
KEY PERFORMANCE INDICATOR TOWER 4G L1800 PADA JARINGAN TELKOMSEL KOTA SEMARANG MENGGUNAKAN DECISION TREE C4.5

Annisa Sulistyaningsih

Fakultas Ilmu Komputer, Program Studi Teknik Informatika

Universitas Dian Nuswantoro

Email: annisasulistyaningsih06@gmail.com

Eko Hari Rachmawanto

Fakultas Ilmu Komputer, Program Studi Teknik Informatika

Universitas Dian Nuswantoro

Email: eko.hari@dsn.dinus.ac.id

Christy Atika Sari

Fakultas Ilmu Komputer, Program Studi Teknik Informatika

Universitas Dian Nuswantoro

Email: christy.atika.sari@dsn.dinus.ac.id

Florentina Esti Nilawati

Fakultas Ilmu Komputer, Program Studi Sistem Informasi

Universitas Dian Nuswantoro

Email: florentina.esti.nilawati@dsn.dinus.ac.id

ABSTRAK

Key Performance Indicator (KPI) digunakan untuk menentukan pengukuran yang objektif, melihat tren, dan mendukung pengambilan keputusan. KPI membutuhkan perencanaan yang matang, sehingga harus didukung oleh ketersediaan data dan informasi yang akurat serta konsisten. PT. Telekomunikasi Selular (Telkomsel) merupakan salah satu perusahaan dalam bidang telekomunikasi yang menggunakan KPI untuk menilai performa tower saat momen Natal dan Tahun Baru (NARU). KPI tersebut nantinya akan menjadi bahan evaluasi kedepannya. Oleh karenanya, dibutuhkan pohon keputusan untuk membantu menampilkan *site* yang mengalami gangguan berupa *redsite* saat momen tersebut berlangsung. Dalam penelitian ini, *Decision Tree* diimplementasikan menggunakan algoritma C4.5. C4.5 dinilai mudah diterapkan dan cocok digunakan untuk pengolahan data sederhana. Hasilnya olah data menunjukkan akurasi optimal yaitu 100 %.

Kata kunci: *key performance indicator, decision tree, C4.5, jaringan, redsite*

ABSTRACT

Key Performance Indicators (KPIs) are used to determine objective measurements, see trends, and support decision making. KPIs require careful planning, so they must be supported by the availability of accurate and consistent data and information. PT. Cellular Telekomunikasi (Telkomsel) is one of the companies in the telecommunications sector that uses KPI to assess tower performance during Christmas and New Year's Day (NARU). These KPIs will be used as material for future evaluations. Therefore, a decision tree is needed to help display sites that are experiencing disturbances in the form of *redsites* when the moment takes place. In this study, *Decision Tree* is implemented using the C4.5 algorithm. C4.5 is considered easy to implement and suitable for simple data processing. The result of data processing shows the optimal accuracy is 100%.

Keywords: *key performance indicator, decision tree, C4.5, network, redsite*

1. PENDAHULUAN

Teknologi yang berkembang sangat pesat di era sekarang memberikan dampak yang sangat besar bagi kehidupan manusia. Salah satunya berpengaruh di bidang informasi yaitu database. Informasi menjadi sangat penting dan dibutuhkan di berbagai perspektif kehidupan, baik dalam dunia perbankan, bisnis, pendidikan, dan lain sebagainya. Sebagai contoh di suatu perusahaan pasti terjadi banyak pertukaran informasi setiap harinya. Baik itu transaksi yang dilakukan oleh perusahaan tersebut, memberikan informasi perihal kinerja karyawan, dan lain sebagainya. Dengan semakin banyaknya informasi, maka jumlah data juga semakin meningkat. Jumlah data yang semakin besar memungkinkan terjadinya akumulasi yang besar pula. Oleh karenanya perusahaan membutuhkan sistem database, yaitu sistem yang berfungsi untuk menampung data, sehingga data yang dikumpulkan dapat diolah dan menghasilkan informasi yang bermanfaat bagi perusahaan [1]–[3]. Manfaat dari informasi dan pengetahuan yang terdapat dalam banyaknya data tersebut, saat ini dikenal dengan istilah data mining. Data mining merupakan sebutan yang digunakan untuk menguraikan pengetahuan yang ada di dalam database.

Data mining didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semiotomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar. Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database [3]–[5]. Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, artificial intelligence, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam database besar.

Telkomsel telah menyebarkan lebih dari 209.000 unit Base Transceiver Station (BTS) yang mencapai 97 % dari penduduk Indonesia. Sebagai operator seluler terbesar ke-7 di dunia, Telkomsel merupakan pemimpin pasar dalam industri telekomunikasi Indonesia dan melayani lebih dari 170 juta pelanggan. Hingga saat ini, Telkomsel telah menggelar lebih dari 209.000 unit BTS di seluruh Indonesia, termasuk di wilayah 3T (Tertinggal, Terdepan, Terluar) dan kawasan perbatasan. Dari jumlah tersebut, lebih dari 77.000 di antaranya merupakan BTS 4G. Lebih lanjut, untuk mengakselerasikan jangkauan serta kualitas jaringan 4G di Indonesia, Telkomsel mengimplementasikan teknologi 4G LTE ini di seluruh BTS yang menggunakan frekuensi 900Mhz. Jangkauan layanan 4G yang telah mencapai 95 % dari total wilayah populasi di Indonesia akan terus Telkomsel tingkatkan demi memberikan akses telekomunikasi terdepan bagi masyarakat di seluruh kota/kabupaten yang ada di Indonesia. Telkomsel telah menyebarkan lebih dari 209.000 unit Base Transceiver Station (BTS) yang mencapai 97 % dari penduduk Indonesia. Sebagai operator seluler terbesar ke-7 di dunia, Telkomsel merupakan pemimpin pasar dalam industri telekomunikasi Indonesia dan melayani lebih dari 170 juta pelanggan.

Bagi perusahaan besar seperti PT. Telekomunikasi Seluler (Telkomsel) yang menyediakan banyak layanan dengan jumlah yang besar memungkinkan terjadinya penumpukan data. Telkomsel adalah perusahaan yang bergerak dalam pengoperasian sistem telekomunikasi Telepon Bergerak Seluler dengan sistem GSM (Global System for Mobile Communication). Dengan database mempermudah Telkomsel dalam hal pengambilan keputusan serta melakukan evaluasi mempermudah untuk peningkatan kualitas jaringan yang mendukung kinerja telepon seluler (ponsel) dengan menggunakan query yang umumnya digunakan, namun jika data yang dimiliki besar dan kompleks maka dibutuhkan query yang kompleks pula agar hasil yang didapat lebih efektif.

Salah satu teknik pengolahan yang ada dalam data mining adalah klasifikasi. Klasifikasi merupakan proses penemuan sebuah model atau fungsi yang mendeskripsikan dan mengelompokkan data ke dalam kelas-kelas. Klasifikasi adalah proses menemukan sekumpulan pola atau fungsi-fungsi yang mendeskripsikan dan memisahkan kelas data satu dengan yang lainnya, dan digunakan untuk memprediksi data yang belum memiliki kelas data tertentu [5], [6]. Terdapat beberapa algoritma dalam teknik klasifikasi, antara lain algoritma C4.5. Algoritma C4.5 merupakan algoritma pada klasifikasi yang membentuk pohon keputusan (decision tree). Penerapan algoritma C4.5 dipilih karena dapat mengolah data dalam jumlah besar dan mengubahnya menjadi pohon keputusan yang mempresentasikan aturan dengan pengembangan metode [2], [7]–[9] menggunakan alat bantu WEKA. Dalam rangka memberikan pelayanan terbaik pada konsumennya, telkomsel memiliki dua agenda besar setiap tahunnya yaitu periode Ramadan dan Idul Fitri (RAFI) dan periode Natal dan Tahun Baru (NARU). Pada dua agenda besar ini biasanya terjadi lonjakan

pengguna data internet melebihi hari biasa. Oleh karenanya, data mining dirasa perlu diimplementasikan untuk membantu mengontrol trafik data pengguna yang terjadi pada saat periode tersebut.

Di penelitian kali ini penulis mengambil salah satu agenda, yaitu NARU. Pada momen NARU, Telkomsel memprediksi trafik akan mengalami peningkatan sebesar 26,34 %. Hal ini didasari perilaku pelanggan yang melakukan aktivitas di media sosial, streaming video, ataupun streaming musik. Untuk mengantisipasi lonjakan trafik tersebut, telkomsel telah melakukan optimalisasi dan drive test jaringan di berbagai daerah di Indonesia. Optimalisasi seluruh jaringan Telkomsel disebar di titik point of interest (POI) yang diprediksi akan menjadi titik berkumpulnya masyarakat dalam jumlah besar. Di Kota Semarang sendiri terdapat beberapa titik yang disinyalir akan menjadi pusat keramaian pada momen Natal dan Tahun Baru. Maka dari itu, Telkomsel telah menyiapkan POI di titik tersebut.

Dalam hal ini, algoritma C4.5 berperan sangat penting dalam mengolah database pelanggan, salah satunya untuk mengolah data Key Performance Indicators (KPI) pada jaringan 4G band L1800 Telkomsel khususnya yang ada di Kota Semarang. Fungsi dari data KPI yang dihimpun pada saat momen NARU yaitu menampilkan informasi berupa performa tower yang mengalami gangguan berupa redsite saat momen berlangsung dan menjadi bahan evaluasi kedepannya.

2. METODOLOGI PENELITIAN

Pada bagian ini akan menampilkan teori pendukung mengenai penelitian yang akan dibuat, kemudian penulis memilih beberapa sumber referensi berupa jurnal yang sudah dipublikasikan sebelumnya serta memberikan informasi mengenai konsep algoritma Decision Tree C4.5.

2.1 Penelitian Terkait

Implementation of Data Mining to Analyze Drug Cases Using C4.5 Decision Tree oleh Sri Wahyuni [3]. Tujuan dari penelitian tersebut yaitu ingin menganalisis data tahanan yang berada di rumah tahanan Labuhan Deli yang diekstraksi dengan decision tree. Dalam pemilihan variabel data tahanan sebagai atribut kondisi dan atribut keputusan, sangat mempengaruhi aturan maupun pengetahuan yang dihasilkan. Semakin banyak data yang diolah, maka semakin banyak pula aturan atau pengetahuan yang akan diperoleh. Dengan menggunakan algoritma C4.5 diharapkan dapat menyediakan informasi atau pengetahuan dari suatu aturan yang mudah dipahami oleh masyarakat luas karena telah dijelaskan dengan pohon keputusan.

Implementasi Algoritma C4.5 Untuk Menentukan Penerima Beasiswa di STT Harapan Medan, oleh Rismayanti [10]. Penelitian tersebut membahas mengenai cara untuk menentukan penerima beasiswa Peningkatan Prestasi Akademik (PPA) di STT Harapan Medan. Dalam menentukan penerima beasiswa, ada 4 parameter yang digunakan, yaitu IPK, Semester, Penghasilan Orangtua (PO), dan Jumlah Tanggungan Orangtua (JTO). Dari penelitian tersebut dihasilkan decision tree bahwa yang lolos sebagai penerima beasiswa adalah mahasiswa yang memiliki predikat IPK Cumlaude dan predikat Very Good dengan JTO yang banyak, sedangkan yang tidak lolos adalah mahasiswa yang memiliki predikat IPK Good dan predikat Very Good dengan JTO bernilai Cukup.

Implementation of Decision Tree Using C4.5 Algorithm in Decision Making of Loan Application by Debtor (Case Study: Bank Pasar of Yogyakarta Special Region), oleh Rafik Khairul Amin [8]. Penelitian ini membahas mengenai proses pemberian pinjaman kepada calon debitur di Bank Pasar Yogyakarta. Bagi pihak bank, pinjaman merupakan sumber pendapatan utama serta sumber resiko operasional bisnis terbesar. Memberikan pinjaman itu mudah, tetapi untuk mengatasi penghentian pembayaran pinjaman dibutuhkan keterampilan, pengalaman, waktu, serta biaya yang besar. Teknik data mining dengan algoritma c4.5 digunakan untuk membantu petugas analis pinjaman (loan analyst) dalam menilai kemampuan finansial calon debitur sebelum menyetujui pinjaman guna menghindari resiko penghentian pembayaran pinjaman. Selama proses penilaian, loan analyst perlu mengumpulkan data calon debitur ini. Kemudian, data akan diproses sesuai dengan prosedur bank terkait. Akhirnya, keputusan akan dibuat apakah pemohon memenuhi syarat untuk mendapatkan pinjaman dari bank. Dari penelitian tersebut, dihasilkan keputusan

berdasarkan pohon keputusan (decision tree) apakah pemohon memenuhi syarat untuk mendapatkan pinjaman dari bank.

Tabel 1. Penelitian Terkait pada Metode C4.5

No	Penulis	Judul	Hasil
1	Sri Wahyuni	Implementation of Data Mining to Analyze Drug Cases Using C4.5 Decision Tree [3]	Metode decision tree dengan algoritma C4.5 dapat menyediakan informasi atau pengetahuan dari suatu aturan yang mudah dipahami oleh masyarakat luas karena telah dijelaskan dengan pohon keputusan (decision tree)
2	Rismayanti	Implementasi Algoritma C4.5 Untuk Menentukan Penerima Beasiswa di STT Harapan Medan [10]	Ada 4 indikator yang digunakan sebagai prediktor yaitu semester, IPK, Penghasilan Orangtua (PO), dan Jumlah Tanggungan Orangtua (JTO). Hasil decision tree menghasilkan keputusan bahwa yang lolos sebagai penerima beasiswa adalah mahasiswa yang memiliki predikat IPK Cumlaude dan predikat Very Good dengan JTO yang banyak, sedangkan yang tidak lolos adalah mahasiswa yang memiliki predikat IPK Good dan predikat Very Good dengan JTO bernilai Cukup.
3	Rafik Khairul Amin, Indwiarti, Yuliant Sibaroni	Implementation of Decision Tree Using C4.5 Algorithm in Decision Making of Loan Application by Debtor (Case Study: Bank Pasar of Yogyakarta Special Region) [8]	Berdasarkan hasil pengujian, didapatkan nilai precision terbesar mencapai 78.08 % dengan rasio 90 % : 10 %. Nilai recall terbesar sebesar 96.4 % dengan rasio data 80 % : 20 %. Rasio 80 % : 20 % dianggap sebagai yang terbaik karena meraih accuracy tertinggi dan recall yang tinggi. Namun, data training yang besar tidak selalu menjamin accuracy, precision, dan recall yang lebih tinggi. Melainkan bergantung pada kualitas data yang digunakan sebagai data training.

2.2 Algoritma C4.5

Algoritma C4.5 digunakan dalam Data Mining sebagai Decision Tree Classifier yang dapat digunakan untuk mengambil keputusan, berdasarkan sample data tertentu (univariate atau multivariate predictors) [11]. C4.5 merupakan algoritma yang dikembangkan oleh Rose Quinlan yang menghasilkan pohon keputusan (Decision Tree) [12], yang dapat digunakan sebagai solusi dari masalah klasifikasi. Algoritma ini memperbaiki algoritma ID3 dengan menangani atribut kontinyu dan diskrit serta menangani nilai yang hilang (missing value). Langkah pertama yang harus dilakukan ketika membuat pohon keputusan adalah memilih variabel yang digunakan sebagai akar dari pohon keputusan. Caranya yaitu dengan menentukan nilai entropy, gain, split info, dan gain ratio [7].

a. Entropy

Entropy merupakan parameter yang digunakan untuk mengukur tingkat keberagaman (heterogeneity) dari suatu kumpulan data. Semakin besar nilai entropy, maka semakin besar pula tingkat keberagaman data. Perhitungan rumus entropy seperti pada (1) dan entropy untuk setiap variabel seperti pada (2), dimana m adalah jumlah klasifikasi dan p_i adalah proporsi sample (peluang) untuk kelas i , A adalah variabel, V adalah nilai yang mungkin untuk variabel A , $|S_v|$ adalah jumlah sampel untuk v , $|S|$ adalah jumlah sampel untuk seluruh data, Entropy (S_v) adalah

$$Entropy(S) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (1)$$

$$Entropy_A(S) = \sum_v \frac{|S_v|}{|S|} Entropy(S_v) \quad (2)$$

b. Gain

Gain dari suatu variabel adalah selisih dari nilai entropy total dengan entropy dari variabel tersebut. Pada algoritma C4.5, gain dipakai untuk menentukan variabel yang menjadi node dari suatu pohon keputusan. Suatu variabel yang memiliki gain tertinggi akan dijadikan node di pohon keputusan. Gain dapat dirumuskan seperti pada (3) sampai (5) dimana D adalah jumlah data sampel dan D_j adalah jumlah masing-masing pada setiap atribut. Split Info berguna sebagai pembagi dari $\text{Gain}(A)$ yang akan menghasilkan Gain Ratio, sedangkan Gain Ratio merupakan hitungan lain yang berguna untuk mengatasi masalah pada atribut atau parameter yang memiliki nilai bervariasi. Gain Ratio yang memiliki nilai tertinggi nantinya akan digunakan sebagai atribut test untuk node.

$$\text{Gain}(A) = \text{Entropy}(S) - \text{Entropy}_A(S) \quad (3)$$

$$\text{SplitInfo}_A(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} \cdot \log_2 \left(\frac{|D_j|}{|D|} \right) \quad (4)$$

$$\text{GainRatio}(A) = \frac{\text{Gain}(A)}{\text{SplitInfo}_A(D)} \quad (5)$$

2.3 Key Performance Indikator (KPI)

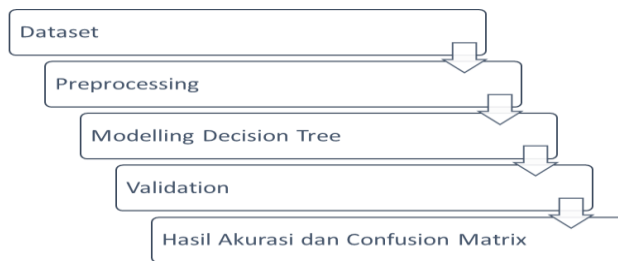
Terdapat beberapa hal yang berhubungan dengan KPI, antara lain:

1. Relevan (Relevant). Indikator harus relevan dengan perusahaan. Salah satu cara untuk memastikan keterkaitannya adalah dengan menghubungkan indikator kinerja (performance indicators) dengan sasaran (goals) yang sesuai rencana dan tujuan perusahaan atau divisi layanan tertentu. Indikator idealnya juga harus relevan bagi orang yang menyediakan data dan bagi pengguna PI, meskipun begitu, indikator tunggal mungkin tidak relevan bagi semua pengguna karena perbedaan kepentingan dan sudut pandang.
2. Ketentuan yang Jelas (Clear Definition). Performance Indicators harus memiliki ketentuan yang jelas dan dapat dipahami untuk memastikan pengelompokan yang konsisten dan perbandingan yang adil. Ketentuan yang ambigu dapat menyebabkan kesalahan penafsiran dan kekacauan. Untuk memperketat atau memperluas ketentuan juga dapat menimbulkan masalah.
3. Mudah dipahami dan digunakan (Easy to understand and use). Merupakan hal penting bahwa indikator menggambarkan aturan yang mudah dipahami oleh pengguna informasi, bahkan jika aturan tersebut menggunakan istilah teknis. Indikator yang difokuskan untuk publik harus menghindari aturan aksen atau konsep yang abstrak.
4. Sebanding (Comparable). Indikator idealnya harus dapat diperbandingkan konsistensi keduanya antara perusahaan dengan waktu lembur. Aspek penting dari indikator kinerja yang dapat disamakan adalah menyesuaikan kondisi ataupun konteks perbandingan yang sedang terjadi. Keadaan eksternal dan internal dapat berbeda sedemikian rupa sehingga perbandingan tidak valid.
5. Teruji (Verifiable). Indikator juga perlu dikumpulkan dan dihitung sedemikian rupa sehingga memungkinkan informasi dan data teruji. Indikator harus didasarkan pada sistem pengumpulan data yang kuat, dan memungkinkan manajer untuk memverifikasi keakuratan informasi dan konsistensi metode yang digunakan.
6. Hemat biaya (Cost Effective). Kriteria penting lainnya adalah menyeimbangkan biaya pengumpulan informasi dengan kegunaannya. Jika memungkinkan, suatu indikator harus didasarkan pada informasi yang sudah tersedia dan dikaitkan dengan aktivitas pengumpulan data yang ada.
7. Responsif (Responsive). Indikator kinerja harus responsif terhadap perubahan.
8. Memperbolehkan adanya perubahan (Allow Innovation). Ketentuan suatu indikator seharusnya tidak boleh menghalangi perusahaan untuk mengembangkan proses inovatif atau menghasilkan metode alternatif, sistem atau prosedur untuk peningkatan layanan. PI idealnya harus dibangun untuk memungkinkan terjadinya inovasi.
9. Valid secara statistik (Statistically Valid). Indikator harus valid secara statistik.

10. Terjadi pada saat yang tepat (Timely). PI harus didasarkan pada data yang tersedia dalam skala waktu yang wajar. Skala waktu ini tergantung pada penggunaan data. Beberapa data dikumpulkan setiap minggu atau bahkan setiap hari, karena diperlukan dalam manajemen operasi layanan, sedangkan yang lain tersedia setahun sekali untuk tujuan yang lebih strategis dan jangka panjang.

2.4 Metode Yang Diusulkan

Pada penelitian ini metode yang digunakan ada decision tree dengan menggunakan algoritma C4.5. Dalam tahap modeling ini nantinya akan dicari akurasi dan confusion matriksnya.



Gambar 1. Tahapan Usulan Metode

3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah sumber data yang didapatkan dari data site yang tersebar selama event NARU 2020 di Kota Semarang sebanyak 2237 data. Data yang didapat terdiri dari beberapa atribut antara lain *sitename*, *siteid*, *LAC*, *CI*, *cellname*, *band*, *Type Band*, *POIname*, *Call Setup Success Rate*, *Service Drop Rate*, *Average User Number*, *Maximum User Number*, *Uplink Traffic Volume*, dan *Downlink Traffic Volume* seperti diilustrasikan pada Tabel 2.

Tabel 2. Atribut Dataset

No	Atribut	Proses
1	Sitename	Data Cleaning
2	SITEID	Data Cleaning
3	LAC	Data Cleaning
4	CI	Data Cleaning
5	Cellname	Data Cleaning
6	Band	Data Cleaning
7	Type Band	Data Cleaning
8	POI Name	Data Cleaning
9	Call Setup Success Rate	Atribut Terpakai
10	Service Drop Rate	Atribut Terpakai
11	Average User Number	Atribut Terpakai
12	Maximum User Number	Atribut Terpakai
13	Uplink Traffic Volume	Atribut Terpakai
14	Downlink Traffic Volume	Atribut Terpakai

Tabel 1 merupakan atribut yang terdapat dalam dataset asli yang didapatkan, Setelah melakukan pemilihan atribut dataset asli yang didapat nantinya akan diolah di data cleaning. Data cleaning dilakukan untuk mengurangi adanya data yang kurang lengkap informasinya maupun data yang bermasalah. Untuk itu dengan pembersihan data atau data cleaning diharapkan dapat memudahkan proses mining nantinya.

Berdasarkan data yang diperoleh, terdapat variable yang tidak berpengaruh seperti sitename, siteid, LAC, CI, cellname, band, Type Band, dan POIname seperti ditunjukkan melalui Tabel 3.

Tabel 3. Sampel data setelah proses penghapusan variabel

No	Call Setup Success Rate	Service Drop Rate	Average User Number	Maximum User Number	Uplink Traffic Volume	Downlink Traffic Volume
1	99.7150064	0.058263635	77.37608333	177	15213.22628	113612.8099
2	99.7150064	0.058263635	77.37608333	177	15213.22628	113612.8099
3	99.7150064	0.058263635	77.37608333	177	15213.22628	113612.8099
4	99.78513745	0.057229778	29.131	69	3403.212564	36014.70031
5	99.78513745	0.057229778	29.131	69	3403.212564	36014.70031
6	99.78513745	0.057229778	29.131	69	3403.212564	36014.70031
7	99.59877568	0.074698477	59.21766667	183	5037.441246	64323.11503
8	99.59877568	0.074698477	59.21766667	183	5037.441246	64323.11503
9	99.59877568	0.074698477	59.21766667	183	5037.441246	64323.11503
10	99.74312056	0.068959922	64.23095833	183	5144.531878	64220.2737
11	99.74312056	0.068959922	64.23095833	183	5144.531878	64220.2737
12	99.74312056	0.068959922	64.23095833	183	5144.531878	64220.2737
13	99.86898314	0.03478088	32.64241667	106	5666.77895	59781.17778
14	99.86898314	0.03478088	32.64241667	106	5666.77895	59781.17778
15	99.86898314	0.03478088	32.64241667	106	5666.77895	59781.17778
16	99.73117598	0.064151163	12.529875	49	1581.743426	18309.24146
17	99.73117598	0.064151163	12.529875	49	1581.743426	18309.24146
18	99.73117598	0.064151163	12.529875	49	1581.743426	18309.24146
19	99.73141234	0.079813705	32.913125	67	2316.929591	37297.64907
20	99.73141234	0.079813705	32.913125	67	2316.929591	37297.64907
21	99.73141234	0.079813705	32.913125	67	2316.929591	37297.64907
22	99.82795115	0.052015383	35.203375	75	2690.625333	45033.39482
23	99.82795115	0.052015383	35.203375	75	2690.625333	45033.39482
24	99.82795115	0.052015383	35.203375	75	2690.625333	45033.39482
25	99.51472357	0.103287844	95.69045833	236	16740.57772	145307.059

Dilihat dari tabel diatas, tabel tersebut berisi variabel variabel yang telah disaring untuk digunakan dalam penelitian ini. Variabel yang tidak digunakan telah dihapus untuk mempermudah dalam mengolah data agar sesuai dengan tema yang dibahas. Tahap selanjutnya adalah transformasi data. Pada Tahap ini, transformasi dilakukan untuk mengubah bentuk data numerik menjadi data yang bertipe kategoris yang mana agar semua variabel menjadi satu tipe. Perubahan tersebut bertujuan untuk memudahkan dalam proses perhitungan selanjutnya dalam mengolah data menggunakan algoritma c4.5. Berikut merupakan proses transformasi data yang dilakukan. Atribut proses tranformasi dapat dilihat pada Tabel 4, sedangkan hasil proses transformasi data dapat dilihat pada Tabel 5. Tujuan dilakukannya transformasi hanya untuk mempermudah pemahaman sewaktu membuat rule pohon keputusan.

Tabel 4. Sampel data setelah proses penghapusan variabel

No	Atribut	Proses
1	Call Setup Success Rate	>=98.40568428, <98.40568428
2	Service Drop Rate	>=0.096507658, <0.096507658
3	Average User Number	>=27.48992355, <27.48992355
4	Maximum User Number	>=83.85650425, <83.85650425
5	Uplink Traffic Volume	>=4123.88685, <4123.88685
6	Downlink Traffic Volume	>=43869.61712, <43869.61712
7	Call Setup Success Rate	>=98.40568428, <98.40568428
8	Service Drop Rate	>=0.096507658, <0.096507658

Tabel 4. Data yang sudah di transformasi

<i>call_action</i>	<i>service_action</i>	<i>average_action</i>	<i>max_action</i>	<i>uplink_action</i>	<i>downlink_action</i>	<i>status</i>
Good	Strong	High	Abnormal	High	Yes	Not Redsite
Good	Strong	High	Abnormal	High	Yes	Not Redsite
Good	Strong	High	Abnormal	High	Yes	Not Redsite
Good	Strong	High	Normal	Low	No	Not Redsite
Good	Strong	High	Normal	Low	No	Not Redsite
Good	Strong	High	Normal	Low	No	Not Redsite
Good	Strong	High	Abnormal	High	Yes	Not Redsite
Good	Strong	High	Abnormal	High	Yes	Not Redsite
Good	Strong	High	Abnormal	High	Yes	Not Redsite
Good	Strong	High	Abnormal	High	Yes	Not Redsite
Good	Strong	High	Abnormal	High	Yes	Not Redsite
Good	Strong	High	Abnormal	High	Yes	Not Redsite
Good	Strong	High	Abnormal	High	Yes	Not Redsite
Good	Strong	High	Abnormal	High	Yes	Not Redsite
Good	Strong	Low	Normal	Low	No	Not Redsite
Good	Strong	Low	Normal	Low	No	Not Redsite
Good	Strong	Low	Normal	Low	No	Not Redsite
Good	Strong	High	Normal	Low	No	Not Redsite
Good	Strong	High	Normal	Low	No	Not Redsite

Sampel data acak digunakan sebagai data awal untuk melakukan perhitungan manual C4.5. Data training yang digunakan yaitu sebanyak 2237 data. Setelah ditentukan data awal, langkah selanjutnya adalah menghitung entropy dari masing-masing data. Atribut terpilih = Service Drop Rate karena memiliki nilai gain paling tinggi dibandingkan atribut lainnya yaitu sebesar 0.094489913970958. Atribut terpilih = Uplink Traffic Volume karena memiliki nilai gain paling tinggi dibandingkan atribut lainnya yaitu sebesar 0.56170233580873. Berikut ini adalah tahapan perhitungan secara manual untuk proses KPI tower yang telah dilakukan.

Entropy Total

Jumlah Kasus (S) = 2237 data

Redsite = 83 Data

Not Redsite = 2154 Data

$$\text{Entropy (S)} = \left(-\left(\frac{2154}{2237}\right) \times \log_2\left(\frac{2154}{2237}\right)\right) + \left(-\left(\frac{83}{2237}\right) \times \log_2\left(\frac{83}{2237}\right)\right) \\ = 0.228849348$$

Entropy Call Setup Success Rate (Good)

Jumlah Kasus (S) = 2094 data

Redsite = 83 Data

Not Redsite = 2011 Data

$$\text{Entropy (S)} = \left(-\left(\frac{2011}{2094}\right) \times \log_2\left(\frac{2011}{2094}\right)\right) + \left(-\left(\frac{83}{2094}\right) \times \log_2\left(\frac{83}{2094}\right)\right) \\ = 0.240625633$$

Entropy Call Setup Success Rate (Bad)

Jumlah Kasus (S) = 143 data

Redsite = 0 Data

Not Redsite = 143 Data

$$\text{Entropy (S)} = \left(-\left(\frac{143}{143}\right) \times \log_2\left(\frac{143}{143}\right)\right) + \left(-\left(\frac{0}{143}\right) \times \log_2\left(\frac{0}{143}\right)\right) \\ = 0.0030491371$$

Entropy Service Drop Rate (Weak)

Jumlah Kasus (S) = 418 data

Redsite = 335 Data

Not Redsite = 83 Data

$$\begin{aligned} \text{Entropy (S)} &= \left(-\left(\frac{335}{418}\right) \times \log_2\left(\frac{335}{418}\right)\right) + \left(-\left(\frac{83}{418}\right) \times \log_2\left(\frac{83}{418}\right)\right) \\ &= 0.719047976 \end{aligned}$$

Entropy Service Drop Rate (Strong)

Jumlah Kasus (S) = 1819 data

Redsite = 0 Data

Not Redsite = 1819 Data

$$\begin{aligned} \text{Entropy (S)} &= \left(-\left(\frac{1819}{1819}\right) \times \log_2\left(\frac{1819}{1819}\right)\right) + \left(-\left(\frac{0}{1819}\right) \times \log_2\left(\frac{0}{1819}\right)\right) \\ &= 0 \end{aligned}$$

Entropy Average User Number (High)

Jumlah Kasus (S) = 936 data

Redsite = 83 Data

Not Redsite = 853 Data

$$\begin{aligned} \text{Entropy (S)} &= \left(-\left(\frac{853}{936}\right) \times \log_2\left(\frac{853}{936}\right)\right) + \left(-\left(\frac{83}{936}\right) \times \log_2\left(\frac{83}{936}\right)\right) \\ &= 0.432032326 \end{aligned}$$

Entropy Average User Number (Low)

Jumlah Kasus (S) = 1301 data

Redsite = 0 Data

Not Redsite = 1301 Data

$$\begin{aligned} \text{Entropy (S)} &= \left(-\left(\frac{1301}{1301}\right) \times \log_2\left(\frac{1301}{1301}\right)\right) + \left(-\left(\frac{0}{1301}\right) \times \log_2\left(\frac{0}{1301}\right)\right) \\ &= 0 \end{aligned}$$

Entropy Max User Number (Normal)

Jumlah Kasus (S) = 1361 data

Redsite = 0 Data

Not Redsite = 1361 Data

$$\begin{aligned} \text{Entropy (S)} &= \left(-\left(\frac{1361}{1361}\right) \times \log_2\left(\frac{1361}{1361}\right)\right) + \left(-\left(\frac{0}{1361}\right) \times \log_2\left(\frac{0}{1361}\right)\right) \\ &= 0 \end{aligned}$$

Entropy Max User Number (Abnormal)

Jumlah Kasus (S) = 876 Data

Redsite = 83 Data

Not Redsite = 793 Data

$$\begin{aligned} \text{Entropy (S)} &= \left(-\left(\frac{793}{876}\right) \times \log_2\left(\frac{793}{876}\right)\right) + \left(-\left(\frac{83}{876}\right) \times \log_2\left(\frac{83}{876}\right)\right) \\ &= 0.452125327 \end{aligned}$$

Entropy Uplink Traffic Volume (High)

Jumlah Kasus (S) = 801 data

Redsite = 83 Data

Not Redsite = 718 Data

$$\begin{aligned} \text{Entropy (S)} &= \left(-\left(\frac{718}{801}\right) \times \log_2\left(\frac{718}{801}\right)\right) + \left(-\left(\frac{83}{801}\right) \times \log_2\left(\frac{83}{801}\right)\right) \\ &= 0.480368274 \end{aligned}$$

Entropy Uplink Traffic Volume (Low)

Jumlah Kasus (S) = 1436 data

Redsite = 0 Data

Not Redsite = 1436 Data

$$\begin{aligned} \text{Entropy (S)} &= \left(-\left(\frac{1436}{1436}\right) \times \log_2\left(\frac{1436}{1436}\right)\right) + \left(-\left(\frac{0}{1436}\right) \times \log_2\left(\frac{0}{1436}\right)\right) \\ &= 0 \end{aligned}$$

Entropy Downlink Traffic Volume (Yes)

Jumlah Kasus (S) = 906 data

Redsite = 83 Data

Not Redsite = 823 Data

$$\text{Entropy (S)} = \left(-\left(\frac{823}{906}\right) \times \log_2\left(\frac{823}{906}\right)\right) + \left(-\left(\frac{83}{906}\right) \times \log_2\left(\frac{83}{906}\right)\right) \\ = 0.441825974$$

Entropy Downlink Traffic Volume (No)

Jumlah Kasus (S) = 1331 data

Redsite = 0 Data

Not Redsite = 1331 Data

$$\text{Entropy (S)} = \left(-\left(\frac{1331}{1331}\right) \times \log_2\left(\frac{1331}{1331}\right)\right) + \left(-\left(\frac{0}{1331}\right) \times \log_2\left(\frac{0}{1331}\right)\right) \\ = 0$$

Menentukan Gain node 1 menggunakan (6) berikut.

$$\text{Gain (S, A)} = \text{Entropy (S)} - \sum_{i=1}^n \frac{|S_i|}{|S|} \times \text{Entropy (S}_i\text{)} \quad (6)$$

Gain Call Setup Success Rate = 0.003605684078907

Gain Service Drop Rate = 0.094489913970958

Gain Average User Number = 0.048079452454427

Gain Max User Number = 0.051798929529573

Gain Uplink Traffic Volume = 0.05684443631154

Gain Downlink Traffic Volume = 0.049906865595641

Atribut terpilih = Service Drop Rate karena memiliki nilai gain paling tinggi dibandingkan atribut lainnya yaitu sebesar 0.094489913970958.

Menghitung Entropy dan Gain mining 2/node 2

Entropy Total = 0.7190479761562691

Entropy Call Setup Success Rate (Good) = 0.7615249385854395

Entropy Call Setup Success Rate (Bad) = 0

Entropy Average User Number (High) = 0.974823751862113

Entropy Average User Number (Low) = 0

Entropy Max User Number (Normal) = 0

Entropy Max User Number (Abnormal) = 0.996719074075869

Entropy Uplink Traffic Volume (High) = 0.6577047787442195

Entropy Uplink Traffic Volume (Low) = 0

Entropy Downlink Traffic Volume (Yes) = 0.9749758489189322

Entropy Downlink Traffic Volume (No) = 0

Gain Call Setup Success Rate = 0.034039897505214

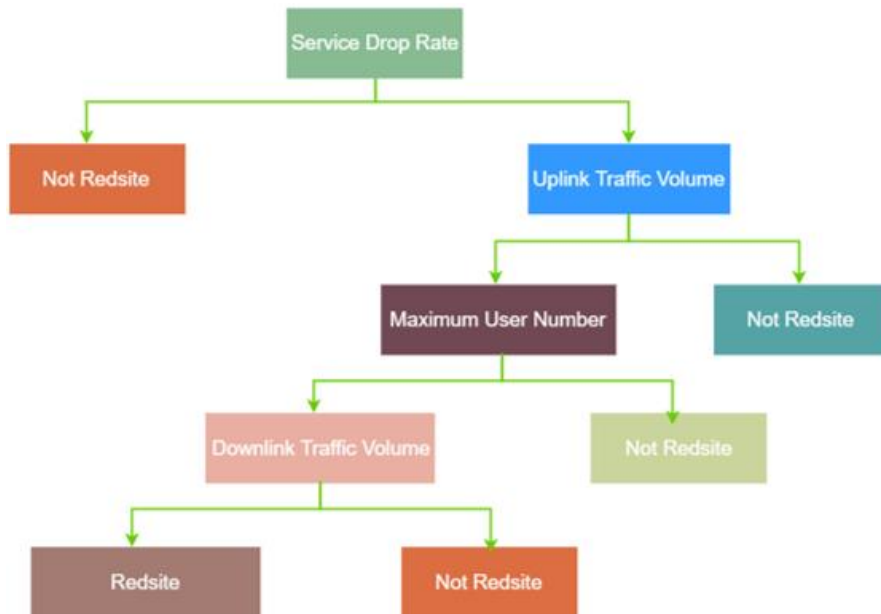
Gain Average User Number = 0.24329667191881

Gain Max User Number = 0.29460779685892

Gain Uplink Traffic Volume = 0.56170233580873

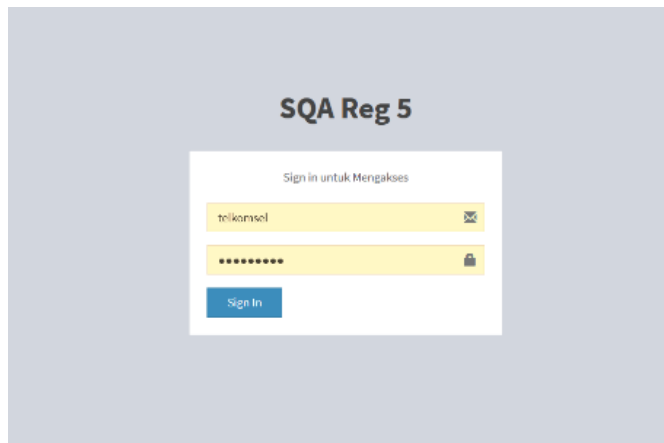
Gain Downlink Traffic Volume = 0.39250104176008

Atribut terpilih = Uplink Traffic Volume karena memiliki nilai gain paling tinggi dibandingkan atribut lainnya yaitu sebesar 0.56170233580873.

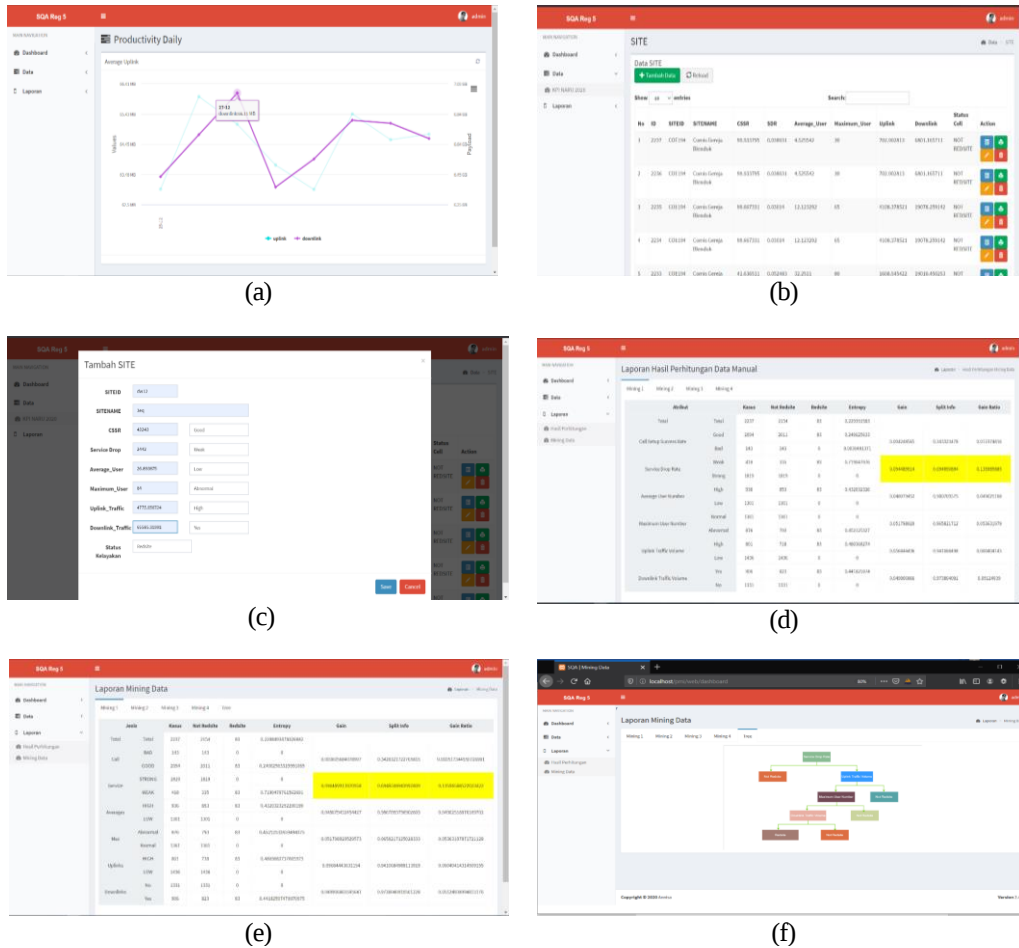


Gambar 1. Pohon keputusan

Pada gambar pohon keputusan diatas dapat diketahui rule-rule yang dapat mengklasifikasikan data. Pengujian algoritma yang digunakan pada penelitian ini menggunakan aplikasi Weka. Weka ini digunakan untuk mengetahui serta mengukur seberapa kuat metode yang digunakan jika diimplementasikan pada tools ini. Aplikasi yang dibuat sebagai pengimplementasian metode yang digunakan sebagai pengklasifikasian menggunakan tools Sublime Text. Gambar 2 dan Gambar 3 merupakan interface dari aplikasi yang telah dibuat.



Gambar 2. Halaman Login



Gambar 3. (a) Tampilan Halaman Dashboard, (b) tampilan data site POI 4G, (c) menu tambah data, (d) halaman laporan perhitungan, (e) halaman hasil perhitungan mining, (f) tampilan pohon keputusan

Berdasarkan Gambar 3, point a merupakan halaman dashboard terdapat chart Traffic yang diambil dari data POI 4G yang ada di Kota Semarang. Lalu, di sebelah kiri terdapat menu dashboard, data, dan laporan. Point b merupakan tampilan data site POI 4G yang terdapat di Kota Semarang. Pada halaman ini juga terdapat fitur cari data. Point c merupakan menu tambah data. Menu tambah data berfungsi untuk menginput data dan nantinya muncul status dari site tersebut sesuai pohon keputusan yang telah dibentuk sebelumnya. Point d merupakan halaman laporan perhitungan. Pada halaman laporan perhitungan, data yang telah dihitung manual dengan excel ditampilkan pada halaman ini. Fungsinya adalah untuk menyesuaikan hitungan manual excel dengan hitungan hasil mining. Point e merupakan halaman hasil perhitungan mining. Pada halaman ini berisi hasil hitungan mining yang berasal dari database, proses mining dilakukan hingga mining ke 4 hingga mendapat hasil akhir gain ratio bernilai 1. Point f merupakan halaman tampilan pohon keputusan. Pada halaman ini berisi pohon keputusan yang terbentuk dari pengolahan data sebelumnya.

4. KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan mengenai implementasi data mining menggunakan algoritma C4.5 untuk menentukan key performance indicator 4G L1800 pada jaringan Telkomsel di Kota

Semarang telah menghasilkan akurasi sebesar 100% dengan menggunakan 2237 data yang digunakan maka penelitian ini dapat dikatakan memiliki akurasi yang sangat baik dalam mengeksekusi sebuah data dalam menentukan status site saat event NARU 2020. Penelitian ini menghasilkan sebuah website untuk PT. Telkomsel Semarang yang dapat digunakan untuk membantu menampilkan informasi Key Performance Indicators (KPI) Telkomsel pada jaringan 4G di Kota Semarang pada saat agenda NARU serta untuk mengetahui penyebab terjadinya redsite. Untuk lebih menyempurnakan penelitian yang telah dibuat adapun saran yang diberikan. Penelitian ini dapat dikembangkan lebih lanjut dengan menggabungkan atau membandingkan dengan algoritma lain sehingga nantinya akan mendapatkan hasil yang lebih baik. Penyempurnaan aplikasi web yang lebih baik agar bisa digunakan oleh pihak yang membutuhkan.

DAFTAR PUSTAKA

- [1] S. Sathyadevan and R. R. Nair, 2015. "Computational Intelligence in Data Mining - Volume 1: Proceedings of the International Conference on CIDM, 20-21 December 2014," *Smart Innov. Syst. Technol.*, vol. 31, pp. 549–562.
- [2] W. Jia and L. J. Huang, 2010. "Improved C4.5 decision tree," *Int. Conf. Internet Technol. Appl. ITAP 2010 - Proc.*, vol. 3, no. 3, pp. 341–345.
- [3] S. Wahyuni, 2018. "Implementation of Data Mining to Analyze Drug Cases Using C4.5 Decision Tree," *J. Phys. Conf. Ser.*, vol. 970, no. 1.
- [4] H. Chauhan and A. Chauhan, 2013. "Implementation of decision tree algorithm c4. 5," *Int. J. Sci. Res. Publ.*, vol. 3, no. 10, pp. 4–6.
- [5] Y. T. Samuel, J. J. Hutapea, and B. Jonathan, 2019. "Predicting the timeliness of student graduation using decision tree c4.5 algorithm in universitas advent Indonesia," *Proc. 2019 Int. Conf. Inf. Commun. Technol. Syst. ICTS 2019*, pp. 276–280.
- [6] W. Dai and W. Ji, 2014. "A mapreduce implementation of C4.5 decision tree algorithm," *Int. J. Database Theory Appl.*, vol. 7, no. 1, pp. 49–60.
- [7] R. Revathy and R. Lawrance, 2017. "Comparative Analysis of C4.5 and C5.0 Algorithms on Crop Pest Data," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 5, no. 1, pp. 50–58.
- [8] R. K. Amin, Indwiarti, and Y. Sibaroni, 2015. "Implementation of decision tree using C4.5 algorithm in decision making of loan application by debtor (Case study: Bank pasar of Yogyakarta Special Region)," in *2015 3rd International Conference on Information and Communication Technology, ICoICT 2015*, vol. 0, pp. 75–80.
- [9] X. Zhu, J. Wang, H. Yan, and S. Wu, 2009. "Research and application of the improved algorithm C4.5 on decision tree," in *Proceedings of the International Symposium on Test and Measurement*, vol. 2, pp. 184–187.
- [10] R. Rismayanti, 2017. "Implementasi Algoritma C4.5 Untuk Menentukan Penerima Beasiswa Di Stt Harapan Medan," *J. Media Infotama*, vol. 12, no. 2, pp. 116–120.
- [11] B. HSSINA, A. MERBOUHA, H. EZZIKOURI, and M. ERRITALI, 2014. "A comparative study of decision tree ID3 and C4.5," *Int. J. Adv. Comput. Sci. Appl.*, vol. 4, no. 2.
- [12] Y. Zheng, L. Peng, L. Lei, and Y. Junjie, 2005. "R-C4.5 decision tree model and its applications to health care dataset," in *2005 International Conference on Services Systems and Services Management, Proceedings of ICSSSM'05*, vol. 2, pp. 1099–1103.