studi : Magsiter Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Software Engineering*
Jenis Karya : Tesis

Demi pengembangan ilmu pengetahuan, dengan ini menyetujui untuk memberikan ijin kepada pihak Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIC Nusa Mandiri) **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalti-Free Right*)** atas karya ilmiah kami yang berjudul : "Penerapan Algoritma Klasifikasi C4.5 Berbasis Adaboost Untuk Prediksi Loyalitas Pelanggan Distributor Pulsa Elektronik" beserta perangkat yang diperlukan (apabila ada).

Dengan **Hak Bebas Royalti Non-Eksklusif** ini pihak STMIC Nusa Mandiri berhak menyimpan, mengalih-media atau *bentuk*-kan, mengelolanya dalam pangkalan data (*database*), mendistribusikannya dan menampilkan atau mempublikasikannya di *internet* atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari kami selama tetap mencantumkan nama kami sebagai penulis/pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak STMIC Nusa Mandiri, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 28 Agustus 2018

Yang menyatakan,



Supriatiningsih

ABSTRAK

Nama : Supriatiningsih
NIM : 14002068
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : *Software Engineering*
Judul : “Penerapan Algoritma Klasifikasi C4.5 Berbasis Adaboost Untuk Prediksi Loyalitas Pelanggan Distributor Pulsa Elektronik”

Pada dasarnya pelanggan mempunyai peranan penting dalam meningkatkan pendapatan perusahaan. Salah satu perusahaan yang memiliki pelanggan dalam jumlah besar adalah perusahaan telekomunikasi seluler. Permasalahan yang muncul dalam loyalitas pelanggan seringkali pelanggan berpindah ke kompetitor lain. Hal ini disebabkan belum terbangunnya hubungan yang baik antara perusahaan dan pelanggan. Untuk itu pada penelitian ini bertujuan untuk membuat pola prediksi pelanggan yang tidak loyal dan berpotensi pindah ke kompetitor lain dengan memanfaatkan proses data mining. Teknik mining yang akan digunakan adalah klasifikasi dengan algoritma C4.5 berbasis AdaBoost untuk meningkatkan akurasi. Identifikasi variabel untuk memprediksi pelanggan yang loyal atau tidak berdasarkan informasi alamat pelanggan, data penjualan pulsa, deposit pulsa dan tanggal mulai berlangganan. Pengujian dilakukan pada salah satu server distributor pulsa elektronik. Dari hasil penelitian yang dilakukan dari tahap awal hingga pengujian, didapatkan pola prediksi loyalitas pelanggan. Akurasi algoritma C4.5 sebesar 69.67% sedangkan akurasi algoritma C4.5 berbasis AdaBoost bernilai 88.00%. Dengan demikian penggunaan algoritma C4.5 berbasis AdaBoost dapat meningkatkan akurasi prediksi loyalitas pelanggan sebesar 18.33% dibandingkan algoritma C4.5.

Kata kunci: Algoritma C4.5, AdaBoost, Prediksi loyalitas pelanggan

ABSTRACT

Name : Supriatiningsih
NIM : 14002068
Study of Program : Master of Computer Science
Levels : Strata Dua (S2)
Concentration : *Software Engineering*
Titel : *“Application-Based AdaBoost C4.5 Algorithm To Predict Customer Loyalty Of Electronic Phone Pulse Distributor”*

Basically customers have an important role in increasing company revenue. One company that has a large number of customers is a cellular telecommunications company. Problems that arise in customer loyalty often move customers to other competitors. This is due to the lack of a good relationship between the company and the customer. For this reason, this study aims to create predictive patterns of customers who are not loyal and have the potential to move to other competitors by utilizing the data mining process. The mining technique that will be used is classification with AdaBoost based C4.5 algorithm to improve accuracy. Variable identification to predict loyal customers or not based on customer address information, pulse sales data, credit deposits and start date. The test was carried out on one of the electronic pulse distributor servers. From the results of research conducted from the initial stage to the testing, a prediction pattern of customer loyalty was obtained. C4.5 algorithm accuracy is 69.67% while AdaBoost based C4.5 algorithm accuracy is 88.00%. Thus the use of AdaBoost based C4.5 algorithm can improve the accuracy of customer loyalty predictions by 18.33% compared to C4.5 algorithm.

Keywords: C4.5 Algorithm, AdaBoost, Prediction of Customer Loyalty

DAFTAR ISI

	Halaman
HALAMAN SAMPUL	i
HALAMAN JUDUL.....	ii
HALAMAN PERNYATAAN ORISINALITAS	iii
HALAMAN PENGESAHAN.....	iv
KATA PENGANTAR.....	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI KARYA.... ..	vi
ILMIAH UNTUK KEPENTINGAN AKADEMIS.....	vii
ABSTRAK.....	viii
ABSTRCT.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN.....	xiv
BAB 1. PENDAHULUAN.....	1
1.1. Latar Belakang Penulisan.....	1
1.2. Identifikasi Masalah.....	3
1.3. Tujuan Penelitian	4
1.4. Ruang Lingkup Penelitian	4
1.5. Hipotesis	4
1.6. Sistematika Penulisan	5
BAB II. LANDASAN TEORI.....	7
2.1. Tinjauan Studi.....	7
2.2. Tinjauan Pustaka.....	9
2.2.1. Loyalitas pelanggan	9
2.2.2. Distributor Pulsa Elektronik	12
2.2.3. Algoritma C4.5	12
2.2.4. AdaBoost.....	15
2.2.5. Cross Validation.....	17
2.3. Tinjauan Organisasi/Obyek Penelitian	18
BAB III. METODOLOGI PENELITIAN.....	20
3.1. Metodologi Penelitian.....	20
3.1.1. Jenis Penelitian	21
3.1.2. Metode Pengumpulan Data.....	22
3.1.3. Eksperimen	22
3.2. Jadwal Penelitian	45
BAB IV. HASIL PENELITIAN DAN PEMBAHASAN.....	46

4.1. Pengukuran Penelitian	46
4.1.1. Hasil Penelitian	46
4.1.2. Pengujian Model	46
4.2. Analisa Hasil Komparasi	55
4.3. Implikasi Penelitian	56
BAB V. PENUTUP	57
5.1. Kesimpulan	57
5.2. Saran	57

DAFTAR PUSTAKA

LEMBAR KONSULTASI

SURAT KETERANGAN RISET/PRAKTEK KERJA LAPANGAN

DAFTAR RIWAYAT HIDUP

LAMPIRAN-LAMPIRAN

DAFTAR TABEL

	Halaman
Tabel I.1. Laporan Pelanggan Non Aktif Pulsa Elektronik.....	2
Tabel II.2. Atribut Prediksi Loyalitas Pelanggan.....	11
Tabel III.1. Atribut dan Data Pelanggan Keseluruhan.....	23
Tabel III.2. Data Eksperimen.....	25
Tabel III.3. Kategori Atribut.....	35
Tabel III.4. Kategori data pelanggan yang tidak aktif.....	27
Tabel III.5. Hasil Perhitungan Gain Information	29
Tabel III.6. Hasil Perhitungan Node 1	29
Tabel III.7. Hasil Perhitungan Node 1.1	31
Tabel III.8. Hasil Perhitungan Node 1.2	32
Tabel III.9. Hasil Perhitungan Node 1.1.1	33
Tabel III.10. Hasil Perhitungan Node 1.2.2	34
Tabel III.11. Bobot Nilai ke 1 Menggunakan AdaBoost.....	42
Tabel III.12. Bobot Nilai ke 2 Menggunakan AdaBoost.....	43
Tabel III.13. Perhitungan nilai class hipotesa 1 dan hipotesa 2.....	43
Tabel III.14. Jadwal Penelitian	45
Tabel IV.1. Model Confusion Matrix.....	47
Tabel IV.2 Hasil Pengujian dengan C4.5 dan C4.5 berbasis AdaBoost	47
Tabel IV.3 Komparasi evaluasi C4.5 dan C4.5 berbasis AdaBoost	55

DAFTAR GAMBAR

	Halaman
Gambar I.1. Grafik Peningkatan Pelanggan Non Aktif.....	3
Gambar II.1. Contoh Klasifikasi Pohon Keputusan	13
Gambar III.1. Kerangka Pemikiran.....	20
Gambar III.2. Desain model decision tree menggunakan Rapid Miner	35
Gambar III.3. Hasil klasifikasi menggunakan algoritma C4.5	35
Gambar III.4. Model pohon keputusan secara lengkap	36
Gambar III.5. Pohon tunggal dari hipotesa 1	37
Gambar III.6. Pohon tunggal dari hipotesa 2	38
Gambar III.7. Pohon tunggal dari hipotesa 3	38
Gambar III.8. Pohon tunggal dari hipotesa 4	39
Gambar III.9. Pohon tunggal dari hipotesa 5	39
Gambar III.10. Pohon tunggal dari hipotesa 6	40
Gambar III.11. Pohon tunggal dari hipotesa 7	40
Gambar III.12. Pohon tunggal dari hipotesa 8	41
Gambar III.13. Pohon tunggal dari hipotesa 9	41
Gambar III.14. Pohon tunggal dari hipotesa 10	42
Gambar III.15. Desain Model algoritma C4.5 berbasis Model AdaBoost.....	44
Gambar III.16. Hasil klasifikasi C4.5 berbasis AdaBoost	45
Gambar IV.1. Pengujian C4.5.....	46
Gambar IV.2. Pengujian C4.5 berbasis Adaboost	46
Gambar IV.3. Hasil <i>Confusion Matrix</i> menggunakan algoritma C4.5	49
Gambar IV.4. Hasil <i>Confusion Matrix</i> C4.5 berbasis Adaboost.....	50
Gambar IV.5. <i>Confusion Matrix</i> menentukan <i>accuracy</i> dengan C4.5.....	51
Gambar IV.6. <i>Confusion Matrix accuracy</i> dengan C4.5 berbasis AdaBoost ..	51
Gambar IV.7. Grafik ROC dengan model algoritma C4.5	52
Gambar IV.8. Grafik ROC dengan model algoritma C4.5 berbasis AdaBoost	53
Gambar IV.9. Komparasi ROC C4.5 dan C4.5 berbasis AdaBoost.....	54

DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Data Real Pelanggan Keseluruhan.....	66
Lampiran 2. Atribut dan Data Pelanggan Keseluruhan.....	70

BAB I

PENDAHULUAN

1.1. Latar Belakang Penulisan

Adanya prospek bisnis, maka distribusi pemasaran pulsa elektronik merupakan peluang bisnis yang menjanjikan. Bisnis pulsa elektronik merupakan salah satu bisnis yang mempunyai tingkat risiko rendah dikarenakan pulsa seluler merupakan kebutuhan utama yang harus ada, tidak mudah rusak, tidak mempunyai waktu kadaluarsa, tidak membutuhkan biaya pengiriman. Salah satu tantangan pada industri telekomunikasi seluler adalah tingginya jumlah pelanggan yang berhenti menggunakan layanan perusahaan dan pindah ke perusahaan kompetitor (Govindaraju, et al, 2008). Kasus pelanggan yang tidak loyal atau berpindah ke kompetitor lain (*churn*) merupakan permasalahan utama yang sering dihadapi perusahaan telekomunikasi karena akan berpengaruh terhadap pendapatan yang diterima oleh perusahaan tersebut. Untuk mendapatkan pelanggan baru membutuhkan biaya lebih besar dibandingkan dengan mempertahankan pelanggan lama (Jahromi, 2009). Perlu adanya satu model prediksi yang akurat sehingga dapat memprediksi pelanggan yang berpindah ke perusahaan lain (Baizal, 2009).

Ada beberapa penelitian dan teknik prediksi pelanggan yang tidak loyal dibuat oleh beberapa peneliti seperti (Liao, 2007) yang memprediksi indikasi pelanggan yang tidak loyal pada beberapa Perusahaan Service Provider menggunakan decision tree. (Jahromi, 2009) yang memprediksi pelanggan yang tidak loyal pada perusahaan provider telekomunikasi di Iran dengan menggunakan model Decision Tree. (Khan, Jamwal, Sepehri, 2010) juga melakukan prediksi pelanggan yang tidak loyal pada perusahaan Internet Service Provider (ISP) dengan menggunakan model Decision Tree dan Neural Network. (Govindaraju, et al, 2008) memprediksi pelanggan yang tidak loyal dengan menggunakan model

decision tree pada perusahaan telekomunikasi yang menyediakan jasa operator telepon seluler di Indonesia yaitu Telkomsel. Dari penelitian sebelumnya telah banyak menggunakan model pohon keputusan untuk memprediksi loyalitas pelanggan pada perusahaan telekomunikasi.

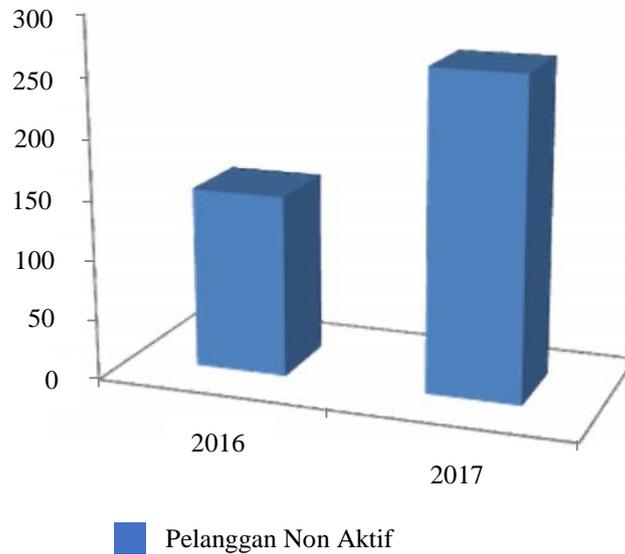
Di dalam penelitian ini penulis menggunakan metode untuk klasifikasi yang nilai akurasi lebih akurat yaitu metode adaboost. Setelah penggunaan citra integral, kedua adalah penggunaan pengklasifikasi yang simple dan efisien yang dibangun dengan memilih sejumlah kecil dari fitur-21 fitur penting yang berasal dari suatu pustaka yang sangat besar dari fitur-fitur potensial menggunakan AdaBoost. Untuk memastikan klasifikasi yang cepat, proses pembelajarannya harus mengecualikan suatu mayoritas besar dari fitur yang tersedia, dan memfokuskan pada suatu set kecil dari fitur yang kritikal. Pemilihan fitur ini dicapai menggunakan algoritma pembelajaran AdaBoost dengan membatasi setiap pengklasifikasi yang lemah untuk bergantung pada fitur tunggal. Sebagai hasilnya pada setiap tingkatan pada proses boosting untuk memilih pengklasifikasi lemah yang baru, dapat ditampilkan sebagai proses seleksi fitur.

Dengan tingginya tingkat persaingan dalam bisnis server pulsa, membuka peluang bagi pelanggan untuk pindah ke kompetitor lain. Di bawah ini merupakan tabel yang menunjukkan data pelanggan yang tidak aktif pada perusahaan telekomunikasi. Data ini diambil dari server distributor pulsa elektronik Sembilan Reload Purwokerto yaitu:

Tabel I.1. Laporan Pelanggan Non Aktif Pulsa Elektronik

Tahun	Jumlah Pelanggan	Pelanggan Non Aktif
2016	155	28%
2017	280	33%

(sumber: server distributor pulsa elektronik Sembilan Reload 2016-2017)



Gambar I.1. Grafik peningkatan pelanggan non aktif

(sumber: server distributor pulsa elektronik Sembilan Reload 2016-2017)

Pada tabel 1.1. menunjukkan mulai tahun 2016 hingga tahun 2017 terjadi kenaikan jumlah pelanggan yang tidak aktif, dikarenakan pelanggan tidak melakukan transaksi pengisian pulsa secara rutin sehingga menjadi penyebab status pelanggan menjadi tidak aktif.

Berdasarkan latar belakang di atas maka peneliti tertarik untuk melakukan penelitian dengan judul “Penerapan Algoritma Klasifikasi C4.5 Berbasis AdaBoots Untuk Prediksi Loyalitas Pelanggan Distributor Pulsa Elektronik”.

1.2. Identifikasi Masalah

Apakah dengan metode klasifikasi metode C4.5 berbasis AdaBoost dapat meningkatkan akurasi untuk memprediksi pola loyalitas pelanggan dilihat dari pelanggan yang aktif dan tidak aktif?

1.3. Tujuan Penelitian

Tujuan dari penelitian ini ingin mengetahui nilai akurasi klasifikasi prediksi loyalitas pelanggan distributor pulsa elektronik menggunakan metode C4.5 berbasis Adaboost. Dapat digunakan untuk memudahkan bagian penjualan

(sales) dalam memberikan keputusan, memberikan sumbangan pemikiran bagi pengembangan teori yang berkaitan dengan prediksi loyalitas pelanggan dengan menggunakan algoritma C4.5 berbasis AdaBoost secara otomatis dan algoritma C4.5 berbasis AdaBoost mampu menjadi alat pendukung pengambil keputusan bagi distributor pulsa elektronik.

1.4. Ruang Lingkup Penelitian

Ruang lingkup berfungsi untuk membatasi pembahasan pada pokok permasalahan saja. Ruang lingkup penelitian dibatasi pada pengujian model algoritma klasifikasi C4.5 dan metode Boosting yaitu AdaBoost. Ruang lingkup data dilakukan pada server distributor pulsa elektronik Sembilan Reload Purwokerto, sedangkan data yang diteliti mulai tahun 2017.

1.5. Hipotesis

Ada 10 hipotesis yang ada pada penelitian ini untuk menentukan prediksi loyalitas pelanggan antara lain sebagai berikut:

1. Jika pelanggan dengan deposit yang banyak maka dinyatakan sebagai pelanggan aktif.
2. Jika pelanggan dengan deposit yang cukup, frekuensinya jarang dan lama langganan baru maka dinyatakan sebagai pelanggan aktif.
3. Jika pelanggan dengan deposit cukup, frekuensi jarang, lama langganan baru maka dinyatakan sebagai pelanggan tidak aktif.
4. Jika pelanggan dengan deposit cukup, frekuensi sedang maka dinyatakan sebagai pelanggan aktif.
5. Jika pelanggan dengan deposit cukup, frekuensi sering maka dinyatakan sebagai pelanggan aktif.

6. Jika pelanggan dengan deposit minim, frekuensi jarang maka dinyatakan sebagai pelanggan tidak aktif.
7. Jika pelanggan dengan deposit minim, frekuensi sedang, area desa maka dinyatakan sebagai pelanggan aktif.
8. Jika pelanggan dengan deposit minim, frekuensi sedang, area kota maka dinyatakan sebagai pelanggan tidak aktif.
9. Jika pelanggan dengan deposit minim, frekuensi sedang, area pinggir kota maka dinyatakan sebagai pelanggan aktif.
10. Jika pelanggan dengan deposit minim, frekuensi sering maka dinyatakan sebagai pelanggan aktif.

1.6. Sistematika Penulisan

Sistematika penulisan tesis ini terdiri dari 5 (lima) bab, dimana tiap bab terdiri dari beberapa sub bab sebagai berikut :

BAB 1 PENDAHULUAN

Pada Bab ini membahas tentang Latar Belakang Penulisan, Identifikasi Masalah, Tujuan Penelitian, Ruang Lingkup Penelitian, Hipotesis dan Sistematika Penulisan.

BAB 2 LANDASAN/KERANGKA PEMIKIRAN

Pada bab ini dibahas teori yang melandasi penelitian, dalam bab ini juga diuraikan Tinjauan Pustaka, Tinjauan Studi, Tinjauan Organisasi/Obyek Penelitian.

BAB 3 METODOLOGI PENELITIAN

Pada bab ini membahas metode pengumpulan data dan eksperimen. Eksperimen merupakan inti pembahasan dari bab ini, yaitu menguji model algoritma C4.5 dan model C4.5 berbasis AdaBost untuk meningkatkan akurasi prediksi loyalitas pelanggan.

BAB 4 HASIL PENELITIAN DAN PEMBAHASAN

Pada bab ini menampilkan hasil dari eksperimen, baik sebelum diterapkan model maupun setelah diterapkan model. Membahas metode, mengukur hasil dengan metode statistik. Hasil kedua model tersebut dibandingkan untuk melihat tingkat akurasi yang paling tinggi. Hasil uji kedua model akan ditampilkan dalam model statistik.

BAB 5 PENUTUP

Pada bab ini membahas kesimpulan dari penelitian, kekurangan serta kelebihan dari model yang digunakan.

BAB II

LANDASAN TEORI

2.1. Tinjauan Studi

Ada beberapa penelitian yang menggunakan model pohon keputusan C4.5 sebagai algoritma untuk memprediksi loyalitas pelanggan berdasarkan data historis:

1. Penelitian yang dilakukan oleh Hehui Qian, Zhiwei Qiu, yaitu penelitian tentang bagaimana memilih fitur yang berguna untuk meningkatkan akurasi dan mengurangi waktu kompleksitas dengan menerapkan algoritma C4.5 untuk memilih fitur harga dan menunjukkan bahwa kinerja C4.5 lebih baik dibandingkan ID3 dalam hal akurasi.(Hehui Qian & Zhiwei Qiu, 2014).
2. Penelitian yang dilakukan oleh Xue Bao, Xin Guan, yaitu penelitian tentang penerapan beberapa model data mining seperti algoritma K-Means dan RS-C4.5 untuk memprediksi akurat dari output minyak mentah. Dari hasil penelitian didapatkan bahwa RS-C4.5 merupakan model yang paling baik untuk memprediksi akurasi output minyak mentah.(Xue Bao & Xin Guan, 2016).
3. Penelitian yang dilakukan oleh Lokesh S. Katore, Bhakti S. Ratnaparkhi, Dr. Jayant S. Umale, yaitu penelitian tentang keakuratan prediksi tergantung pada himpunan parameter keterampilan dan analisis yang relevan pada ini. Dalam karya ini kami mengusulkan keakuratan yang ditingkatkan pendekatan untuk prediksi di mana algoritma C4.5 untuk aturan generasi dan menerapkan korelasi statistik pada hasil.(Lokesh S. Katore, et al, 2015).
4. Penelitian yang dilakukan oleh Aldi Nurzahputra Much Aziz Muslim yaitu penelitian tentang data pemohon kredit yang besar dapat diolah menjadi informasi yang dapat digunakan sebagai pendukung keputusan dalam menentukan permohoanan kredit. Terdapat beberapa algoritma klasifikasi

salah satunya yaitu pohon keputusan atau decision tree. Algoritma C4.5 dapat diterapkan dalam mengklasifikasi permohonan kredit. Penelitian ini menggunakan German Credit Card dataset. Adapun tujuan penelitian ini yaitu meningkatkan akurasi dari algoritma C4.5 dengan menerapkan adaboost dalam mengklasifikasi permohonan kredit dengan membandingkan hasil sebelum dan sesudah diterapkan adaboost. Validasi dalam penelitian ini menggunakan 10 fold cross validation. Sedangkan pengukuran akurasi diukur dengan confusion matrix. Hasil percobaan menunjukkan terdapat peningkatan akurasi 3.7%. Akurasi penerapan algoritma C4.5 saja mencapai 70.5%. Sedangkan akurasi penerapan algoritma C4.5 dengan adaboost mencapai 74.2%. (Aldi Nurzahputra & Much Aziz Muslim, 2017).

5. Penelitian yang dilakukan oleh Kraljevic, Gotovac, yaitu penelitian tentang penerapan beberapa model data mining seperti neural network dan decision tree untuk memprediksi pelanggan yang potensial tidak loyal terhadap perusahaan HT Mostar, yaitu sebuah perusahaan telekomunikasi prabayar di Bosnia & Herzegovina. Beberapa model data mining akan dibandingkan untuk memprediksi pelanggan yang tidak loyal. Dari hasil penelitian didapatkan bahwa decision tree merupakan model yang paling baik untuk memprediksi pelanggan yang tidak loyal pada perusahaan telekomunikasi prabayar (Kraljevic & Gotovac, 2010).
6. Penelitian yang dilakukan oleh Oseman, et al tentang analisa loyalitas pelanggan pada industri telekomunikasi menggunakan decision tree. Penelitian ini dilakukan sebagai salah satu cara untuk mencegah berkurangnya pendapatan Perusahaan dengan menggunakan teknik data mining. Variabel prediksi pelanggan yang loyal dan tidak loyal berdasarkan data pelanggan (usia dan tempat tinggal pelanggan) serta aktivitas pelanggan menggunakan pelayanan (Oseman, 2010).
7. Penelitian yang dilakukan oleh Jin-Hui & Jian-Jun dimana melakukan perbandingan dengan beberapa model peramalan data yang ada seperti decision tree, neural network, SVM, Naive Bayes untuk memprediksi pelanggan yang loyal dan tidak loyal. Dari beberapa model kemudian akan ditentukan model mana yang terbaik. Dari semua penelitian dan model yang

digunakan, terbukti bahwa pohon keputusan dan algoritma C4.5 mampu menyelesaikan masalah yang berhubungan dengan prediksi loyalitas pelanggan dan membuktikan pohon keputusan menghasilkan akurasi yang lebih baik dibandingkan dengan model lain (Jin-Hui & Jian-Jun, 2009).

2.2. Tinjauan Pustaka

Dalam tesis ini, penulis menggunakan referensi dari buku, jurnal dan prosiding untuk menjelaskan model algoritma klasifikasi C4.5 berbasis AdaBoost.

2.2.1. Loyalitas Pelanggan (Customer Loyalty)

Pelanggan (customer) adalah seseorang yang menjadi terbiasa untuk membeli produk. Kebiasaan itu terbentuk melalui pembelian dan interaksi yang sering selama periode waktu tertentu (Griffin, 2005). Orang yang melakukan pembelian tanpa ada hubungan yang kuat dan pembelian berulang tidak dapat dikategorikan sebagai pelanggan melainkan sebagai pembeli.

Loyalitas pelanggan didefinisikan berdasarkan perilaku membeli. Pelanggan yang loyal adalah orang yang melakukan pembelian berulang secara teratur, membeli antar lini produk atau jasa, mereferensikan kepada orang lain, menunjukkan kekebalan terhadap tarikan dari pesaing (Griffin, 2005). Bagi Perusahaan menjaga pelayanan dengan sangat baik akan meningkatkan rasa kepuasan pelanggan sehingga pada akhirnya pelanggan akan memberikan kesetiaan. Kesetiaan pelanggan tidak dapat dipaksakan oleh Perusahaan, kesetiaan diperoleh dari kepuasan terhadap pelayanan yang diberikan Perusahaan seiring berjalannya waktu dan usaha.

Dalam membangun loyalitas pelanggan diperlukan landasan yang kuat. Terdapat beberapa faktor untuk mengukur loyalitas pelanggan (Griffin, 2005) yaitu:

1. Basis klien, merupakan seluruh jumlah pelanggan yang aktif dilihat dari jumlah pelanggan pertama-kali, pelanggan berulang.

2. Tingkat retensi pelanggan baru, menunjukkan persentase pelanggan pertama kali yang kembali melakukan pembelian kedua dalam periode waktu tertentu.
3. Pangsa pelanggan, menunjukkan persentase jumlah pembelian pelanggan terhadap produk atau jasa yang dibelanjakan ke Perusahaan.
4. Jumlah rata-rata pelanggan baru per bulan, merupakan rata-rata jumlah pelanggan pertama-kali membeli di Perusahaan setiap bulannya.
5. Frekuensi pembelian, merupakan frekuensi rata-rata seorang pelanggan membeli dari Perusahaan per tahunnya.
6. Jumlah pembelian rata-rata, menunjukkan jumlah rata-rata yang dibayar atas produk atau jasa pada setiap bulannya.
7. Tingkat peralihan, merupakan persentase tahunan rata-rata pelanggan yang hilang atau menjadi tidak aktif karena suatu alasan.

Kesalahan dalam pemberian layanan kepada pelanggan berdampak terhadap perpindahan pelanggan ke Perusahaan lain. Perilaku pelanggan yang meninggalkan layanan yang diberikan perusahaan disebut churn. Pelanggan yang melakukan churn pada industri telekomunikasi di Eropa berkisar 8-12 % setiap tahun, sedangkan di Amerika berkisar 2-3% per bulannya (Govindaraju, etc all, 2008).

Dalam hal ini pelanggan yang tidak loyal dapat dibagi menjadi dua kelompok utama (Rob, 2005), yaitu:

1. Sukarela (Voluntary Churners)

Sukar untuk ditentukan karena pelanggan yang tidak loyal jenis ini membuat keputusan secara sadar untuk mengakhiri layanan yang digunakan.

2. Tidak Sukarela (Involuntary Churners)

Mudah untuk diidentifikasi karena pelanggan yang menggunakan jasa ditarik sengaja oleh Perusahaan karena beberapa alasan.

Variabel untuk memprediksi pelanggan yang berpindah ke perusahaan lain pada pelanggan provider telepon selular pasca bayar di Bandung (Govindaraju, et al, 2008) antara lain:

1. Data Informasi Pelanggan
 - a. Data Demografi. Berisikan data-data yang dibawa oleh pelanggan saat pertama kali mereka memulai menggunakan layanan dan mendeskripsikan karakteristik pelanggan.
 - b. Data Kontrak. Berisikan data-data kontrak pelanggan yang diperoleh saat kontrak dimulai.
2. Data transaksi atau indeks pelanggan. Dideskripsikan sebagai perilaku pelanggan melakukan transaksi pembelian oleh pelanggan dalam berbagai cara setiap periode tertentu.
3. Data Pembayaran. Dideskripsikan sebagai perilaku pelanggan melakukan pembayaran setiap periode tertentu.
4. Data Call. Dideskripsikan sebagai perilaku komunikasi (incoming atau outgoing).

Dalam penelitian sebelumnya untuk memprediksi pelanggan yang loyal atau tidak pada perusahaan provider di Malaysia menggunakan beberapa atribut sebagai parameter pengukuran antara lain (Oseman, 2010):

Tabel II.1. Atribut Prediksi Loyalitas Pelanggan

Nama Atribut	Sifat Nilai
Lama berlangganan	Kategorikal
Area	Kategorikal
Durasi penggunaan telepon	Kategorikal
Class	Kategorikal

Pada tabel diatas sifat nilai adalah kategorikal. Dimana kategorikal merupakan nilai yang bukan berbentuk angka. Nilai kategorikal merupakan nilai yang diambil dari angka, namun untuk memudahkan proses analisa maka angka tersebut dibagi menjadi beberapa kategori.

2.2.2. Distributor Pulsa Elektronik

Distributor adalah badan perseorangan yang mempunyai kemampuan membeli produk kita dalam jumlah besar dan menyalurkan kepada sub distributor atau pengecer langsung. Dan oleh karena itu distributor tersebut diberi komisi atau diperkenankan menaikkan harga jual untuk mengambil keuntungan (Sameto, 2004). Distributor diperlukan untuk menunjang keberhasilan pemasaran dari suatu produk. Pulsa elektronik adalah dealer resmi voucher elektronik area nasional merupakan distributor penjualan pulsa isi ulang atau voucher elektronik GSM atau CDMA dengan harga rata-rata lebih. Industri telekomunikasi yang berkembang pesat sejalan dengan kebutuhan masyarakat akan pulsa elektronik. Kemudahan dalam pengisian pulsa membuat masyarakat lebih memilih pulsa elektronik daripada menggunakan voucher pulsa. Jumlah pengguna telepon seluler di Indonesia hingga Juni 2010 pelanggan, atau sekitar 80 persen populasi penduduk Indonesia dan 95% pengguna adalah pelanggan prabayar. (Atmosutarno, 2010).

2.2.3. Algoritma C4.5

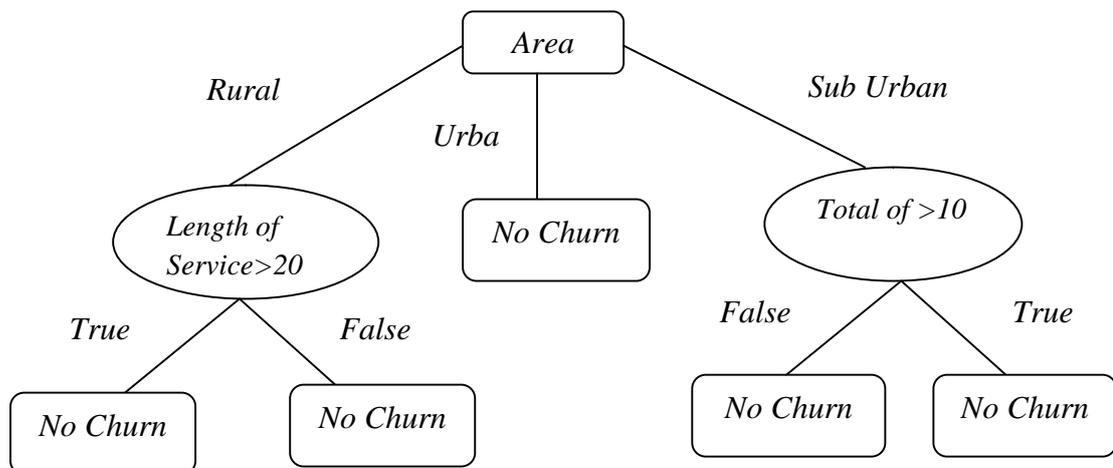
Untuk mengklasifikasikan obyek diperlukan urutan pertanyaan sebelum dapat dibuat kelompoknya. Jawaban dari pertanyaan pertama akan mempengaruhi pertanyaan berikutnya dan seterusnya. Dalam pohon keputusan (decision tree), pertanyaan pertama akan ditanyakan pada simpul akar. Jawaban dari pertanyaan ini dikemukakan dalam cabang-cabang. Jawaban dalam cabang akan disusul dengan pertanyaan kedua lewat simpul berikutnya. Langkah ini akan berakhir disuatu simpul jika sudah jelas kelas atau jenis obyek yang kita cari. Langkah dari pohon keputusan mirip sebuah struktur pohon dimana terdapat node internal (bukan daun) yang mendeskripsikan atribut-atribut, setiap cabang

menggambarkan hasil dari dari atribut yang diuji, dan setiap daun menggambarkan kelas.

Secara umum terdapat beberapa ciri kasus yang cocok untuk diterapkan dalam pohon keputusan (Santosa, 2007) antara lain:

1. Data/Example dinyatakan dengan pasangan atribut dan nilainya.
2. Misalnya atribut satu example adalah temperatur dan nilainya adalah dingin. Biasanya untuk satu example nilai dari satu atribut tidak terlalu banyak jenisnya.
3. Label/Output data biasanya bernilai diskrit.
4. Output ini biasanya bernilai ya atau tidak, sakit atau tidak sakit, diterima atau ditolak. Dalam beberapa kasus mungkin saja outputnya tidak hanya dua kelas. Tetapi penerapan pohon keputusan lebih banyak untuk kasus binary.
5. Data mempunyai missing value.
6. Misalkan untuk beberapa contoh, nilai dari suatu atributnya tidak diketahui. Dalam keadaan seperti ini pohon keputusan masih mampu memberi solusi yang baik.

Berikut ini merupakan contoh pohon keputusan untuk penentuan loyalitas pelanggan (Oseman, et al, 2010):



Gambar II.1. Contoh Klasifikasi Pohon Keputusan

Gambar diatas menjelaskan pohon keputusan yang dibuat berdasarkan hasil klasifikasi dengan terdiri dari tiga atribut yaitu length of service, area, durasi penggunaan telepon.

Pohon keputusan dengan menggunakan algoritma C4.5 dikembangkan pada akhir tahun 1970 oleh seorang peneliti bernama Ross Quinlan. Model pohon keputusan yang dibuat pertama kali adalah ID3 (*Iterave Dichotomiser*) yang kemudian pada tahun 1990 dikembangkan menjadi algoritma C4.5. Algoritma C4.5 merupakan model dari data mining yang berbasis supervised learning.

Tahapan dalam pembuatan sebuah pohon keputusan dengan algoritma C4.5 (Santoso, 2007) yaitu:

1. Menentukan atribut / data training. Atribut biasanya diambil dari data histori yang pernah terjadi sebelumnya yang bisa memisahkan obyek menurut kelasnya. Atribut yang dipilih adalah atribut yang menghasilkan simpul yang paling bersih. Semakin bersih suatu cabang semakin baik.
2. Menentukan atribut mana yang akan dipilih terlebih dahulu, dengan cara memilih atribut yang menghasilkan information gain paling besar. Sebelum menghitung information gain terlebih dahulu menghitung entropi yaitu :

$$\text{Entropy}(y) = \sum_{i=1}^n -p_i * \log^2 p_i \quad \dots\dots\dots (1)$$

Keterangan:

y : sekumpulan obyek dengan label/output y

n : jumlah partisi S

p_i : proporsi dari S_i terhadap S

3. Hitung nilai Gain dengan metode information gain :
4. Untuk menghitung information gain dari output data atau variabel dependent y yang dikelompokkan berdasarkan atribut A, dinotasikan dengan gain (y,A).

.....(2)

$$5. \text{Gain}(y,A) = \text{Entropi}(y) - \sum_{i=0}^n \frac{|y_c|}{|y|} \text{entropi}(y_c)$$

Keterangan:

y : sekumpulan obyek dengan label/output y

A : atribut

n : jumlah nilai atribut

$|y_c|$: jumlah kasus pada masing-masing pecahan atribut

$|y|$: jumlah label/output

6. Ulangi langkah ke-2 hingga semua atribut terpisah atau terpecah.
7. Kriteria pemisahan obyek di setiap simpul berdasarkan information gain yang paling besar.

2.2.4. C4.5 berbasis *AdaBoost*

Boosting adalah pendekatan pada *machine learning* untuk meningkatkan peraturan prediksi yang akurat dengan menggabungkan banyak peraturan yang relatif lemah dan tidak akurat. *Adaptive boosting (adaboost)* merupakan salah satu dari beberapa varian pada algoritma *boosting* (Liu, 2015). *Adaboost* merupakan *ensemble learning* yang sering digunakan pada algoritma *boosting*.

Algoritma *AdaBoost* dari Freund dan Schapire (1995) merupakan algoritma penguat praktis pertama, dan tetap menjadi salah satu yang paling banyak digunakan dan dipelajari, dengan aplikasi di berbagai bidang. *Boosting* bisa dikombinasikan dengan *classifier* algoritma yang lain untuk meningkatkan performa klasifikasi. Tentunya secara intuitif, penggabungan beberapa model akan membantu jika model tersebut berbeda satu sama lain. *Adaboost* dan variannya telah sukses diterapkan pada beberapa bidang (*domain*) karena dasar teorinya yang kuat, prediksi yang akurat, dan kesederhanaan yang besar. Terdapat beberapa jenis dari metode *boosting*, salah satu nya adalah *AdaBoost* yang

kependekan dari Adaptive Boosting. Boosting diperkenalkan pertama kali oleh Freund and Schapire pada tahun 1995. *Boosting* merupakan metode yang dapat digunakan untuk meningkatkan ketelitian dalam proses klasifikasi dan prediksi. Metode *boosting* mengkombinasikan beberapa model yang digunakan untuk saling melengkapi dari masing-masing model yang digunakan sehingga dapat meningkatkan akurasi (Freund & Schapire, 1999).

Metode boosting dapat digunakan pada model klasifikasi atau model penentuan prediksi sehingga menghasilkan output yang lebih baik dari kedua model dasarnya (Witten, 2011). Metode boosting dilakukan secara berulang-ulang (iterative), dimana antara satu iterasi dengan iterasi selanjutnya saling berhubungan. Metode lain yang dapat digunakan untuk meningkatkan akurasi adalah metode Bagging. Dari kedua metode tersebut metode Boosting merupakan metode yang dapat memperbaiki tingkat kesalahan klasifikasi. AdaBoost merupakan algoritma boosting yang paling sering digunakan. Berbeda dengan metode untuk meningkatkan akurasi lainnya, metode AdaBoost bekerja dengan data yang sama. Setiap kali pembuatan pohon, data yang digunakan tetap seperti semula tetapi memiliki sebaran bobot yang berbeda dalam tiap iterasi. Penggunaan bobot dilakukan pada saat proses penggabungan dari dugaan akhir dari banyak pohon yang dihasilkan (Sartono & Syafitri, 2010).

Sebagai contoh penggunaan metode boosting, misalkan seorang pasien menderita sakit dengan gejala-gejala tertentu kemudian melakukan konsultasi pada beberapa dokter. Masing-masing dokter memberikan diagnosa. Diagnosa tersebut diberikan bobot nilai berdasarkan keakuratan diagnosa yang diambil dari masa lalu. Diagnosa akhir dibuat berdasarkan bobot nilai dari penggabungan beberapa diagnosa. Metode boosting bekerja dengan menggunakan data yang sama. Setiap pembuatan pohon keputusan menggunakan data yang sama tetapi memiliki bobot atau nilai yang berbeda pada setiap iterasinya. Pemberian bobot juga dilakukan pada saat proses penggabungan diagnosa akhir dari beberapa pohon keputusan yang dihasilkan.

Tahapan algoritma C4.5 berbasis AdaBoost menggunakan data training terdiri atas n pengamatan, dengan y sebagai peubah respon dan memiliki k kelas.

Apabila ingin menggunakan pohon keputusan dengan algoritma boosting sebanyak M iterasi. Tahapan Algoritma C4.5 berbasis AdaBoost (Sartono & Syafitri, 2010) sebagai berikut:

1. Penentuan awal bobot setiap pengamatan yaitu $w_i = 1/n$ untuk semua $i=1,2,\dots,n$
2. Andaikan m adalah nomor iterasi, maka untuk $m = 1,2,\dots,M$, lakukan proses berikut :
 - a. susun pohon tunggal dengan memperhatikan bobot sebesar w_i
 - b. hitung tingkat kesalahan klasifikasi (catatan: I adalah fungsi indikator bernilai 1 atau 0, dan diasumsikan $e_m < 0.5$ dan jika tidak maka proses berhenti)

$$e = \frac{\sum_{i=1}^n w_i I(y_i \neq \hat{y}_i)}{\sum_{i=1}^n w_i} \dots\dots\dots (3)$$

hitung $\alpha_m = \frac{1}{e_m} e_m$

- c. tentukan bobot yang baru untuk setiap pengamatan menjadi $w_i = w_i \alpha_m$ untuk pengamatan yang salah klasifikasi, sedangkan untuk pengamatan yang diduga dengan tepat maka bobotnya tetap.
3. Dugaan akhir adalah kelas k yang memiliki nilai terbesar dari

$$m: y_{im} = y_i^{\log \alpha_m} \dots\dots\dots (4)$$

Dikarenakan nilai $e_m > 0.5$ maka α_m akan selalu lebih dari satu. Dengan demikian pada setiap penambahan iterasi, bobot dari pengamatan yang mengalami salah klasifikasi akan membesar. Hal tersebut merupakan ide dasar dari boosting, yaitu bahwa pada setiap iterasi model pohon memberikan bobot yang lebih besar kepada data yang lebih sulit. Pada iterasi berikutnya pohon yang dibentuk akan terkonsentrasi.

2.2.5. Cross Validation

Validasi silang (cross Validation) merupakan suatu metode statistik yang digunakan menganalisa dan mengukur keakuratan hasil percobaan pada data yang independen. Metode ini membagi sebuah data menjadi beberapa subdata yang selanjutnya subdata satu digunakan untuk mengkonfirmasi kebenaran subdata yang lain.

K-fold cross validation adalah salah satu metode cross validation yang membagi data menjadi k subdata. Salah satu subbagian data dijadikan sebagai validator dan testing sedangkan K-1 data digunakan sebagai data pelatihan. Proses diatas dilakukan berulang sebanyak K kali untuk setiap subbagian data. Hasil dari pengujian adalah rata-rata dari K kali pengujian pada data tersebut. Teknik validasi silang yang memiliki akurasi paling baik menggunakan 10 fold cross validation (Witten, et al, 2011). penelitian ini kami menggunakan 10 fold cross validation.

2.3. Tinjauan Organisasi/Obyek Penelitian

Sembilan Reload Purwokerto merupakan salah satu perusahaan distributor server pulsa elektronik yang berdiri pada tanggal 25 Oktober 2015 , didirikan oleh Rifai Abdul Ghani, yang berada pada perumahan Purwokerto. Perusahaan ini menggunakan multi chip server guna mendukung proses pengisian pulsa secara otomatis kepada pelanggan. Dengan multi chip server, perusahaan dapat menyediakan semua produk dari pulsa elektronik dengan sistem satu deposit untuk pengisian banyak operator dimana transaksi langsung dilakukan oleh server dan dapat dilakukan 24 jam setiap hari dengan berbasis sms.

Konsumen yang berminat menjadi pelanggan Sembilan Reload melakukan pendaftaran dengan mengisi formulir yang berisi data-data calon pelanggan atau agen dan bersedia mengikuti aturan yang diberikan oleh perusahaan. Pelanggan dapat menggunakan pelayanan dari Sembilan Reload untuk kepentingan usaha atau pribadi. Setiap pelanggan diminta menyimpan sejumlah dana (deposit) di

awal sebesar Rp. 100.000. Setiap transaksi pengisian pulsa yang dilakukan oleh pelanggan akan mengurangi saldo deposit yang dimiliki oleh pelanggan. Apabila saldo deposit pelanggan tidak mencukupi maka transaksi pengisian pulsa akan gagal.

Transaksi pengisian pulsa membutuhkan peralatan telekomunikasi guna mendukung keberhasilan pengiriman pulsa, seperti berikut ini:

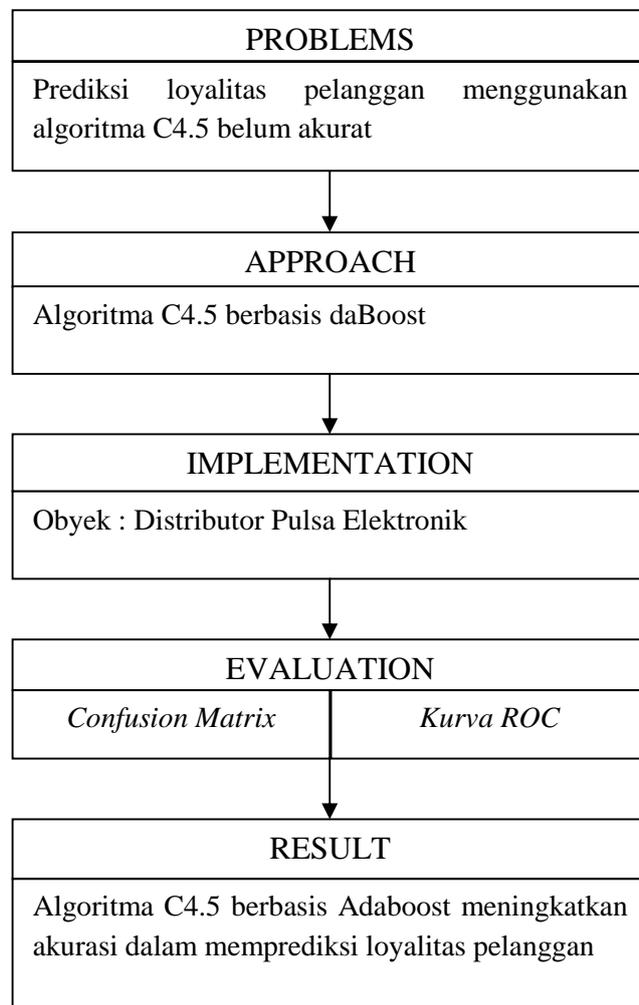
1. Router: D-Link 10 port
2. Modem : Internet D-Link, Modempool Wavecom Q2406 8Port
3. Handphone: Nokia 6100, Siemens C55, Siemens M55
4. Kabel Data: DKU-5, Modem USB
5. Koneksi Internet: Speedy unlimited

BAB III

METODOLOGI PENELITIAN

3.1. Metodologi Penelitian

Penelitian ini dilakukan dalam beberapa tahapan mulai dari pra-proses data, klasifikasi, validasi, akurasi. Berikut kerangka pemikiran pada Gambar III.1.



Gambar III.1 Kerangka Pemikiran

3.1.1. Jenis Penelitian

Jenis penelitian yang digunakan dalam penelitian ini adalah model penelitian eksperimen. Model eksperimen merupakan metode penelitian yang digunakan untuk mencari pengaruh perlakuan tertentu (Sugiyono, 2010). Penelitian ini bertujuan untuk melakukan perbandingan dan evaluasi pada algoritma klasifikasi data mining. Penelitian eksperimen ini menggunakan penerapan algoritma C4.5 berbasis AdaBoost untuk meningkatkan hasil akurasi prediksi pelanggan yang loyal terhadap perusahaan berdasarkan parameter letak demografi dan aktifitas transaksi pada PT. Sembilan Reload Purwokerto.

3.1.2. Metode Pengumpulan Data

Metode pengumpulan data yang digunakan oleh penulis dalam memprediksi loyalitas pelanggan adalah data primer : sumber data yang diperoleh secara langsung dari sumber, misalnya data - data pelanggan, aktivitas transaksi dan hasil interview.

3.1.3. Eksperimen

Dalam melakukan eksperimen digunakan data historis untuk membangun model prediksi. Penulis melakukan pengumpulan data untuk melakukan eksperimen berawal dari data bulan juni, juli dan agustus 2017 sebagai data eksperimen. Penulis menggunakan 643 data untuk data eksperimen.

Selain data, dalam melakukan eksperimen, penulis menggunakan perangkat keras dan perangkat lunak untuk membantu pengolahan data, dengan spesifikasi

komputer sebagai berikut:

1. Hardware :

Processor	Intel® Atom™
RAM	1 GB
Harddisk	160 GB

2. Software :

Sistem Operasi Windows XP Home Edition Version 2002

Service Pack 3

Tools RapidMiner

Ada beberapa tahap yang dilakukan dalam melakukan eksperimen model ini, yaitu:

1. Pengumpulan data
2. Pengembangan model
3. Pengujian model pada data eksperimen
4. Evaluasi dan validasi data eksperimen

Dibawah ini merupakan penjabaran mengenai tahapan melakukan eksperimen data:

1. Pengumpulan Data

Data pelanggan diambil dari lembar kerja pada database Distributor Pulsa Elektronik tahun 2016-2017. Apabila dilihat pada data tersebut terdapat beberapa atribut yang terduplikasi dan tidak digunakan. Berikut Tabel data pelanggan yang telah diketahui status aktif atau tidak sebagai pelanggan.

Tabel III.1. Atribut dan data pelanggan keseluruhan

Area	Deposit	LamaLgn	Frekuensi	Class
Desa	banyak	baru	sedang	aktif
Desa	banyak	baru	sering	aktif
Kota	banyak	baru	sering	aktif
Kota	banyak	baru	sedang	aktif
Kota	banyak	baru	jarang	aktif
Desa	cukup	baru	sedang	aktif
Desa	cukup	baru	jarang	aktif
Kota	cukup	baru	sedang	aktif
Kota	cukup	baru	sering	aktif

Kota	cukup	baru	jarang	aktif
Desa	minim	baru	sedang	aktif
Kota	minim	baru	sering	aktif
Desa	banyak	lama	sering	aktif
Desa	banyak	lama	sedang	aktif
Kota	banyak	lama	sedang	aktif
Kota	banyak	lama	jarang	aktif
Kota	banyak	lama	sering	aktif
Desa	cukup	lama	sering	aktif
Kota	cukup	lama	sering	aktif
Kota	minim	lama	sering	aktif
Desa	minim	baru	jarang	aktif
Desa	banyak	baru	jarang	aktif
Desa	cukup	lama	sedang	aktif
Pinggirkota	banyak	lama	sedang	aktif
Pinggirkota	banyak	lama	sering	aktif
Pinggirkota	banyak	lama	jarang	aktif
Pinggirkota	banyak	baru	jarang	aktif
Pinggirkota	banyak	baru	sedang	aktif
Pinggirkota	banyak	baru	sering	aktif
Pinggirkota	cukup	lama	sedang	aktif
Pinggirkota	cukup	baru	sedang	aktif
Pinggirkota	cukup	baru	jarang	aktif
Pinggirkota	cukup	lama	jarang	aktif
Pinggirkota	cukup	baru	sering	aktif
Pinggirkota	minim	lama	sedang	aktif
Pinggirkota	minim	baru	sedang	aktif
Pinggirkota	minim	lama	sering	aktif
Pinggirkota	minim	baru	sering	aktif
Pinggirkota	minim	baru	jarang	aktif
Desa	minim	baru	jarang	tidak
Kota	minim	baru	sedang	tidak
Kota	minim	baru	jarang	tidak
Kota	cukup	lama	jarang	tidak
Desa	minim	lama	jarang	tidak
Desa	minim	lama	sedang	tidak
Kota	minim	lama	jarang	tidak
Desa	cukup	lama	jarang	tidak
Kota	minim	lama	sedang	tidak
Desa	cukup	baru	jarang	tidak
Kota	cukup	lama	sedang	tidak

Pinggirkota	cukup	lama	jarang	tidak
Pinggirkota	minim	lama	jarang	tidak
Pinggirkota	minim	baru	jarang	tidak
Pinggirkota	minim	baru	sedang	tidak
Kota	banyak	lama	jarang	aktif
pinggirKota	banyak	lama	jarang	aktif
Desa	banyak	baru	jarang	aktif
Desa	banyak	baru	sedang	aktif
Desa	cukup	lama	sering	aktif
Kota	cukup	baru	sering	aktif
Desa	cukup	baru	sering	aktif
Desa	cukup	baru	sedang	aktif
Kota	cukup	lama	jarang	aktif
Kota	cukup	baru	jarang	aktif
Desa	cukup	baru	jarang	aktif
pinggirKota	minim	baru	sedang	aktif
Kota	minim	baru	jarang	tidak
Kota	minim	lama	jarang	tidak
Desa	minim	lama	jarang	tidak
Desa	minim	lama	sedang	aktif
Desa	minim	baru	sedang	aktif
pinggirKota	minim	lama	sedang	aktif
pinggirKota	minim	lama	sering	aktif
Kota	minim	baru	sering	aktif
Desa	minim	baru	sering	aktif
pinggirKota	minim	lama	jarang	aktif
Kota	minim	baru	jarang	tidak
Desa	minim	baru	jarang	aktif
Kota	banyak	lama	jarang	aktif
pinggirKota	banyak	lama	jarang	aktif
Desa	banyak	baru	jarang	aktif
Desa	banyak	baru	sedang	aktif
Desa	cukup	lama	sering	aktif
Kota	cukup	baru	sering	aktif
Desa	cukup	baru	sering	aktif
Desa	cukup	baru	sedang	aktif
Kota	cukup	lama	jarang	aktif
Kota	cukup	baru	jarang	aktif
Desa	cukup	baru	jarang	aktif
pinggirKota	minim	baru	sedang	aktif
Kota	minim	baru	jarang	tidak

Kota	minim	lama	jarang	tidak
Desa	minim	lama	jarang	tidak
Desa	minim	lama	sedang	aktif
Desa	minim	baru	sedang	aktif
pinggirKota	minim	lama	sedang	aktif
pinggirKota	minim	lama	sering	aktif
Kota	minim	baru	sering	aktif
Desa	minim	baru	sering	aktif
pinggirKota	minim	lama	jarang	aktif
Kota	minim	baru	jarang	aktif
Desa	minim	baru	jarang	aktif
Desa	banyak	baru	sedang	aktif
Desa	banyak	baru	sering	aktif
Kota	banyak	baru	sering	aktif
Kota	banyak	baru	sedang	aktif
Kota	banyak	baru	jarang	aktif
Desa	cukup	baru	sedang	aktif
Desa	cukup	baru	jarang	aktif
Kota	cukup	baru	sedang	aktif
Kota	cukup	baru	sering	aktif
Kota	cukup	baru	jarang	aktif
Desa	minim	baru	sedang	aktif
Kota	minim	baru	sering	aktif
Desa	banyak	lama	sering	aktif
Desa	banyak	lama	sedang	aktif
Kota	banyak	lama	sedang	aktif
Kota	banyak	lama	sering	aktif
Desa	cukup	lama	sering	aktif
Kota	cukup	lama	sering	aktif
Kota	minim	lama	sering	aktif
Desa	minim	baru	jarang	aktif
Desa	banyak	baru	jarang	aktif
Desa	cukup	lama	sedang	aktif
Pinggirkota	banyak	lama	sedang	aktif
Pinggirkota	banyak	lama	sering	aktif
Pinggirkota	banyak	lama	jarang	aktif
Pinggirkota	banyak	baru	jarang	aktif
Pinggirkota	banyak	baru	sedang	aktif
Pinggirkota	banyak	baru	sering	aktif
Pinggirkota	cukup	lama	sedang	aktif
Pinggirkota	cukup	baru	sedang	aktif

Pinggirkota	cukup	baru	jarang	aktif
Pinggirkota	cukup	lama	jarang	aktif
Pinggirkota	cukup	baru	sering	aktif
Pinggirkota	minim	lama	sedang	aktif
Pinggirkota	minim	baru	sedang	aktif
Pinggirkota	minim	lama	sering	aktif
Pinggirkota	minim	baru	sering	aktif
Pinggirkota	minim	baru	jarang	aktif
Desa	minim	baru	jarang	tidak
Kota	minim	baru	sedang	tidak
Kota	minim	baru	jarang	tidak
Kota	cukup	lama	jarang	tidak
Desa	minim	lama	jarang	tidak
Desa	minim	lama	sedang	tidak
Kota	minim	lama	jarang	tidak
Desa	cukup	lama	jarang	tidak
Desa	banyak	baru	sedang	aktif
Desa	banyak	baru	sering	aktif
Kota	banyak	baru	sering	aktif
Kota	banyak	baru	sedang	aktif
Kota	banyak	baru	jarang	aktif
Desa	cukup	baru	sedang	aktif
Desa	cukup	baru	jarang	aktif
Kota	cukup	baru	sedang	aktif
Kota	cukup	baru	sering	aktif
Kota	cukup	baru	jarang	aktif
Desa	minim	baru	sedang	aktif
Kota	minim	baru	sering	aktif
Desa	banyak	lama	sering	aktif
Desa	banyak	lama	sedang	aktif
Kota	banyak	lama	sedang	aktif
Kota	banyak	lama	jarang	aktif
Kota	banyak	lama	sering	aktif
Desa	cukup	lama	sering	aktif
Kota	cukup	lama	sering	aktif
Kota	minim	lama	sering	aktif
Desa	minim	baru	jarang	aktif
Desa	banyak	baru	jarang	aktif
Desa	cukup	lama	sedang	aktif
Pinggirkota	banyak	lama	sedang	aktif
Pinggirkota	banyak	lama	sering	aktif

Pinggirkota	banyak	lama	jarang	aktif
Pinggirkota	banyak	baru	jarang	aktif
Pinggirkota	banyak	baru	sedang	aktif
Pinggirkota	banyak	baru	sering	aktif
Pinggirkota	cukup	lama	sedang	aktif
Pinggirkota	cukup	baru	sedang	aktif
Pinggirkota	cukup	baru	jarang	aktif
Pinggirkota	cukup	lama	jarang	aktif
Pinggirkota	cukup	baru	sering	aktif
Pinggirkota	minim	lama	sedang	aktif
Pinggirkota	minim	baru	sedang	aktif
Pinggirkota	minim	lama	sering	aktif
Pinggirkota	minim	baru	sering	aktif
Pinggirkota	minim	baru	jarang	aktif
Desa	minim	baru	jarang	tidak
Kota	minim	baru	sedang	tidak
Kota	minim	baru	jarang	tidak
Kota	cukup	lama	jarang	tidak
Desa	minim	lama	jarang	tidak
Desa	minim	lama	sedang	tidak
Kota	minim	lama	jarang	tidak
Desa	cukup	lama	jarang	tidak
Kota	minim	lama	sedang	tidak
Desa	cukup	baru	jarang	tidak
Kota	cukup	lama	sedang	tidak
Pinggirkota	cukup	lama	jarang	tidak
Pinggirkota	minim	lama	jarang	tidak
Pinggirkota	minim	baru	jarang	tidak
Pinggirkota	minim	baru	sedang	tidak
Kota	banyak	lama	jarang	aktif
pinggirKota	banyak	lama	jarang	aktif
Desa	banyak	baru	jarang	aktif
Desa	banyak	baru	sedang	aktif
Desa	cukup	lama	sering	aktif
Kota	cukup	baru	sering	aktif
Desa	cukup	baru	sering	aktif
Desa	cukup	baru	sedang	aktif
Kota	cukup	lama	jarang	aktif
Kota	cukup	baru	jarang	aktif
Desa	cukup	baru	jarang	aktif
pinggirKota	minim	baru	sedang	aktif

Kota	minim	baru	jarang	tidak
Kota	minim	lama	jarang	tidak
Desa	minim	lama	jarang	tidak
Desa	minim	lama	sedang	aktif
Desa	minim	baru	sedang	aktif
pinggirKota	minim	lama	sedang	aktif
pinggirKota	minim	lama	sering	aktif
Kota	minim	baru	sering	aktif
Desa	minim	baru	sering	aktif
pinggirKota	minim	lama	jarang	aktif
Kota	minim	baru	jarang	tidak
Desa	minim	baru	jarang	aktif
Kota	minim	lama	sedang	tidak
Desa	cukup	baru	jarang	tidak
Kota	cukup	lama	sedang	tidak
Pinggirkota	cukup	lama	jarang	tidak
Pinggirkota	minim	lama	jarang	tidak
Pinggirkota	minim	baru	jarang	tidak
Pinggirkota	minim	baru	sedang	tidak
Desa	cukup	baru	sering	aktif
Kota	cukup	lama	jarang	aktif
Kota	cukup	baru	jarang	aktif
pinggirKota	minim	baru	sedang	aktif
Kota	minim	baru	jarang	tidak
Kota	minim	lama	jarang	tidak
Desa	minim	lama	jarang	tidak
Desa	minim	lama	sedang	aktif
Desa	minim	baru	sedang	aktif
pinggirKota	minim	lama	sedang	aktif
Desa	minim	baru	sering	aktif
pinggirKota	minim	lama	jarang	aktif
Kota	minim	baru	jarang	tidak
Desa	minim	baru	jarang	aktif
Desa	minim	lama	jarang	tidak
Desa	minim	lama	sedang	aktif
Desa	minim	baru	sedang	aktif
pinggirKota	minim	lama	jarang	aktif
Pinggirkota	minim	lama	jarang	tidak

Untuk itu diperlukan teknik untuk mem-filter data training atau dikenal dengan teknik preprocessing (Witten, 2011). Dimana hasil dari teknik preprocessing menghasilkan atribut yang berbeda. Berikut tabel hasil teknik preprocessing (Han & Kamber, 2006) yaitu:

1. Data cleansing digunakan untuk membersihkan nilai yang kosong, tidak konsisten atau ada data yang kosong (missing value dan noisy).
2. Data integration digunakan untuk menyatukan tempat penyimpanan yang berbeda-beda ke dalam satu tempat penyimpanan yang sama. Dalam hal ini data diambil dari data warehouse yaitu data pelanggan dan data transaksi penjualan.
3. Data reduction, dimana jumlah data tupel yang digunakan untuk data training terlalu besar dan hanya beberapa atribut yang diperlukan, sehingga atribut yang tidak perlu akan dihapus. Apabila terdapat tupel yang sama maka hanya menggunakan satu tupel saja. Dimana akan terlihat seperti tabel di bawah ini:

Tabel III.2. Data eksperimen

area	deposit	lamaLgn	frekuensi	class
desa	banyak	baru	sedang	aktif
desa	banyak	baru	sering	aktif
kota	banyak	baru	sering	aktif
kota	banyak	baru	sedang	aktif
kota	banyak	baru	jarang	aktif
desa	cukup	baru	sedang	aktif
desa	cukup	baru	jarang	aktif
kota	cukup	baru	sedang	aktif
kota	cukup	baru	sering	aktif
kota	cukup	baru	jarang	aktif
desa	minim	baru	sedang	aktif
kota	minim	baru	sering	aktif
desa	banyak	lama	sering	aktif

desa	banyak	lama	sedang	aktif
kota	banyak	lama	sedang	aktif
kota	banyak	lama	jarang	aktif
kota	banyak	lama	sering	aktif
desa	cukup	lama	sering	aktif
kota	cukup	lama	sering	aktif
kota	minim	lama	sering	aktif
desa	minim	baru	jarang	aktif
desa	banyak	baru	jarang	aktif
desa	cukup	lama	sedang	aktif
pinggirkota	banyak	lama	sedang	aktif
pinggirkota	banyak	lama	sering	aktif
pinggirkota	banyak	lama	jarang	aktif
pinggirkota	banyak	baru	jarang	aktif
pinggirkota	banyak	baru	sedang	aktif
pinggirkota	banyak	baru	sering	aktif
pinggirkota	cukup	lama	sedang	aktif
pinggirkota	cukup	baru	sedang	aktif
pinggirkota	cukup	baru	jarang	aktif
pinggirkota	cukup	lama	jarang	aktif
pinggirkota	cukup	baru	sering	aktif
pinggirkota	minim	lama	sedang	aktif
pinggirkota	minim	baru	sedang	aktif
pinggirkota	minim	lama	sering	aktif
pinggirkota	minim	baru	sering	aktif
pinggirkota	minim	baru	jarang	aktif
desa	minim	baru	jarang	tidak
kota	minim	baru	sedang	tidak
kota	minim	baru	jarang	tidak
kota	cukup	lama	jarang	tidak
desa	minim	lama	jarang	tidak
desa	minim	lama	sedang	tidak
kota	minim	lama	jarang	tidak
desa	cukup	lama	jarang	tidak
kota	minim	lama	sedang	tidak
desa	cukup	baru	jarang	tidak
kota	cukup	lama	sedang	tidak
pinggirkota	cukup	lama	jarang	tidak
pinggirkota	minim	lama	jarang	tidak
pinggirkota	minim	baru	jarang	tidak
pinggirkota	minim	baru	sedang	tidak

desa	banyak	baru	sering	aktif
desa	banyak	lama	sering	aktif
desa	cukup	lama	sering	aktif

Nilai dari semua atribut yang ada di tabel 3.2, merupakan nilai kategorikal dan bukan nilai angka. Berikut ini tabel kategori atribut:

Tabel III.3. Kategori atribut

Atribut	Nilai Angka	Kategori
Area	Kota	Kota
	Pinggir kota	pinggirKota
	Desa	Desa
Lama berlangganan	< 1 tahun	Baru
	>= 1 tahun	Lama
Frekuensi transaksi	1-10	Sering
	11-150	Sedang
	>150	Jarang
Saldo Deposit	<= Rp.10.000	Minim
	>Rp.10.000, < Rp.100.000	Cukup
	>=Rp.100.000	Banyak

Pada tabel di atas disebutkan bahwa ada 15 pelanggan yang tidak aktif . Berikut kategori data pelanggan yang tidak aktif dalam bentuk tabel.

Tabel III.4. Kategori data pelanggan yang tidak aktif

No	Frekuensi	Deposit	LamaLgn	area	Class
1	Jarang	Minim	Baru	desa	Tidak
2	Sedang	Minim	Baru	kota	Tidak
3	Jarang	Minim	Baru	kota	Tidak
4	Jarang	Cukup	Lama	kota	Tidak
5	jarang	Minim	lama	desa	Tidak
6	sedang	Minim	lama	desa	Tidak
7	jarang	Minim	lama	kota	Tidak
8	jarang	Cukup	lama	Desa	Tidak
9	sedang	Minim	lama	Kota	Tidak
10	jarang	Cukup	baru	Desa	Tidak
11	sedang	Cukup	lama	Kota	Tidak
12	jarang	Cukup	lama	pinggirkota	Tidak
13	jarang	Minim	lama	pinggirkota	Tidak
14	jarang	Minim	baru	pinggirkota	Tidak
15	sedang	Minim	baru	pinggirkota	Tidak

2. Tahap Modelling

Model yang digunakan untuk memprediksi pelanggan yang tidak loyal menggunakan algoritma C4.5 dan model AdaBoost.

A. Algoritma C4.5

Seperti yang telah dijelaskan sebelumnya, ada beberapa tahap yang harus dilalui dalam membentuk pohon keputusan, dan algoritma yang digunakan adalah algoritma C4.5

Untuk membuat pohon keputusan, dibutuhkan data training. Dimana data training diambil dari bulan juni sampai agustus 2017 dimana terdapat 54 data.

- 1) Menghitung jumlah kasus class aktif dan jumlah class tidak aktif. Entropi dari semua kasus dan kasus yang dibagi berdasarkan atribut area, deposit, lama berlangganan, frekuensi transaksi. Kemudian lakukan penyeleksian atribut dengan menghitung Gain tertinggi.

Baris Total kolom Entropi dihitung sebagai berikut:

$$\text{Entropi (Total)} = \left(- \frac{39}{54} * \log_2 \left(\frac{39}{54} \right) \right) + \left(- \frac{15}{54} * \log_2 \left(\frac{15}{54} \right) \right)$$

$$\text{Entropi (Total)} = 0.852$$

- 2) Dengan menggunakan persamaan Gain, dapat menghitung jumlah kasus dari masing-masing atribut dan hitung gain. Hasil perhitungan dapat dilihat pada tabel di bawah ini:

Tabel III.5. Hasil perhitungan gain information

Atribut	Entropy	Information Gain
Area	0.8389722	0.013432982
Deposit	0.63528352	0.21712166
Lama Berlangganan	0.83142275	0.020982427
Frekuensi	0.68080834	0.171596836

Dari tabel di atas diketahui bahwa atribut deposit memiliki nilai entropi terendah atau information gain tertinggi sehingga akan menempati node tertinggi yaitu node 1. Atribut deposit terdiri dari 3 bagian yaitu banyak, minim dan cukup. Untuk menentukan bagian mana yang akan menempati node berikutnya, maka data akan diolah kembali

Tabel III.6. Hasil perhitungan node 1

Node			Jumlah kasus (S)	Aktif (S1)	Tidak (S2)	Entropy	Gain	Ket
1.1	Deposit – Cukup		18	13	5	0.85241		
	Frekuensi	Jarang	8	4	4	1	0.191	Gain Tertinggi Node 1.1
		Sedang	6	5	1	0.65002		
		Sering	4	4	0	0		
		Desa	6	4	2	0.9183		
		PinggirKota	6	5	1	0.65002		

		Kota	6	4	2	0.9183		
	Lama	Baru	9	8	1	0.50326	0.105	
	Berlangganan	Lama	9	5	4	0.99108		
1.2	Deposit – Minim		19	9	10	0.998		
	Frekuensi	Jarang	8	2	6	0.81128	0.293	Gain Tertin ggi Node
		Sedang	7	3	4	0.98523		
		Sering	4	4	0	0		
	Area	Desa	5	2	3	0.97095	0.051	
		PinggirKota	8	5	3	0.95443		
		Kota	6	2	4	0.9183		
	Lama	Baru	11	6	5	0.99403	0.021	
	Berlangganan	Lama	8	3	5	0.95443		
1.3	Deposit– Banyak		17	17	0	0	0	Aktif

Dari tabel di atas dapat diketahui bahwa yang akan menjadi node 1.1 dan node 1.2 adalah atribut frekuensi dilihat dari nilai information gain tertinggi. Untuk atribut deposit–banyak secara otomatis menghasilkan class aktif dikarenakan semua atribut bernilai aktif, sedangkan untuk atribut deposit-cukup dan deposit-minim akan dipecah menjadi node 1.1 dan node 1.2 berdasarkan atribut frekuensi.

Berikut perhitungan node 1.1 dan 1.2 berdasarkan data sebelumnya, yaitu:

Tabel III.7. Hasil perhitungan node 1.1

Node			Jumlah kasus (S)	Aktif (S1)	Tidak (S2)	Entropy	Gain	Ket
1.1	Deposit - Cukup							

1.1.1	Frekuensi	Jarang	8	4	4	1		
	Area	Desa	3	1	2	0.9183	0.061	
		PinggirKota	3	2	1	0.9183		
		Kota	2	1	1	1		
	Lama Berlangg anan	Baru	4	3	1	0.81128	0.189	Gain Tertinggi Node 1.1.1
		Lama	4	1	3	0.81128		
1.1.2	Frekuensi	Sedang	6	5	1	0.65002		
	Area	Desa	2	2	0	0	0.317	
		PinggirKota	2	2	0	0		
		Kota	2	1	1	1		
	Lama Berlangganan	Baru	3	3	0	0	0.191	Aktif
		Lama	3	2	1	0.9183		
1.1.3	Frekuensi	Sering	4	4	0	0	0	Aktif

Dari tabel diatas diketahui untuk node 1.1 pecahan dari atribut deposit - cukup yang memiliki nilai gain tertinggi adalah atribut frekuensi sehingga cabang berikutnya dipecah berdasarkan atribut frekuensi. Untuk frekuensi -sering secara otomatis akan menjadi class aktif, untuk frekuensi-sedang tidak akan dipecah lagi maka secara otomatis akan menjadi class aktif sedangkan frekuensi -jarang akan dipecah lagi.

Tabel III.8. Hasil perhitungan node 1.2

Node			Jumlah kasus (S)	Aktif (S1)	Tidak (S2)	Entropy	Gain	Ket
1.2	Deposit - Minim							
1.2.1	Frekuensi	Jarang	8	2	6	0.81128		
	Area	Desa	3	1	2	0.9183	0.123	
		PinggirKota	3	1	2	0.9183		
		Kota	2	0	2	0		

	Lama	Baru	5	2	3	0.97095	0.204	Tidak
	Berlangganan	Lama	3	0	3	0		
1.2.2	Frekuensi	Sedang	7	3	4	0.98523		
	Area	Desa	2	1	1	1	0.306	Gain Tertinggi Node 1.2.2
		PinggirKota	3	2	1	0.9183		
		Kota	2	0	2	0		
	Lama	Baru	4	1	3	0.81128	0.128	
	Berlangganan	Lama	3	1	2	0.9183		
1.2.3	Frekuensi	Sering	4	4	0	0		Aktif

Dari tabel diatas diketahui untuk node 1.2 pecahan dari atribut deposit - minim yang memiliki nilai gain tertinggi adalah atribut frekuensi sehingga cabang berikutnya dipecah berdasarkan atribut frekuensi. Dimana akan terbentuk kode 1.2.1 untuk atribut frekuensi-jarang, 1.2.2 untuk atribut frekuensi-sedang dan node 1.2.3 untuk atribut frekuensi-sering.

Node 1.2.2 dimana berisi atribut frekuensi jarang akan menjadi class tidak aktif, demikian pula pada node 1.2.3 akan menjadi class aktif, dilihat dari nilai gain information nya. Sedangkan untuk node 1.2.2 akan dipecah kembali, setelah dilakukan perhitungan maka mode 1.2.2 akan dipecah berdasarkan atribut desa, pinggirKota dan kota.

Node 1.1 akan dipecah berdasarkan atribut deposit -cukup dan frekuensi-jarang, berikut hasil perhitungannya:

Tabel III.9. Hasil perhitungan node 1.1.1

Node			Jumlah kasus (S)	Aktif (S1)	Tidak (S2)	Entropy	Gain	Ket
1.1.1.1	Lama Berlangganan	Baru	4	3	1	0.81128		
		Desa	2	1	1	1		
		PinggirKota	1	1	0	0		

	Area	Kota	1	1	0	0	0.3113	Aktif
1.1.1.2	Lama Berlangganan	Lama	4	1	3	0.81128		
		Desa	1	0	1	0	0.8113	Tidak
		PinggirKota	2	1	1	1		
	Area	Kota	1	0	1	0		

Dari perhitungan diatas maka atribut frekuensi-jarang akan dipecah kembali berdasarkan atribut Lama Berlangganan dikarenakan memiliki nilai gain tertinggi. Node 1.1.1 akan dipecah menjadi node 1.1.1.1 yang mewakili atribut area -desa, node 1.1.1.2 mewakili atribut area-pinggir kota dan node 1.1.1.3 mewakili atribut area - kota. Ketiga node tersebut akan dihitung menggunakan perhitungan gain information. Hasil perhitungannya akan membentuk class aktif atau tidak aktif.

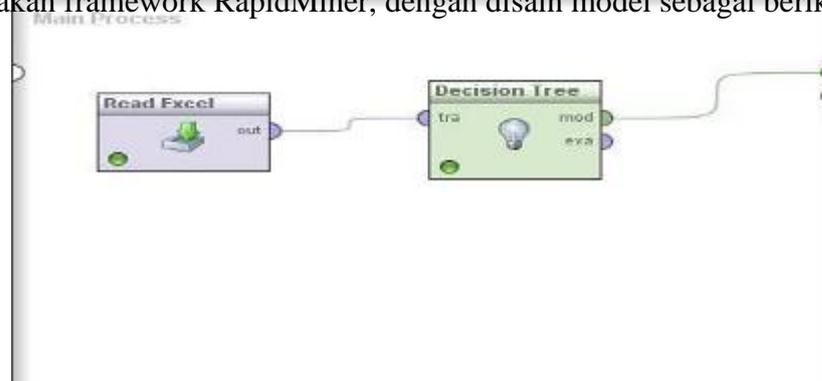
Dari perhitungan sebelumnya node 1.2 akan dipecah menjadi node 1.2.1 mewakili atribut deposit-minim dan frekuensi-sedang, berikut hasil perhitungannya:

Tabel III.10. Hasil perhitungan node 1.2.2

Node			Jumlah kasus (S)	Aktif (S1)	Tidak (S2)	Entropy	Gain	Ket
1.2.2.1	Area	Desa	2	1	1	1		
	Lama	Baru	1	1	0	0	0	Aktif
	Berlangganan	Lama	1	0	1	0		
1.2.2.2	Area	PinggirKota	3	2	1	0.9183		
	Lama	Baru	2	1	1	1	0.252	Aktif
	Berlangganan	Lama	1	1	0	0		
1.2.2.3	Area	Kota	2	0	2	0		
	Lama	Baru	1	0	1	0	0	Tidak
	Berlangganan	Lama	1	0	1	0		

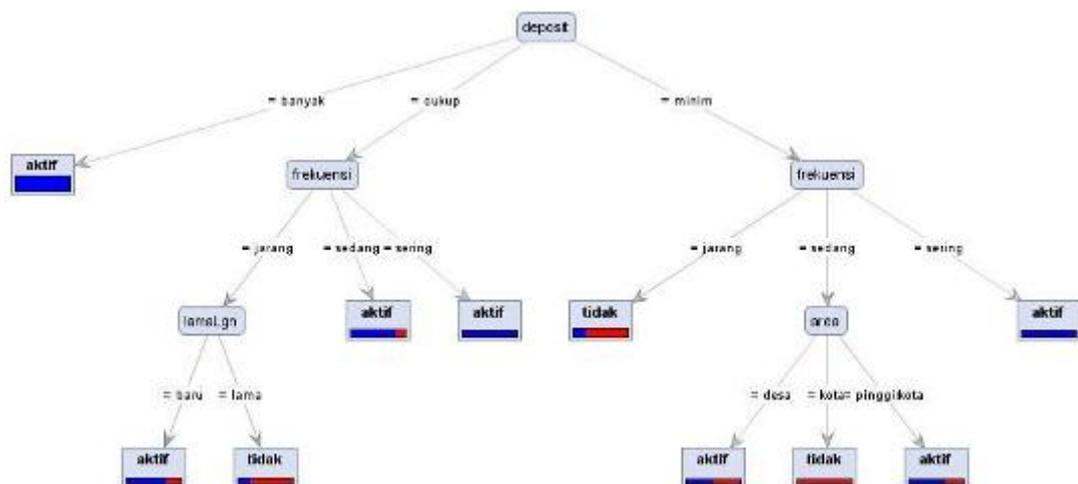
Dari tabel diatas diketahui untuk node 1.2.1 atribut deposit -minim untuk frekuensi-sedang yang memiliki nilai gain tertinggi adalah atribut area sehingga cabang berikutnya dipecah berdasarkan atribut area.

Perhitungan diatas dapat dibuat suatu model pohon keputusan dengan menggunakan framework RapidMiner, dengan disain model sebagai berikut:



Gambar III.2. Desain model decision tree menggunakan Rapid Miner

Dengan desain diatas menghasilkan sebuah pohon keputusan sebagai berikut:



Gambar III.3. Hasil klasifikasi menggunakan algoritma C4.5

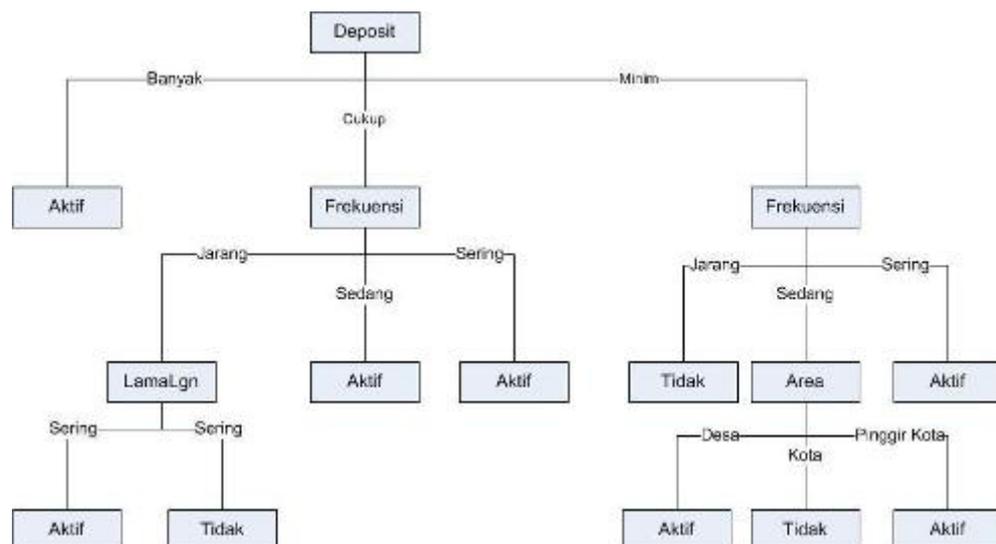
Gambar diatas merupakan hasil output dari prediksi loyalitas pelanggan dengan menggunakan model algoritma C4.5.

B. Penerapan Model AdaBoost

Model lain yang digunakan untuk meningkat akurasi dari pohon keputusan adalah AdaBoost. Berikut penggunaan algoritma AdaBoost dengan menggunakan decision tree, diketahui $n=59$ pengamatan, $y=4$ atribut (Deposit, Frekuensi, Lama Langgan, Area), $k=2$ class (aktif atau tidak), menggunakan algoritma Boosting dari sebanyak M iterasi.

- 1) Penentuan awal bobot $W_i = 1/n = 1/59 = 0.071$ untuk semua $i=1,2,3,\dots,n$
- 2) Penentuan iterasi dimana nomor iterasi $m = 1,2,\dots,M$, kemudian Lakukan proses berikut:
 - a. Susun pohon tunggal dengan n memperhatikan bobot W_i
 - b. Hitung tingkat kesalahan klasifikasi
 - c. Hitung m
 - d. Tentukan bobot baru untuk setiap pengamatan
- 3) Menentukan nilai akhir class.

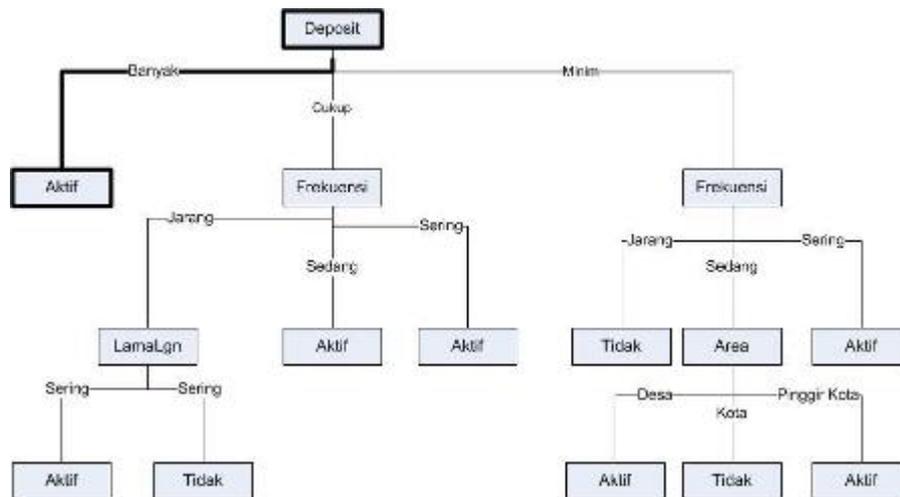
Berikut gambar susunan secara lengkap pohon tunggal yang terdiri dari beberapa hipotesa yang terbentuk dari perhitungan dengan menggunakan algoritma C4.5 berbasis AdaBoost:



Gambar III.4. Model pohon keputusan secara lengkap

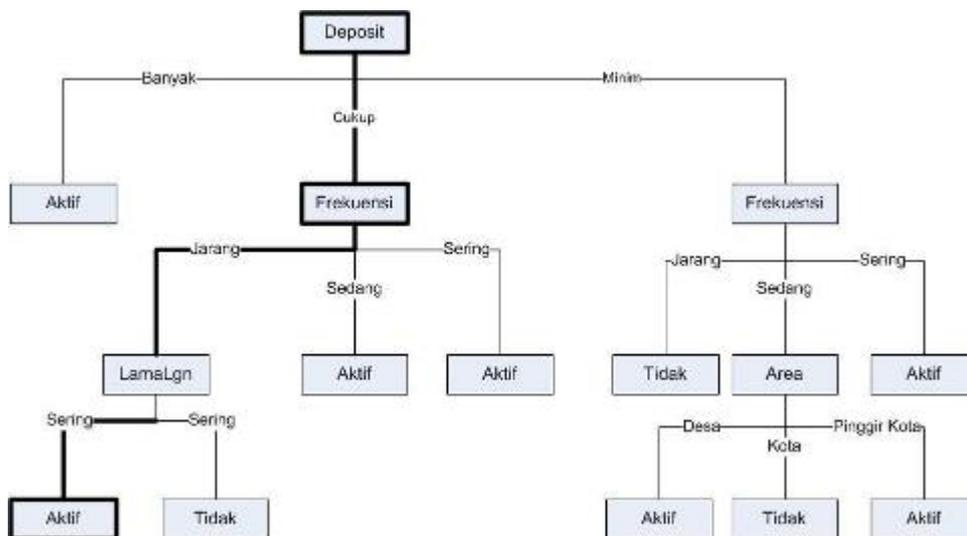
Pada gambar III.4. merupakan sebuah model pohon keputusan secara lengkap yang akan dibagi menjadi beberapa pohon keputusan tunggal berdasarkan class aktif atau tidak.

Perhitungan menggunakan model AdaBoost, setiap pohon tunggal diasumsikan sebagai sebuah iterasi, sehingga dari pohon keputusan secara lengkap diatas dapat menghasilkan 14 pohon tunggal sehingga akan terjadi 14 iterasi.



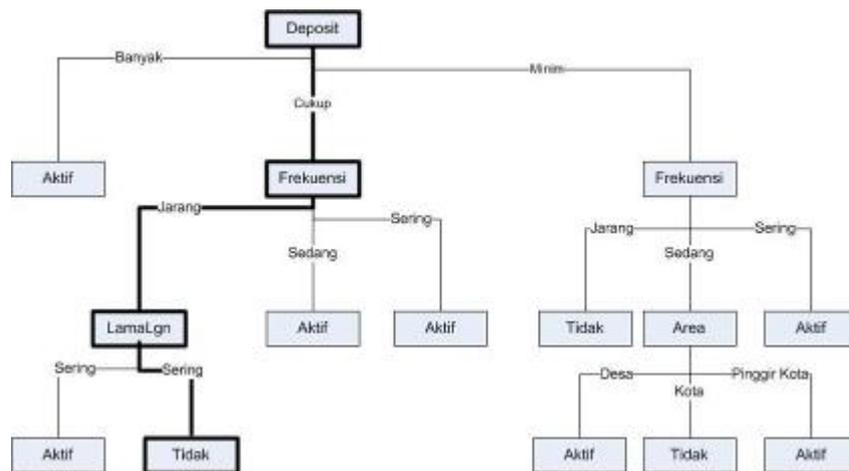
Gambar III.5. Pohon tunggal dari hipotesa 1

Pada gambar III.5. merupakan bentuk pohon tunggal dari hipotesa pertama yang mempunyai nilai class aktif.



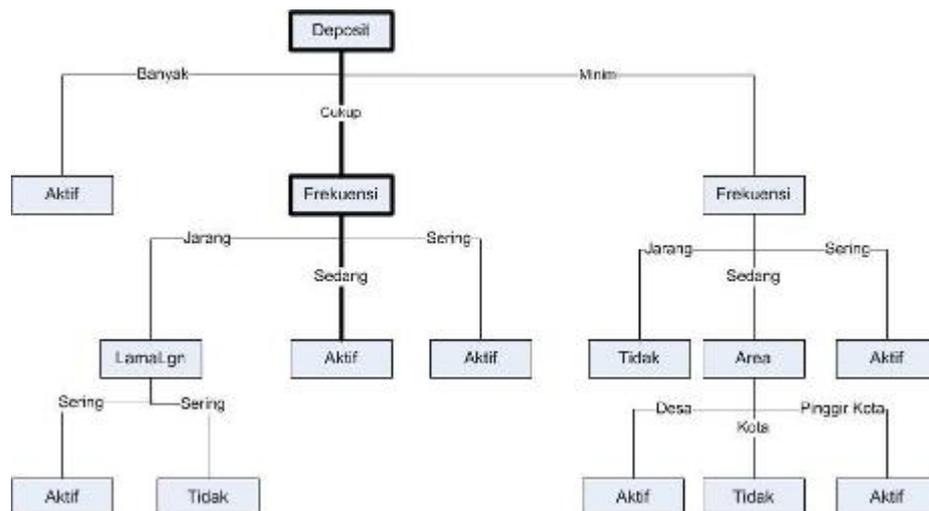
Gambar III.6. Pohon tunggal dari hipotesa 2

Pada gambar III.6. merupakan pohon tunggal hipotesa ke-2 yang mempunyai nilai class aktif.



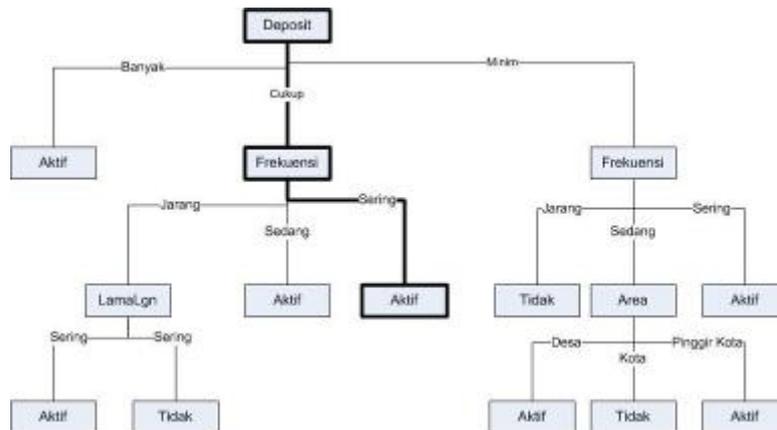
Gambar III.7. Pohon tunggal dari hipotesa 3

Pada gambar III.7 merupakan pohon tunggal hipotesa ke-3 yang mempunyai nilai class tidak.



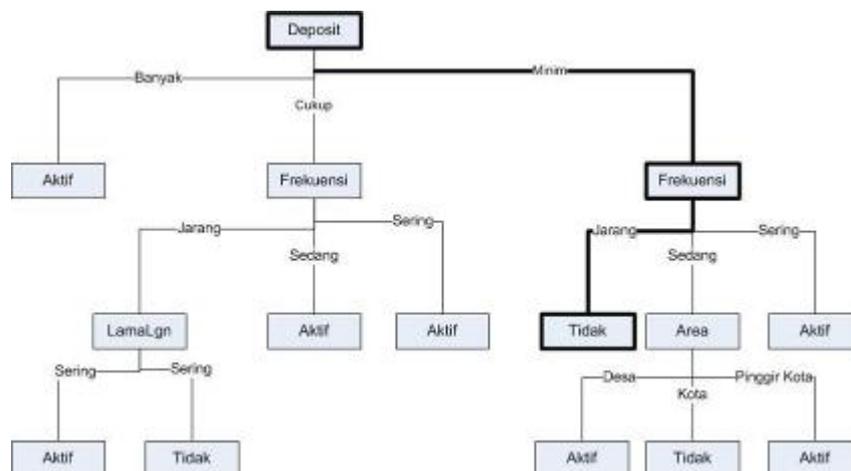
Gambar III.8. Pohon tunggal dari hipotesa 4

Pada gambar III.8. merupakan pohon tunggal hipetasa ke-4 yang mempunyai nilai class aktif.



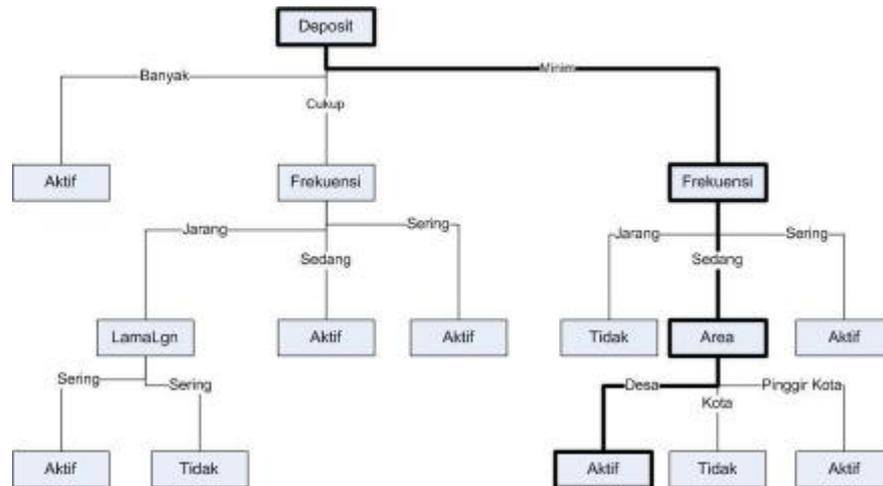
Gambar III.9. Pohon tunggal dari hipotesa 5

Pada gambar III.9. merupakan pohon tunggal hipetasa ke-5 yang mempunyai nilai class aktif.



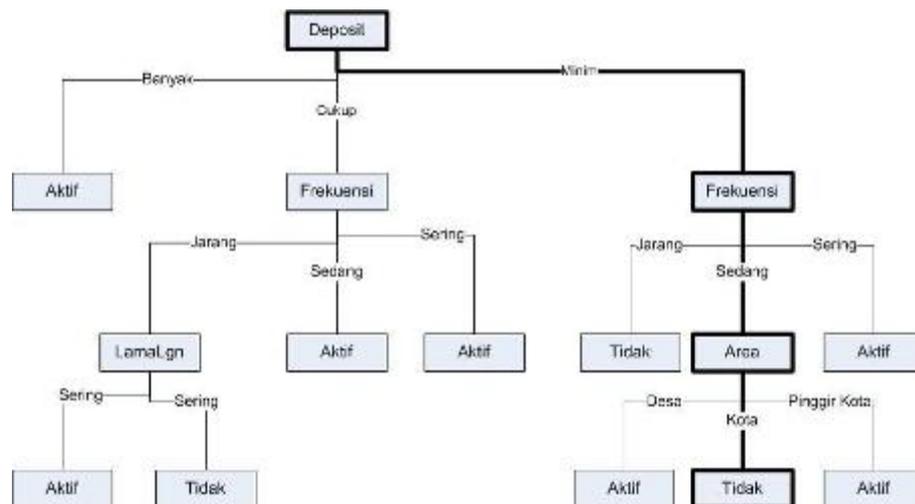
Gambar III.10. Pohon tunggal dari hipotesa 6

Pada gambar III.10. merupakan pohon tunggal hipetasa ke-6 yang mempunyai nilai class tidak.



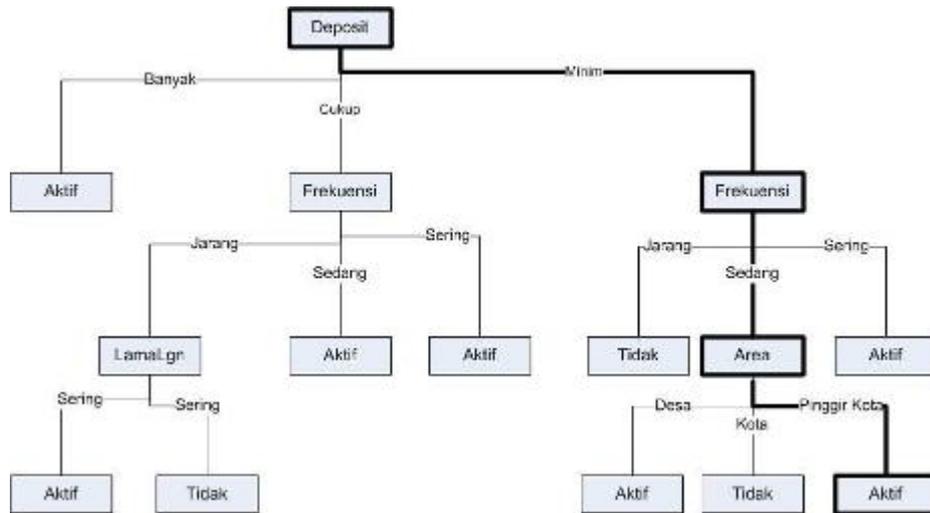
Gambar III.11 Pohon tunggal dari hipotesa 7

Pada gambar III.11. merupakan pohon tunggal hipotesa ke-7 yang mempunyai nilai class aktif.



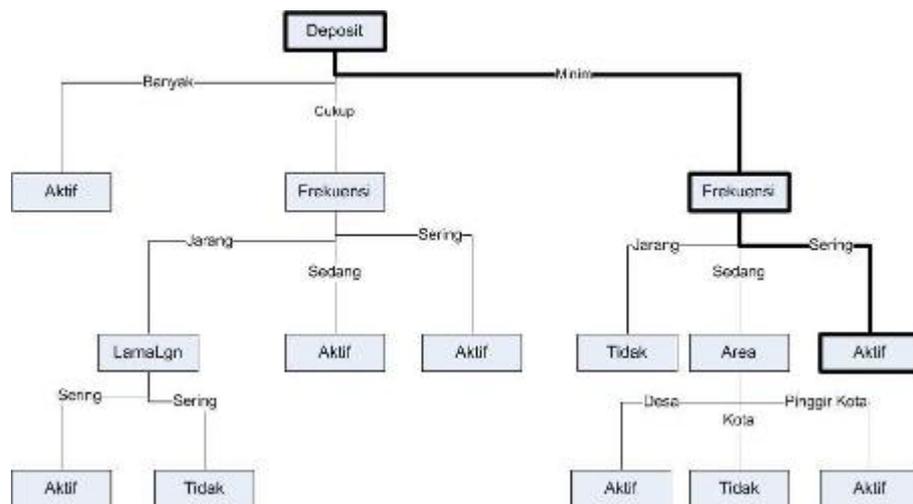
Gambar III.12. Pohon tunggal dari hipotesa 8

Pada gambar III.12. merupakan pohon tunggal hipotesa ke-8 yang mempunyai nilai class tidak.



Gambar III.13. Pohon tunggal dari hipotesa 9

Pada gambar III.13. merupakan pohon tunggal hipotesa ke-9 yang mempunyai nilai class aktif.



Gambar III.14. Pohon tunggal dari hipotesa 10

Pada gambar III.14 merupakan pohon tunggal hipotesa ke-10 yang mempunyai nilai class aktif.

Pohon tunggal di atas yang terdiri dari sepuluh hipotesa pohon keputusan akan dihitung besar bobot dari masing-masing hipotesa, berikut ini perhitungan untuk iterasi pertama yaitu:

Tabel III.11. Bobot nilai iterasi ke-1 menggunakan C4.5 berbasis AdaBoost

Hipotesa	Error (m)	m	Wi	hasil hipotesa
1	0	0	1	0
2	0.25	2.1972246	0.111111	0.244136064
3	0.25	2.1972246	0.111111	0.244136064
4	0.16	3.3164562	0.03628	0.12032494
5	0	0	1	0
6	0.25	2.1972246	0.111111	0.244136064
7	0.5	0	1	0
8	0	0	1	0
9	0.33	1.4163701	0.24259	0.343601483
10	0	0	1	0

Pada iterasi yang ke-1, hipotesa yang mempunyai nilai kesalahan=0 maka pada iterasi berikutnya akan bernilai tetap sedangkan nilai bobot yang baru didapat dari $W_i = W_i * m$. Sehingga untuk iterasi ke-2 didapat hasil perhitungan sebagai berikut:

Tabel III.12. Bobot nilai iterasi ke-2 menggunakan C4.5 berbasis AdaBoost

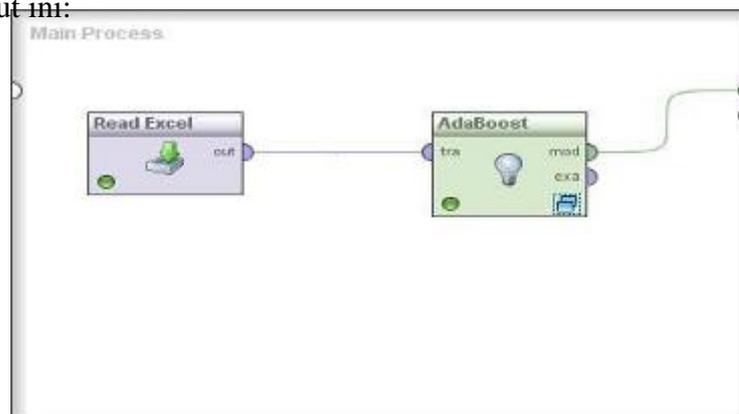
Hipotesa	Error (m)	m	Wi	hasil hipotesa
1	0	0	1	0
2	0.2441361	2.260271	0.104322	0.23579642
3	0.2441361	2.260271	0.104322	0.23579642
4	0.1203249	3.978713	0.01871	0.074444052
5	0	0	1	0
6	0.2441361	2.260271	0.104322	0.23579642
7	0	0	1	0
8	0	0	1	0
9	0.3436015	1.294571	0.274015	0.354732364
10	0	0	1	0

Semakin besar bobot sebuah pohon tunggal maka semakin kecil kesalahan (error) dari pohon tunggal. Pada iterasi berikutnya akan dihasilkan klasifikasi yang lebih baik. Iterasi akan berhenti apabila total nilai m bernilai 1 atau -1. Berikut perhitungannya:

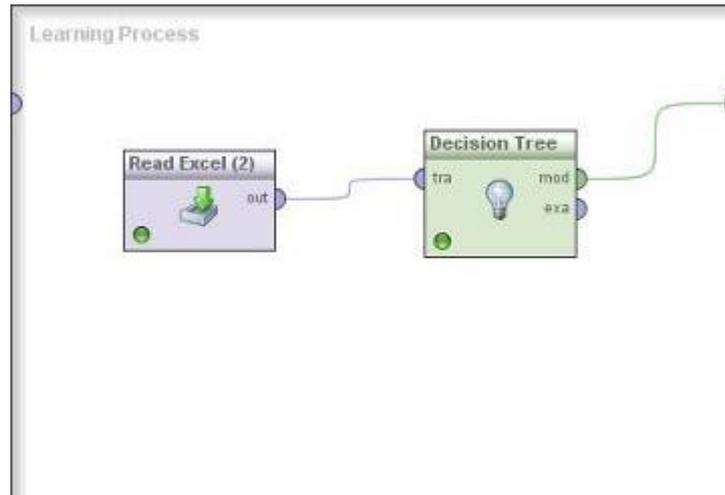
Tabel III.13. Perhitungan nilai class hipotesa 1 dan hipotesa 2

Hipotesa	m (iterasi 1)	m (iterasi 2)	sign (m1+ m2)
2	2.197224577	2.260271342	1
3	2.197224577	2.260271342	1
4	3.316456153	3.978713349	1
6	2.197224577	2.260271342	1
7	0	2.260271342	1
9	1.416370116	1.29457119	1

Tujuan utama penggunaan AdaBoost pada algoritma C4.5 adalah untuk meningkatkan akurasi prediksi loyalitas pelanggan. Dengan menggunakan framework RapidMiner di desain model algoritma C4.5 berbasis AdaBoost pada gambar berikut ini:

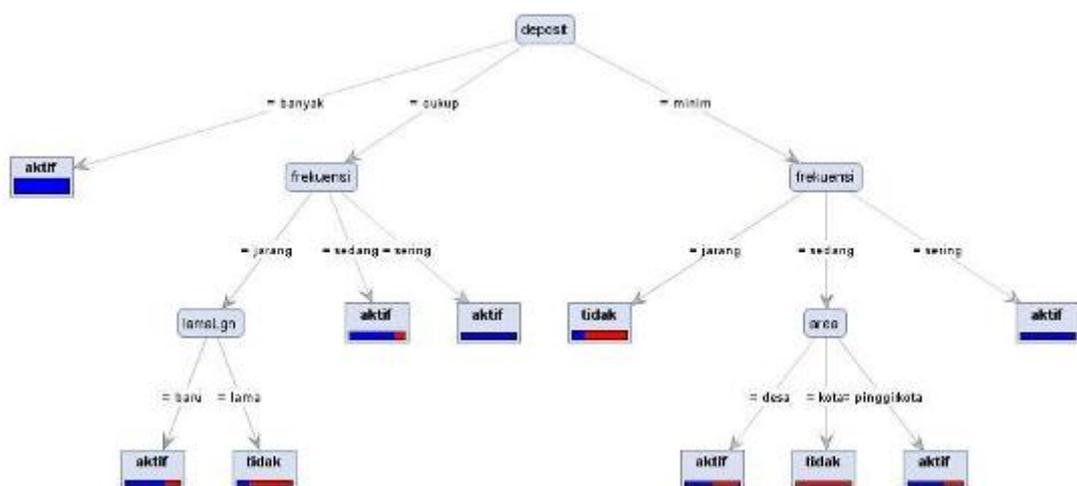


Gambar III.16. Desain model AdaBoost



Gambar III.17. Desain Model algoritma C4.5 dalam Model AdaBoost

Pada gambar diatas merupakan bentuk desain model untuk algoritma C45 berbasis AdaBoost menggunakan Rapid Miner.



Gambar III.18. Hasil klasifikasi menggunakan algoritma C4.5 berbasis AdaBoost

Pada gambar diatas merupakan output dari hasil klasifikasi dengan menggunakan model C4.5 berbasis AdaBoost.

Dari gambar hasil klasifikasi algoritma C4.5 dan algoritma C4.5 berbasis AdaBoost mempunyai kemiripan hasil pohon keputusan.

3.2. Jadwal Penelitian

Jadwal penelitian dalam penyusunan tesis ini tercantum dalam tabel III.14.berikut :

Tabel III.14. Jadwal Penelitian

No	Kegiatan	April	Mei	Juni	Juli	Agust
1	Pencarian dan Pemilihan Obyek Penelitian	■				
2	Pengambilan data		■	■		
3	Persiapan data			■	■	
4	Pembuatan model				■	■
5	Pengujian model				■	■
6	Pembuatan laporan tesis				■	■

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1. Pengukuran Penelitian

Pengukuran penelitian ini membahas tentang hasil penelitian dan pengujian model yang dipakai dalam penelitian sehingga hasilnya sesuai penelitian yang dilakukan.

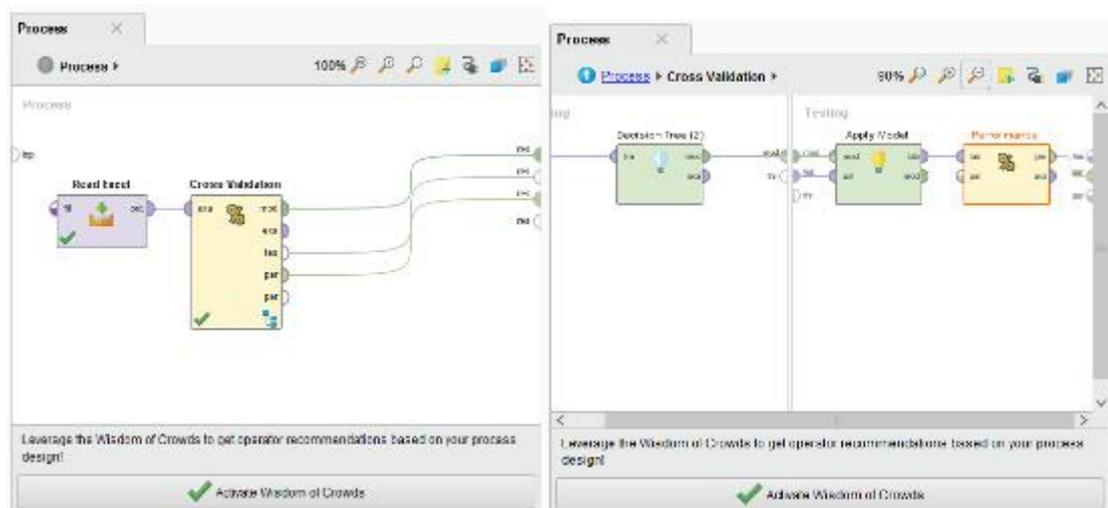
4.1.1. Hasil Penelitian

Hasil dari penelitian ini menguji keakuratan prediksi loyalitas pelanggan dengan menggunakan algoritma C4.5 berbasis AdaBoost. Data yang dianalisa adalah data pelanggan dari server distributor pulsa elektronik baik data pelanggan yang masih aktif atau yang sudah tidak aktif lagi melakukan transaksi.

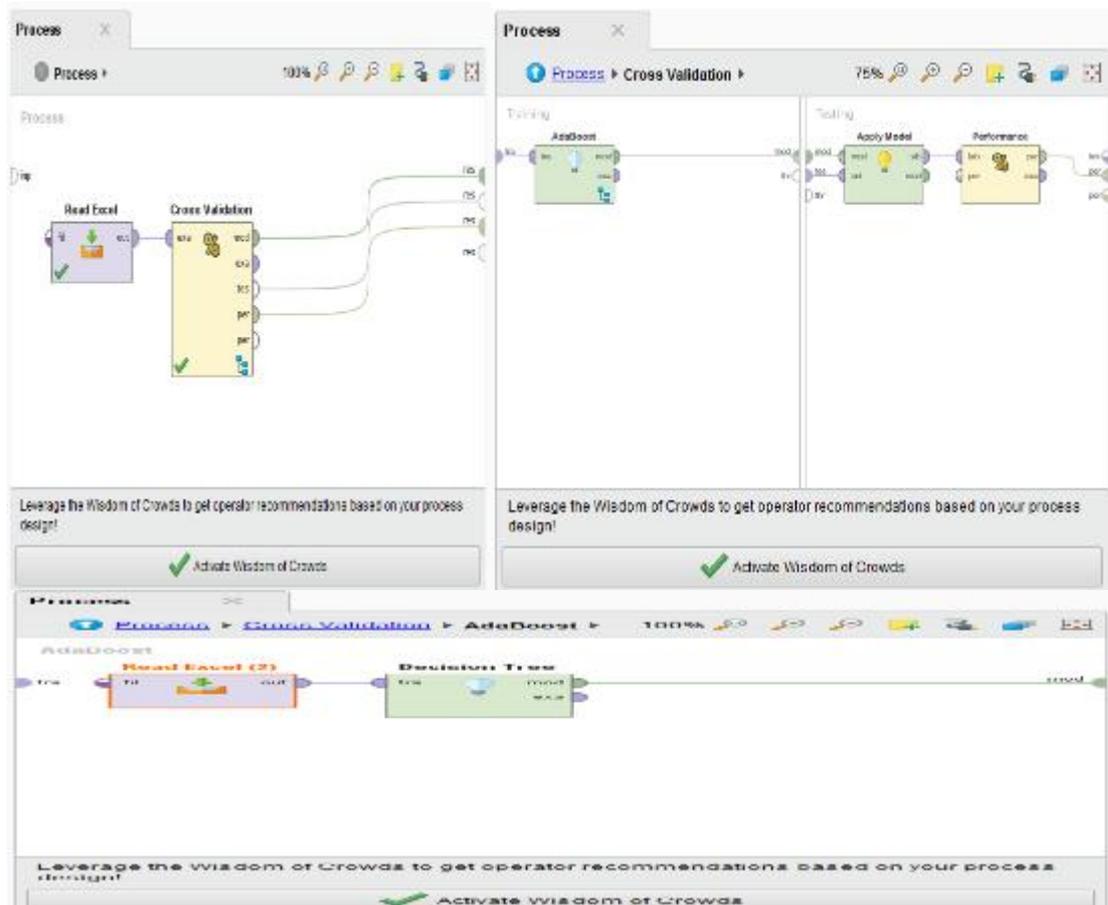
4.1.2. Pengujian Model

Pengujian model dihasilkan dari data eksperimen akan diujikan dengan menggunakan metode *10-fold cross-validation*, dimana data secara acak (*random*) akan dibagi menjadi 10 bagian. Pembagian menjadi 10 bagian merupakan metode yang paling tepat untuk mendapatkan estimasi terbaik menentukan kesalahan. Setiap bagian akan dihitung tingkat kesalahan setelah itu secara keseluruhan akan di hitung rata-rata nya (Witten, et al, 2010).

Setelah dilakukan klasifikasi model data, maka tahap selanjutnya melakukan pengujian data untuk memprediksi akurasi data uji.



Gambar IV.1. Pengujian C4.5



Gambar IV.2. Pengujian C4.5 berbasis Adaboost

Metode yang digunakan untuk menganalisa model klasifikasi yaitu:

1. Confusion Matrix

Confusion matrix merupakan metode yang digunakan mengevaluasi prediksi benar atau tidak data uji berdasarkan model klasifikasi (Gorunescu, 2011). Evaluasi model ini menggunakan tabel berikut ini:

Tabel IV.1. Model *Confusion Matrix*

<i>Classification</i>	<i>Predicted Class</i>		
		<i>Class=Yes</i>	<i>Class=No</i>
<i>Observed Class</i>	<i>Class=Yes</i>	<i>True Positives</i>	<i>False Negatives</i>
	<i>Class=No</i>	<i>False Positives</i>	<i>True Negatives</i>

Perhitungan dibuat dalam bentuk tabel dimana sisi kiri tabel menunjukkan *class* yang diamati, sedangkan sisi atas menunjukkan *class* yang diprediksi. Setiap sel dari tabel menunjukkan banyaknya data yang digunakan untuk diprediksi.

True positives merupakan tupel positif di data set yang diklasifikasikan positif. *True negatives* merupakan tupel negatif di data set yang diklasifikasikan negatif. *False positives* adalah tupel positif di data set yang diklasifikasikan negatif sedangkan *false negatives* merupakan jumlah tupel negatif yang diklasifikasikan positif.

Model yang telah dikembangkan akan diuji keakuratannya dengan memasukkan sejumlah data uji (*test set*) ke dalam model. Data uji diambil dari data pelanggan pada bulan juni, juli, agustus 2017. Berikut hasil pengujiannya:

Tabel IV.2. Hasil Pengujian dengan C4.5 dan C4.5 berbasis AdaBoost

deposit	Frekuensi	Area	lamaLgn	Class	C4.5	C4.5 adaboost
banyak	Sedang	Desa	baru	Aktif	aktif	aktif
banyak	Sering	Desa	baru	Aktif	aktif	aktif
banyak	Sering	Kota	baru	Aktif	aktif	aktif
banyak	Sedang	Kota	baru	Aktif	aktif	aktif
banyak	Jarang	Kota	baru	Aktif	aktif	aktif
banyak	Sering	Desa	lama	Aktif	aktif	aktif
banyak	Sedang	Desa	lama	Aktif	aktif	aktif

banyak	Sedang	Kota	lama	Aktif	aktif	aktif
banyak	Jarang	Kota	lama	Aktif	aktif	aktif
banyak	Sering	Kota	lama	Aktif	aktif	aktif
banyak	Jarang	Desa	baru	Aktif	aktif	aktif
banyak	Sedang	Pinggirkota	lama	Aktif	aktif	aktif
banyak	Sering	Pinggirkota	lama	Aktif	aktif	aktif
banyak	Jarang	Pinggirkota	lama	Aktif	aktif	aktif
banyak	Jarang	Pinggirkota	baru	Aktif	aktif	aktif
banyak	Sedang	Pinggirkota	baru	Aktif	aktif	aktif
banyak	Sering	Pinggirkota	baru	Aktif	aktif	aktif
banyak	Sering	Desa	baru	Aktif	aktif	aktif
banyak	Sering	Desa	lama	Aktif	aktif	aktif
cukup	Sedang	Desa	baru	Aktif	aktif	aktif
cukup	Jarang	Desa	baru	Aktif	aktif	aktif
cukup	Sedang	Kota	baru	Aktif	aktif	aktif
cukup	Sering	Kota	baru	Aktif	aktif	aktif
cukup	Jarang	Kota	baru	Aktif	aktif	aktif
cukup	Sering	Desa	lama	Aktif	aktif	aktif
cukup	Sering	Kota	lama	Aktif	aktif	aktif
cukup	Sedang	Desa	lama	Aktif	aktif	aktif
cukup	Sedang	Pinggirkota	lama	Aktif	aktif	aktif
cukup	Sedang	Pinggirkota	baru	Aktif	aktif	aktif
cukup	Jarang	Pinggirkota	baru	Aktif	aktif	aktif
cukup	Jarang	Pinggirkota	lama	Aktif	tidak	tidak
cukup	Sering	Pinggirkota	baru	Aktif	aktif	aktif
cukup	Jarang	Kota	lama	Tidak	tidak	tidak
cukup	Jarang	Desa	lama	Tidak	tidak	tidak
cukup	Jarang	Desa	baru	Tidak	aktif	aktif
cukup	Sedang	Kota	lama	Tidak	aktif	aktif
cukup	Jarang	Pinggirkota	lama	Tidak	tidak	tidak
cukup	Sering	Desa	lama	Aktif	aktif	aktif
minim	Sering	Kota	baru	Aktif	aktif	aktif
minim	Sering	Kota	lama	Aktif	aktif	aktif
minim	Jarang	Desa	baru	Aktif	aktif	tidak
minim	Sedang	Pinggirkota	lama	Aktif	aktif	aktif
minim	Sedang	Pinggirkota	baru	Aktif	aktif	aktif
minim	Sering	Pinggirkota	lama	Aktif	aktif	aktif
minim	Sering	Pinggirkota	baru	Aktif	aktif	aktif
minim	Jarang	Pinggirkota	baru	Aktif	aktif	tidak
minim	Jarang	Desa	baru	Tidak	tidak	tidak
minim	Sedang	Kota	baru	Tidak	tidak	tidak

minim	Jarang	Kota	baru	Tidak	tidak	tidak
minim	Jarang	Desa	lama	Tidak	tidak	tidak
minim	Sedang	Desa	lama	Tidak	aktif	aktif
minim	Jarang	Kota	lama	Tidak	tidak	tidak
minim	Sedang	Kota	lama	Tidak	tidak	tidak
minim	Jarang	Pinggirkota	lama	Tidak	tidak	tidak
minim	Jarang	Pinggirkota	baru	Tidak	tidak	tidak
minim	Sedang	Pinggirkota	baru	Tidak	aktif	aktif
minim	Jarang	Desa	lama	Tidak	tidak	tidak
minim	Jarang	Kota	lama	Tidak	tidak	tidak
minim	Jarang	Kota	lama	Tidak	tidak	tidak

Kemudian masukkan data yang telah uji menggunakan algoritma C4.5 berbasis AdaBoost yang ada ke dalam model *confusion matrix* maka akan didapatkan hasil sebagai berikut:

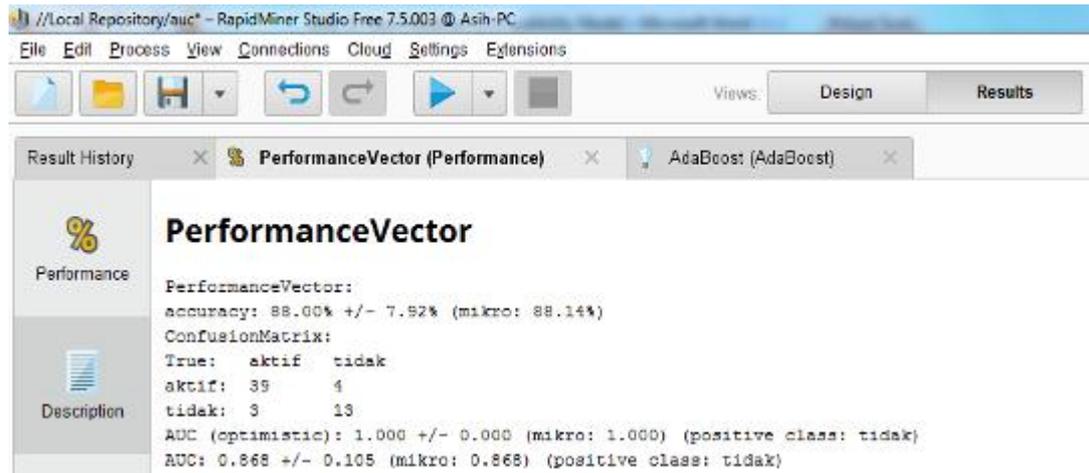
The screenshot shows the RapidMiner Studio interface. The main window displays the results for a **PerformanceVector** model. The results are as follows:

```

PerformanceVector:
accuracy: 69.67% +/- 17.73% (mikro: 69.49%)
ConfusionMatrix:
True:  aktif  tidak
aktif:  34    10
tidak:  8     7
AUC (optimistic): 0.815 +/- 0.162 (mikro: 0.815) (positive class: tidak)
AUC: 0.722 +/- 0.203 (mikro: 0.722) (positive class: tidak)

```

Gambar IV.3. Hasil *Confusion Matrix* menggunakan algoritma C4.5



Gambar IV.4. Hasil *Confusion Matrix* C4.5 berbasis Adaboost

Setelah data uji dimasukkan ke dalam *confusion matrix*, hitung nilai-nilai yang telah dimasukkan tersebut untuk menghitung jumlah *sensitivity*, *specificity*, *precision* dan *accuracy*. *Sensitivity* digunakan untuk membandingkan jumlah *true positives* terhadap jumlah tupel yang *positives* sedangkan *specificity* adalah perbandingan jumlah *true negatives* terhadap jumlah tupel yang *negatives*. Untuk menghitung digunakan persamaan di bawah ini:

$$\begin{aligned}
 \text{sensitivity} &= \frac{t_pos}{pos} \\
 \text{specificity} &= \frac{t_neg}{neg} \\
 \text{precision} &= \frac{t_pos}{t_pos + f_pos} \\
 \text{accuracy} &= \frac{pos}{(pos - neg)} \text{sensitivity} + \frac{neg}{(pos + neg)} \text{specificity}
 \end{aligned}$$

dimana:

t_pos = jumlah *true positives*

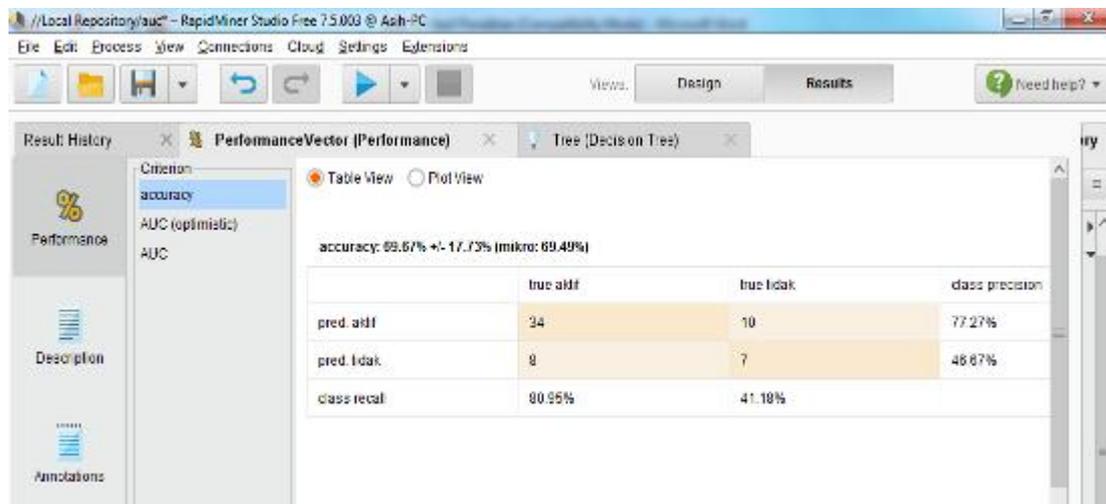
t_neg = jumlah *true negatives*

pos = jumlah tupel positif

neg = jumlah tupel negatif

f_pos = jumlah *false positives*

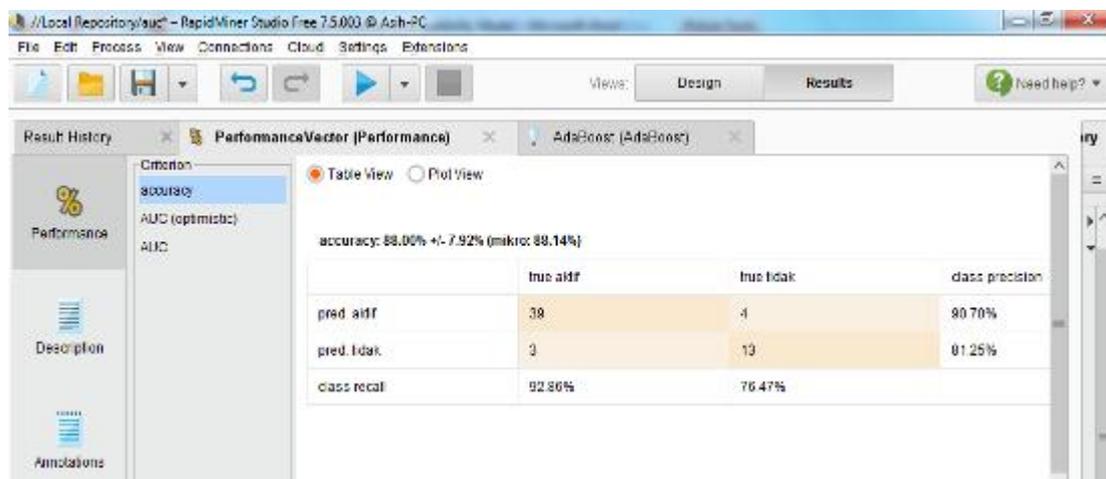
Kemudian masukkan nilai yang ada di dalam *confusion matrix* ke dalam persamaan di atas ke dalam algoritma C4.5, sehingga akan menghasilkan nilai seperti di bawah ini:



	true aidi	true tidak	class precision
pred. aidi	24	10	77.27%
pred. tidak	8	7	46.87%
class recall	80.95%	41.18%	

Gambar IV.5. *Confusion Matrix* menentukan *accuracy* dengan C4.5

Sedangkan untuk algoritma C4.5 berbasis AdaBoost akan menghasilkan nilai seperti di bawah ini:



	true aidi	true tidak	class precision
pred. aidi	38	4	90.70%
pred. tidak	3	13	81.25%
class recall	92.96%	76.47%	

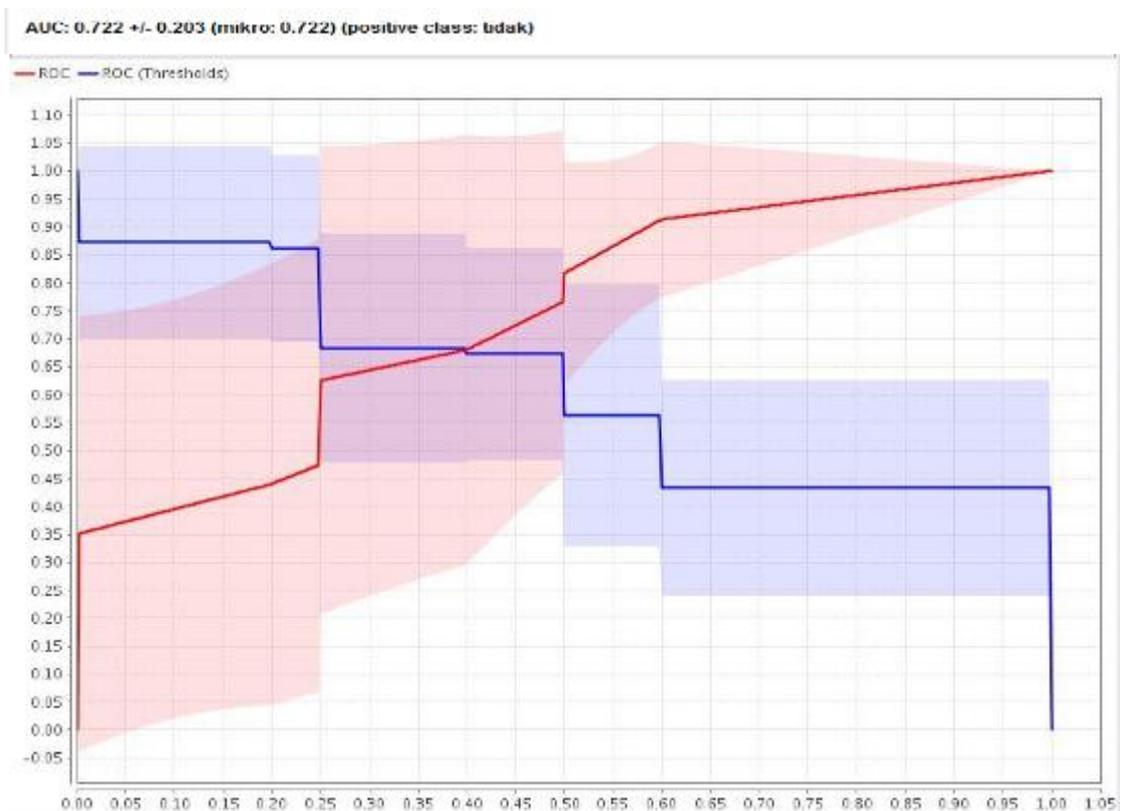
Gambar IV.6. *Confusion Matrix accuracy* dengan C4.5 berbasis AdaBoost

Hasil pengujian *confusion matrix* di atas diketahui menggunakan model algoritma C4.5 mempunyai akurasi 69.67% sedangkan model algoritma C4.5 berbasis Adaboost memiliki akurasi 88.00%. tingkat akurasi yang lebih tinggi dibandingkan algoritma C4.5 sebesar 18.33%.

2. Kurva ROC (Receiver Operating Characteristic)

ROC merupakan sebuah grafik yang digunakan untuk menilai hasil prediksi. Dalam model klasifikasi ROC merupakan teknik visualisasi, pengaturan dan pemilihan klasifikasi berdasarkan hasil *performance*. Kurva ROC merupakan *tools* untuk membandingkan model klasifikasi (Gorunescu, 2011).

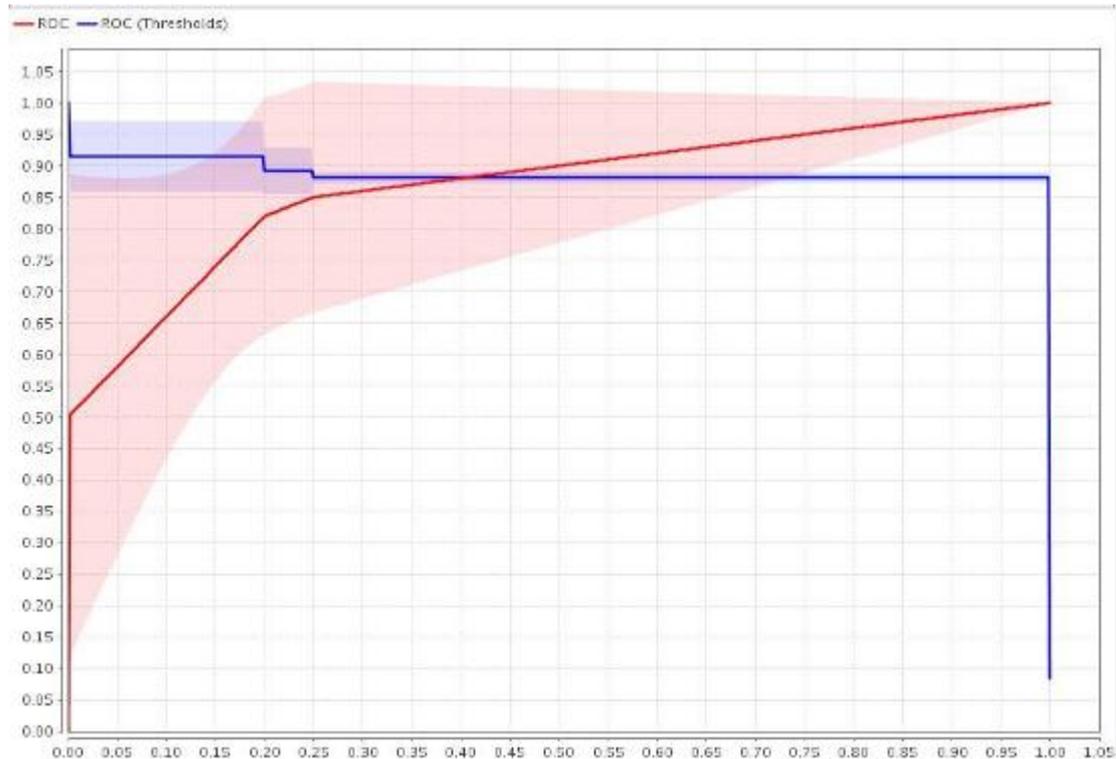
Data uji diatas akan dinilai hasil prediksi dengan menggunakan grafik ROC untuk algoritma C4.5, visualisasi dari grafik ROC yaitu:



Gambar IV.7. Grafik ROC dengan model algoritma C4.5

Sedangkan visualisasi grafik ROC dengan model algoritma C4.5 berbasis AdaBoost sebagai berikut:

AUC: 0.868 +/- 0.105 (mikro: 0.868) (positive class: tidak)



Gambar IV.8. Grafik ROC dengan model algoritma C4.5 berbasis AdaBoost

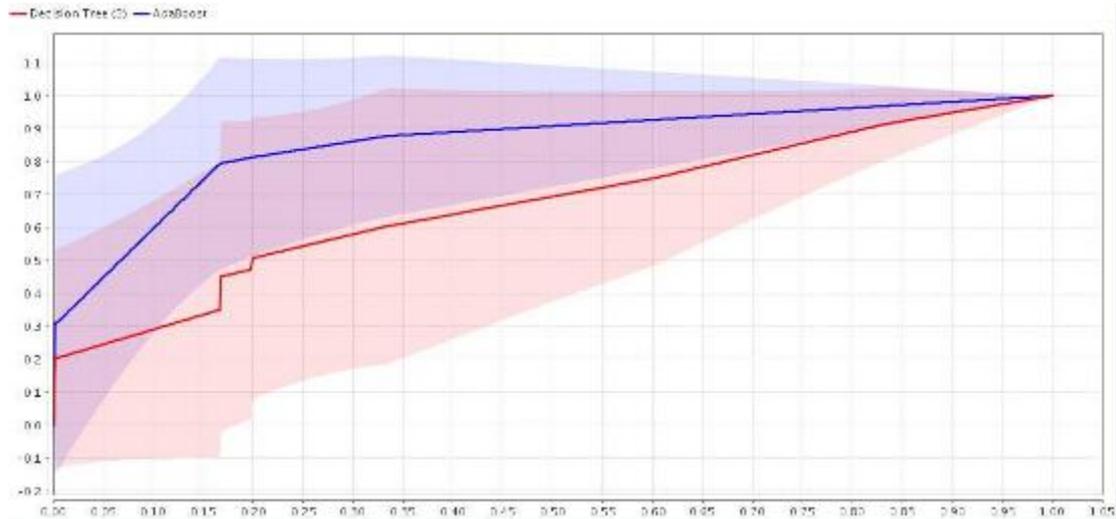
Grafik ROC akan membentuk garis dimana garis tersebut menunjukkan hasil prediksi dari model klasifikasi yang digunakan. Apabila garis tersebut berada di atas diagonal grafik maka hasil klasifikasi bernilai baik (*good classification*), sedangkan garis yang berada di bawah diagonal grafik menghasilkan nilai klasifikasi yang buruk (*poor classification*). Garis yang menempel pada sumbu Y menunjukkan grafik tersebut menunjukkan klasifikasi yang baik (Gorunescu, 2011).

Dari grafik ROC didapatkan pula nilai AUC (*Area Under the ROC Curve*) untuk menganalisa hasil prediksi klasifikasi. Penentuan hasil prediksi klasifikasi dilihat dari batasan nilai AUC sebagai berikut (Gorunescu, 2011):

- a. Nilai AUC 0.90-1.00 = *excellent classification*
- b. Nilai AUC 0.80-0.90 = *good classification*
- c. Nilai AUC 0.70-0.80 = *fair classification*
- d. Nilai AUC 0.60-0.70 = *poor classification*

e. Nilai AUC 0.50-0.60 = *failure*

Kedua grafik ROC tersebut akan dibandingkan untuk menentukan model mana yang mempunyai nilai prediksi yang paling baik, berikut grafik perbandingannya:



Gambar IV.9. Komparasi ROC C4.5 dan C4.5 berbasis AdaBoost

Grafik diatas menunjukkan bahwa algoritma C4.5 berbasis AdaBoost memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan algoritma C4.5. Dengan demikian algoritma C4.5 berbasis AdaBoost akurat untuk memprediksi loyalitas pelanggan.

4.2. Analisa Hasil Komparasi

Hasil evaluasi dengan menggunakan *confusion matrix* dan grafik ROC untuk model klasifikasi algoritma C4.5 dan C4.5 berbasis AdaBoost sebagai berikut:

Tabel IV.3. Komparasi evaluasi C4.5 dan C4.5 berbasis AdaBoost

	Algoritma C4.5	C4.5 berbasis AdaBoost
Accuracy	69.67%	88.00%
AUC	0.722	0.868

Dari hasil evaluasi di atas diketahui bahwa metode Boosting mampu meningkatkan tingkat akurasi prediksi loyalitas pelanggan. Hasil dari grafik AUC menunjukkan bahwa model C4.5 berbasis AdaBoost menunjukkan model tersebut mempunyai *good classification*. Dengan demikian algoritma C4.5 berbasis AdaBoost mampu memberikan pola prediksi dengan akurat.

4.3. Implikasi Penelitian

Dari hasil penelitian implikasi kebijakan yang dirujuk kepada manajerial perusahaan antara lain:

1. Memberikan bonus atau potongan harga bagi pelanggan yang mempunyai deposit dalam jumlah banyak atau frekuensi aktifitas pengisian pulsa elektronik yang tinggi.
2. Memberikan harga kompetitif bagi pelanggan yang tinggal di perkotaan dimana tingkat persaingannya lebih tinggi.
3. Memberikan pelayanan yang lebih cepat dan baik terutama kepada pelanggan lama.

BAB V

PENUTUP

5.1. Kesimpulan

Penelitian ini dibuat untuk memprediksi pola loyalitas pelanggan pada distributor pulsa elektronik dengan tingkat akurasi yang tinggi. Untuk memprediksi permasalahan tersebut digunakan model klasifikasi yaitu algoritma C4.5 berbasis AdaBoost dimana hasil evaluasi model tersebut akan dilihat nilai akurasinya.

Kesimpulan dari penelitian ini adalah:

1. Algoritma C4.5 berbasis AdaBoost meningkatkan akurasi dalam memprediksi pola dan perilaku loyalitas pelanggan sehingga perusahaan dapat memberikan pelayanan yang lebih baik kepada pelanggan.
2. Berdasarkan penelitian tingkat akurasi prediksi dengan model C4.5 berbasis AdaBoost sebesar 88.00% sedangkan algoritma C4.5 sebesar 69.67%. Hal ini membuktikan bahwa hasil evaluasi penelitian algoritma C4.5 berbasis AdaBoost mampu memberikan perbaikan tingkat kesalahan klasifikasi sebesar 18.33%.

5.2. Saran

Meskipun penggunaan algoritma C4.5 berbasis AdaBoost sudah diterapkan dan dapat meningkatkan tingkat akurasi, ada beberapa hal yang harus ditambahkan untuk meningkatkan akurasi algoritma C4.5 yaitu: Pada penelitian selanjutnya akan dikembangkan dalam menguji ulang model penelitian dengan menambah atribut yang lain seperti data keluhan pembayaran, data kepuasan pelanggan dan dapat dilakukan pengujian dengan model lain.

DAFTAR PUSTAKA

- Aldi Nurzahputra, Much Aziz Muslim. 2017. Peningkatan Akurasi Pada Algoritma C4.5 Menggunakan Adaboost Untuk Meminimalkan Resiko Kredit. Prosiding SNATI F Ke - 4 Tahun 2017. ISBN: 978-602-1180-50-1.
- Atmosutarno. 2010. Jumlah Pelanggan Seluler Tembus 180 juta. <http://www.antaraneews.com/berita/12791080087/atsi-jumlah-pelanggan-seluler-tembus-180-juta/>.
- Freud, Yoav & Schapire, Robert. 1999. Short Introduction to Boosting. Journal of Japanese Society for Artificial Intelligence. 14(5):771-780.
- Freund, Y. and Schapire, R.E., 1995, March. A decision-theoretic generalization of on-line learning and an application to boosting. In *European conference on computational learning theory* (pp. 23-37). Springer Berlin Heidelberg.
- Goran, Kraljevic & Gotovac, Sven. 2010. Modelling Data Mining Applications for Prediction of Prepaid Churn in Telecommunication Services. ATKAFF 51(3), 275-283. ISSN 0005-1144.
- Gorunescu. 2011. Data Mining Concepts, Models and Techniques. Romania:Springer-Verlag Berlin Heidelberg.
- Govindaraju,R & Simatupang, T & Samadhi, Ari. 2008. Perancangan Sistem Prediksi Churn Pelanggan PT. Telekomunikasi Seluler Dengan Memanfaatkan Proses Data Mining. Jurnal Informatika Vol 9 No 1, 33-42
- Griffin, Jill. 2005. Customer Loyalty. Jakarta: Erlangga
- Han, J & Kamber, M. 2006. Data Mining Concept and Techniques. India: New Age International Limited
- Hehui Qian & Zhiwei Qiu. 2014. Feature Selection Using C4.5 Algorithm For Electricity Price Prediction. Proceedings of the 2014 International Conference on Machine Learning and Cybernetics, Lanzhou, 13-16 July, 2014.
- Jahromi, A. Tamaddoni & Moesini, Mehrad & Akbari, Issrar. A Dual-Step Multi-Algorithm Approach For Churn Prediction In Pre-Paid Telecommunications Service Providers. Journal of Innovation and Sustainability. ISSN:2179-3565

- Jin-hui, Lei & Jian-jun, He. 2009. The Comparative analysis and Study of Mobile-based Customer Data Churn Prediction Model. World Congress on Software Engineering. IEEE:978-0-7695-3570-8/09
- Kothari, C. R. (2004). *Research Methology Methods and Techniques*. India: New Age International Limited.
- Liao T, Warren. 2007. Recent Advances in Data Mining of Enterprise Data Algorithms and Applications. Lousiana: World Scientific
- Liu, H., Tian, H.Q., Li, Y.F. and Zhang, L., 2015. Comparison of four Adaboost algorithm based artificial neural networks in wind speed predictions. *Energy Conversion and Management*, 92, pp.67-81.
- Lokesh S. Katore. 2015. Novel Professional Career Prediction And Recommendation Method For Individual Through Analytics On Personal Traits Using C4.5 Algorithm. Proceeding 2015 of Global Conference on Comunnication Technologies(GCCT 2015)
- Moore, Patricia. 2005. *Menguasai CRM*. Jakarta: Prestasi Pustaka Publisher
- Oseman, Khalida. 2010. Data Mining in Churn Analysis Model for Telecommunication Industry. *Journal of Statistical and Analytics Vol 1 No 19-27*
- Rahmayanty, Nina. 2010. *Manajemen Pelayanan Prima*. Yogyakarta: Graha Ilmu
- Xue Bao & Xin Guan. 2016. A Method of Predicting Crude Oil Output Based on RS-C4.5 Algorithm IWEI QIU. International Conference on Information Science and Control Engineering.

Lampiran 1. Data Real Pelanggan Keseluruhan

No	Area	Deposit	LamaLgn	Frekuensi
1	Cilampuyang	238,400	5/10/2017	50 kali
2	Banyumas	187,800	15/11/2017	250 kali
3	Purbalingga	293,485	3/9/2017	200 kali
4	Baturaden	304,160	12/9/2017	140 kali
5	Purwokerto	324,460	12/9/2017	9 kali
6	Bukateja	96,384	3/9/2017	130 kali
7	Cimanggu	82,159	4/10/2017	3 kali
8	Kalimantan	91,299	21/9/2017	150 kali
9	Baturaden	100,000	26/9/2017	250 kali
10	Ajibarang	97,070	22/9/2017	10 kali
11	Jakarta	5,150	27/9/2017	120 kali
12	Purbalingga	7,447	29/9/2017	340 kali
13	Karangpucung	142,959	5/7/2016	330 kali
14	Ciporos	1,289,920	5/7/2016	123 kali
15	Purbalingga	148,975	5/7/2016	115 kali
16	Purbalingga	264,227	5/7/2016	5 kali
17	Kedung benda	327,870	5/7/2016	334 kali
18	Kedung reja	57,447	5/7/2016	760 kali
19	Cimanggu	59,575	5/7/2016	220 kali
20	Papua	7,565	5/7/2016	340 kali
21	Sukabumi	5,890	5/10/2017	4 kali
22	Tegal	72,786,565	8/10/2017	3 kali
23	Baturaden	100,020	5/7/2016	115 kali
24	Cilacap	111,900	5/7/2016	145 kali

25	Majenang	148,759	5/7/2016	340 kali
26	Tegal	271,200	5/7/2016	10 kali
27	Bukateja	350,825	9/10/2017	3 kali
28	Purbalingga	274,827	11/10/2017	115 kali
29	Bukateja	185,427	17/10/2017	445 kali
30	Purbalingga	96,102	5/7/2016	140 kali
31	Purbalingga	61,644	17/10/2017	130 kali
32	Cimanggu	59,459	10/10/2017	3 kali
33	Karangpucung	70,159	5/7/2016	5 kali
34	Baturaden	96,402	10/10/2017	445 kali
35	Baturaden	9,969	5/7/2016	140 kali
36	Purwokerto	3,576	17/10/2017	130 kali
37	Banyumas	5,671	5/7/2016	223 kali
38	Banyumas	1,800	17/10/2017	215 kali
39	Purwokerto	5,934	27/10/2017	5 kali
40	Cipawon	6,774	17/10/2017	4 kali
41	Purbalingga	7,175	7/10/2017	135 kali
42	Banjarnegara	3,920	3/10/2017	3 kali
43	Margahayu	30,905	5/7/2016	5 kali
44	Cipawon	6,630	5/7/2016	6 kali
45	Cikarang barat	8,549	5/7/2016	140 kali
46	Baturaden	3,272	5/7/2016	3 kali
47	Merden	80,775	5/7/2016	3 kali
48	Kalimantan	4,260	5/7/2016	115 kali
49	Baturaden	79,960	5/11/2017	5 kali

50	Tegal	97,077	5/7/2016	3 kali
51	Palembang	92,669	5/7/2016	7 kali
52	Purwosari	8,369	5/7/2016	2 kali
53	Cikarang	7,949	5/11/2017	9 kali
54	Cikarang	7,325	5/11/2017	145 kali
55	Purbalingga	277,000	5/7/2016	5 kali
56	Karangpucung	117,050	5/7/2016	3 kali
57	Cilacap	342,355	5/11/2017	10 kali
58	Banyumas	138,645	5/11/2017	115 kali
59	Purwokerto	78,550	5/7/2016	445 kali

Lampiran 2. Atribut dan Data Pelanggan Keseluruhan

No	Area	Deposit	LamaLgn	Frekuensi	Class
1	desa	banyak	baru	sedang	aktif
2	desa	banyak	baru	sering	aktif
3	kota	banyak	baru	sering	aktif
4	kota	banyak	baru	sedang	aktif
5	kota	banyak	baru	jarang	aktif
6	desa	cukup	baru	sedang	aktif
7	desa	cukup	baru	jarang	aktif
8	kota	cukup	baru	sedang	aktif
9	kota	cukup	baru	sering	aktif
10	kota	cukup	baru	jarang	aktif
11	desa	minim	baru	sedang	aktif
12	kota	minim	baru	sering	aktif
13	desa	banyak	lama	sering	aktif
14	desa	banyak	lama	sedang	aktif
15	kota	banyak	lama	sedang	aktif
16	kota	banyak	lama	jarang	aktif
17	kota	banyak	lama	sering	aktif
18	desa	cukup	lama	sering	aktif
19	kota	cukup	lama	sering	aktif
20	kota	minim	lama	sering	aktif
21	desa	minim	baru	jarang	aktif
22	desa	banyak	baru	jarang	aktif
23	desa	cukup	lama	sedang	aktif
24	pinggirkota	banyak	lama	sedang	aktif
25	pinggirkota	banyak	lama	sering	aktif
26	pinggirkota	banyak	lama	jarang	aktif
27	pinggirkota	banyak	baru	jarang	aktif
28	pinggirkota	banyak	baru	sedang	aktif
29	pinggirkota	banyak	baru	sering	aktif
30	pinggirkota	cukup	lama	sedang	aktif
31	pinggirkota	cukup	baru	sedang	aktif
32	pinggirkota	cukup	baru	jarang	aktif
33	pinggirkota	cukup	lama	jarang	aktif
34	pinggirkota	cukup	baru	sering	aktif
35	pinggirkota	minim	lama	sedang	aktif
36	pinggirkota	minim	baru	sedang	aktif
37	pinggirkota	minim	lama	sering	aktif

38	pinggirkota	minim	baru	sering	aktif
39	pinggirkota	minim	baru	jarang	aktif
40	desa	minim	baru	jarang	tidak
41	kota	minim	baru	sedang	tidak
42	kota	minim	baru	jarang	tidak
43	kota	cukup	lama	jarang	tidak
44	desa	minim	lama	jarang	tidak
45	desa	minim	lama	sedang	tidak
46	kota	minim	lama	jarang	tidak
47	desa	cukup	lama	jarang	tidak
48	kota	minim	lama	sedang	tidak
49	desa	cukup	baru	jarang	tidak
50	kota	cukup	lama	sedang	tidak
51	pinggirkota	cukup	lama	jarang	tidak
52	pinggirkota	minim	lama	jarang	tidak
53	pinggirkota	minim	baru	jarang	tidak
54	pinggirkota	minim	baru	sedang	tidak
55	kota	banyak	lama	jarang	aktif
56	pinggirKota	banyak	lama	jarang	aktif
57	desa	banyak	baru	jarang	aktif
58	desa	banyak	baru	sedang	aktif
59	desa	cukup	lama	sering	aktif



SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
PROGRAM PASCA SARJANA
MAGISTER ILMU KOMPUTER

NIM : 14002068
 Nama Mahasiswa : Supriatiningsih
 Dosen Pembimbing : Dr. Didi Rosiyadi, S.Kom, M.Kom
 Konsentrasi : Data Mining
 Judul Tesis : Penerapan Algoritma Klasifikasi C4.5 Berbasis AdaBoost Untuk Prediksi Loyalitas Pelanggan Distributor Pulsa Elektronik



No	Tanggal Bimbingan	Pokok Bahasan	Paraf Dosen Pembimbing
1	22 April 2018	Pembahasan Bab I	
2	22 April 2018	Pembahasan Bab II	
3	8 Juli 2018	Pembahasan Bab III Atribut	
4	8 Juli 2018	Pembahasan Bab III Klasifikasi	
5	15 Juli 2018	Pembahasan Bab IV	
6	15 Juli 2018	Pembahasan Bab V	
7	2 Agustus 2018	Abstrak dan Lampiran	
8	8 Agustus 2018	ACC Keseluruhan	

Bimbingan Tesis

- Dimulai pada tanggal : 22 April 2018
- Diakhiri pada tanggal : 8 Agustus 2018
- Jumlah pertemuan bimbingan : delapan kali

Disetujui Oleh,
 Dosen Pembimbing

[Dr. Didi Rosiyadi, S.Kom, M.Kom]

DAFTAR RIWAYAT HIDUP

A. Biodata Mahasiswa

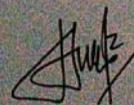
Nama : Supriatiningsih
Nim : 14002068
Tempat / Tanggal Lahir : Cilacap, 18 Juni 1993
Alamat : JL.Dr.Soetomo no:283 RT 02/04,Kec.Majenang
Kab. Cilacap, 53257

B. Riwayat Pendidikan Formal

1. SD NEGERI JENANG 04 Tahun Lulus 2006
2. SMP NEGERI 1 MAJENANG Tahun Lulus 2009
3. SMK KOMPUTAMA MAJENANG Tahun Lulus 2012
4. AMIK BSI PURWOKERTO Tahun Lulus 2015
5. STMIK NUSA MANDIRI Tahun Lulus 2016

Purwokerto, 28 Agustus 2018

Hormat Saya,



Supriatiningsih





**DISTRIBUTOR SERVER PULSA ELEKTRONIK DAN PPOB
SEMILAN RELOAD PURWOKERTO**

Jalan Gunung Wilis RT05 RW03, Kelurahan Bobosan, Purwokerto
Utara, Kab. Banyumas, 53127

SURAT KETERANGAN
No: 09/S.Ket/VIII/2018

Yang bertanda tangan dibawah ini :

Nama : Rifai Abdul Ghani

Jabatan : Pemilik Sembilan Reload Purwokerto

Dengan ini menerangkan bahwa yang tersebut dibawah ini :

Nama : Supriatiningsih

Nim : 14002068

Kampus : STMIK NUSA MANDIRI JAKARTA

Adalah benar telah melaksanakan riset atau penelitian pada Sembilan Reload Purwokerto dari tanggal 10 Mei s/d 10 Juni 2018.

Demikian surat keterangan riset ini dibuat untuk dapat digunakan sebagaimana mestinya.

Purwokerto, 10 Juni 2018



Pemilik

Rifai Abdul Ghani

