

Penerapan Algoritma C.45 Untuk Memprediksi Indikator Website Yang Baik

Rika Harman

Universitas Putera Batam Jl R Soeprapto Muka Kuning – Batam 29434 Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 06 Juli 2018

Revisi Akhir: 01 September 2018

Diterbitkan Online: 15 September 2018

KATA KUNCI

Decision tree, Algoritma C.45, Website, indikator

KORESPONDENSI

No HP: +62811.696.784

E-mail: rika.harman@gmail.com

A B S T R A C T

The website for a company for now is not just a trend but it is a necessity that absolutely must not be, because with the website all the general information of the company can be shared to its customers directly tanpa must be hampered by time and distance. as an example for a particular class of customers can find the information he wants without having to come directly keperusahaan. For that a website must contain important information about the company, the website should be able to represent customer service if asked by customers about company information related to customer needs. In this journal will be discussed one website of the company with 7 (seven) indicators, then of the seven indicators will be seen which indicators most influential on the website itself, this is certainly in the direction of customer satisfaction using the website. so it seems that the existing indicators have different effects between each other. To make this happen the researcher uses clustering method by using algoritma C.45 with WEKA software version 4.0.2. The expected results will be to help the company in the process of developing a better website again.

1. PENDAHULUAN

Berbicara masalah teknologi yang paling mutakhir pada abad ini jelas hal ini tidak bisa dipisahkan dari internet, karena internet merupakan jaringan yang terhubung menggunakan standar TCP/IP sebagai protocol pertukaran paket untuk melayani miliaran pengguna diseluruh dunia. Sudah bukan merupakan rahasia lagi bahwasannya untuk saat sekarang internet telah menjangkau seluruh aspek kehidupan, bahkan sampai dalam kehidupan sehari-hari kita pun tak lepas dari yang namanya internet. Berbagai contoh dikehidupan dapat kita lihat seperti banyak diantara kita yang tak bisa lepas dari handphone setiap harinya, setiap jam bahkan setiap menitnya kita tidak lepas dari penggunaan *smartphone* untuk keperluan tertentu dan hal ini juga tidak lepas dari yang namanya internet. Terkadang dalam realitasnya banyak dari diantara kita duduk bersamaan tapi asyik dengan *smartphone* sendiri-sendiri, atau juga banyak dari diantara kita yang resah bila ketinggalan *smartphone* ketimbang ketinggalan dompet atau peralatan lainnya. Ketinggalan *smartphone* akan kebingungan dan tidak tau harus mengerjakan apa karena sudah menjadi kebiasaan. Sedangkan jika ketinggalan peralatan lain tidak menjadi masalah.

Mengutip dari perkataan Bill gates dalam *chairman of Microsoft Corp* "Internet akan menyapu kita semua seperti sebuah gelombang pasang yang akan menenggelamkan setiap orang yang dilewatinya yang tidak siap untuk hidup dalam sebuah masyarakat informasi". Jadi apabila kita atau bisnis yang sedang geluti tidak ingin tenggelam akibat sapuan dari gelombang pasang yang bernama internet maka kita harus mulai

mulai belajar dan harus menguasai tentang internet apabila ini kita lakukan secara tidak langsung kita telah bisa berselancar ditengah derasnya obak dari gelombang yang diakibatkan oleh internet.

Hal serupa juga berimbas dengan dunia organisasi dan perusahaan dibelahan dunia saat ini, perusahaan dibelahan dunia berlomba-lomba untuk menguasai dan mempergunakan internet, karena untuk zaman sekarang perusahaan yang tidak menggunakan internet akan sangat ketinggalan dan tidak akan dikenal oleh masyarakat zaman sekarang yang sudah melek teknologi apalagi teknologi yang berkaitan dengan internet. Sudah tidak bisa dipungkiri lagi bahwasannya perusahaan yang menguasai teknologi seperti internet tidak perlu repot-repot memperkenalkan perusahaan kepada konsumen tapi masyarakatlah dalam hal ini konsumen yang nantinya akan mencari informasi tentang perusahaan tersebut. Pada ininya jika perusahaan menguasai teknologi internet akan mudah merebut pasar dibanding dengan perusahaan lain yang tidak kenal dengan internet hal ini berbanding searah dengan pola hidup masyarakat zaman sekarang yang lebih memahami teknologi. Salah satu cara bagi perusahaan dalam melayani konsumen melalui fasilitas internet adalah dengan adanya *website* (seperti diketahui suatu hal yang identik dengan internet adalah *website*), melalui fasilitas ini maka dapat terjadi komunikasi antara konsumen dengan perusahaan, dimana konsumen tidak harus datang ke perusahaan tersebut jika membutuhkan jasa dari perusahaan itu melainkan proses pelayanannya bisa dilakukan dari jarak jauh yang nantinya konsumen tetap akan mendapatkan apa yang dia inginkan tentang informasi serta layanan dari perusahaan tersebut. Hal ini jelas cocok dengan model

masyarakat zaman sekarang karena ingin sesuatu tapi tak harus keluar uang dan tenaga banyak untuk mewujudkan keinginan tersebut.

Seperti diketahui juga bahwa secara umum *website* adalah sebuah halaman yang menyajikan informasi baik dalam bentuk tulisan, gambar suara atau video yang diletakkan didalam sebuah *server hosting* yang mana untuk mengaksesnya diperlukan jaringan internet atau *website* juga yang dikenal dengan web, site, situs, atau situs web, untuk lebih mengenal tentang jenis-jenis *website* dapat dikategorikan berdasarkan penggunaan maupun berdasarkan perusahaan yang menggunakannya seperti website untuk perusahaan yang dikenal dengan istilah *Company Profile* atau *website* untuk keperluan lain yang dikenal dengan istilah toko *online*, *blog*, *portal*, *search engine* dan masih banyak lagi. Khusus dalam jurnal ini *website* yang akan diteliti adalah *website* dari perusahaan (*company profile*) yang berarti adalah sebuah *website* yang berisikan tentang informasi tentang perusahaan tersebut hal ini wajib ada atau dimiliki oleh sebuah perusahaan jika ingin berkembang ditengah gelombang pesatnya arus informasi saat ini, dengan *website* ini perusahaan bisa memberikan informasi kepada pihak lain yang membutuhkan.

Untuk melihat sebuah *website* yang baik maka diperlukan indikator-indikator tertentu yang menjadi tolak ukurnya dalam hal ini baik buruknya sebuah *website*, sebuah *website* menurut sudut pandang berbagai user berbeda-beda tergantung kepada kebutuhan user terhadap *website* tersebut, ada yang lebih mengutamakan konten, ada juga yang lebih mengutamakan tampilan, ada juga yang lebih mengutamakan kecepatan akses dari *website* tersebut serta masih banyak lagi tergantung kebutuhan user, untuk melihat hal tersebut maka dalam jurnal ini akan diteliti sebuah *website* dalam hal ini adalah *website* dari PT BRIGHT Batam dan akan dipilih tujuh indikator sebagai pendukung dari *website* tersebut, dari tujuh indikator tersebut akan dilihat indikator mana yang menjadi prioritas utama bagi user dalam melihat baik buruknya sebuah *website*. Artinya satu buah *website* bisa dilihat manakah dari indikator dari yang tujuh tersebut dapat dijadikan sebagai faktor penentu tentang baik dan buruk nya sebuah *website*.

2. TINJAUAN PUSTAKA

Berikut adalah uraian dari teori yang digunakan dalam jurnal ini, teori ini dimaksudkan sebagai acuan jurnal yang dibuat hal ini dilakukan supaya keabsahan dari jurnal bisa dipertanggung jawabkan.

2.1. Knowledge Discovery in Database (KDD)

KDD (*Knowledge Discovery in Database*) merupakan proses terorganisir untuk mengidentifikasi pola dalam data yang besar dan kompleks di mana pola data tersebut ditemukan yang bersifat sah, baru, dan dapat bermanfaat serta dapat dimengerti [3]. Menurut pendapat [4] *Data Mining*, juga dikenal sebagai *Knowledge Discovery in Database* yang mengacu penggalian atau penambangan pengetahuan dari data dalam jumlah besar Jadi dari kedua pendapat di atas dapat disimpulkan bahwa KDD adalah proses pencarian informasi atau pengetahuan dari kumpulan data yang besar (*database*).

Menurut Feelders, Daniels, dan Holsheimer dalam [1] *Data Mining* adalah proses ekstraksi informasi dari kumpulan data melalui penggunaan algoritma dan teknik yang melibatkan bidang teknik statistik, matematika, mesin pembelajaran, dan sistem manajemen *database*. *Data Mining* dapat digunakan untuk mengekstrak informasi dan pengetahuan penting yang

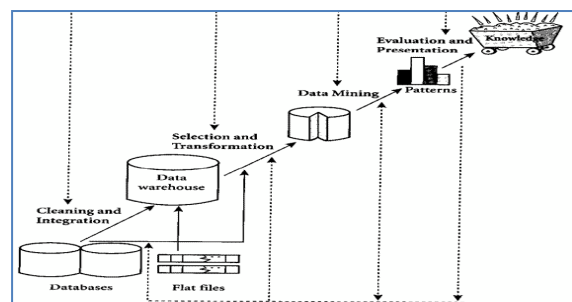
tersembunyi dari *dataset* yang besar. Dengan adanya *Data Mining*, maka akan dimungkinkan untuk mendapatkan harta berharga berupa pengetahuan didalam kumpulan data. Hampir senada dengan [1] mendefinisikan *Data Mining* sebagai kegiatan menemukan pola yang menarik dari data yang berjumlah besar, data dapat disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya. Pola yang ditemukan merupakan pola yang menarik, berharga dan penting dalam memberikan keuntungan kepada pemilik data. Terdapat yang mirip diungkapkan juga oleh [2] bahwa *Data mining* adalah proses yang menggunakan berbagai alat analisis data untuk menemukan pola dan hubungan dalam data yang dapat digunakan untuk membuat prediksi yang *valid*.

Dari berbagai pendapat di atas dapat disimpulkan bahwa *data mining* adalah serangkaian proses dalam menemukan pengetahuan tersembunyi dan berharga dari sekumpulan data yang dapat digunakan untuk membuat prediksi yang *valid* pada berbagai bidang, baik sosial, ekonomi, pendidikan, budaya bahkan sampai bidang pemerintahan. Menurut Feelders, Daniels, dan Holsheimer dalam [1] *Data Mining* adalah proses ekstraksi informasi dari kumpulan data melalui penggunaan algoritma dan teknik yang melibatkan bidang teknik statistik, matematika, mesin pembelajaran, dan sistem manajemen *database*. *Data Mining* dapat digunakan untuk mengekstrak informasi dan pengetahuan penting yang tersembunyi dari *data set* yang besar.

Dengan adanya *Data Mining*, maka akan dimungkinkan untuk mendapatkan harta berharga berupa pengetahuan didalam kumpulan data. Hampir senada dengan [1] mendefinisikan *Data Mining* sebagai kegiatan menemukan pola yang menarik dari data yang berjumlah besar, data dapat disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya. Pola yang ditemukan merupakan pola yang menarik, berharga dan penting dalam memberikan keuntungan kepada pemilik data. Terdapat yang mirip diungkapkan juga oleh [2] bahwa *Data mining* adalah proses yang menggunakan berbagai alat analisis data untuk menemukan pola dan hubungan dalam data yang dapat digunakan untuk membuat prediksi yang *valid*. Dari berbagai pendapat di atas dapat disimpulkan bahwa *data mining* adalah serangkaian proses dalam menemukan pengetahuan tersembunyi dan berharga dari sekumpulan data yang dapat digunakan untuk membuat prediksi yang *valid* pada berbagai bidang, baik sosial, ekonomi, pendidikan, budaya bahkan sampai bidang pemerintahan.

2.2 Tahapan Data Mining

Karena *Data Mining* merupakan rangkaian proses, maka *Data Mining* dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat *interaktif* di mana pemakai terlibat langsung atau dengan perantaraan *knowledge base*. Tahap-tahap ini diilustrasikan pada Gambar 1.



Gambar 1. Tahap-Tahap Data Mining

Tahap-tahap tersebut yaitu:

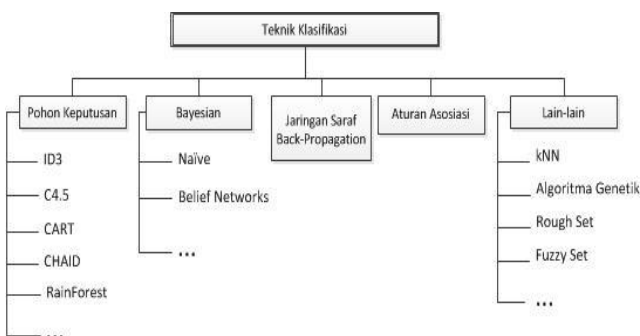
1. Pembersihan data (untuk membuang data yang tidak konsisten dan *noise*)
2. Integrasi data (penggabungan data dari beberapa sumber)
3. Transformasi data (data diubah menjadi bentuk yang sesuai untuk di-*mining*)
4. Aplikasi teknik *Data Mining*
5. Evaluasi pola yang ditemukan (untuk menemukan yang menarik/bernilai)
6. Presentasi pengetahuan (dengan teknik visualisasi).

2.3. Pengelompokan *Data Mining*

Pengelompokan *Data Mining* dibagi menjadi beberapa kelompok, menurut [5] dalam [6] yaitu:

1. Deskripsi
Deskripsi merupakan cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data yang dimiliki.
2. Estimasi
Estimasi hampir sama dengan klasifikasi, kecuali *variable* target estimasi lebih ke arah numerik daripada ke arah kategori. Model yang dibangun menggunakan *record* lengkap yang menyediakan nilai *variable* target sebagai nilai prediksi.
3. Prediksi
Prediksi menerka sebuah nilai yang belum diketahui dan juga memperkirakan nilai untuk masa mendatang.
4. Klasifikasi, dalam klasifikasi terdapat target *variable* kategori, misal penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu tinggi, sedang, dan rendah.
5. Pengklasteran, merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.
6. Asosiasi merupakan bertugas menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

Teknik klasifikasi adalah pendekatan sistematis untuk membangun model klasifikasi dari kumpulan data masukan. Misalnya, teknik pohon keputusan, *Bayesian (Naive Bayesian dan Bayesian Belief Networks)*, Jaringan Saraf Tiruan (*Backpropagation*), teknik yang berbasis konsep dari penambangan aturan-aturan asosiasi, dan teknik lain (*K-Nearest Neighbor*, algoritma genetik, teknik dengan pendekatan himpunan *rough* dan *fuzzy*). Setiap teknik mengadopsi algoritma pembelajaran untuk mengidentifikasi model yang paling sesuai dengan hubungan antara label atribut set dan kelas input data [8]. Setiap teknik juga memiliki kelebihan dan kekurangannya sendiri. Data dengan profil tertentu mungkin paling optimal jika diklasifikasi dengan teknik tertentu, atau dengan kata lain, profil data tertentu dapat mendukung termanfaatkannya kelebihan dari teknik ini.



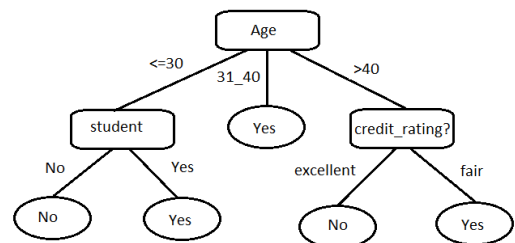
Gambar 2. Pengelompokan Teknik Klasifikasi

Secara umum, proses klasifikasi dapat dilakukan dalam dua tahap, yaitu proses belajar dari data pelatihan dan klasifikasi kasus. Pada proses belajar, Algoritma klasifikasi mengolah data *training* untuk menghasilkan sebuah model. Setelah model diuji dan dapat diterima, pada tahap klasifikasi, model tersebut digunakan untuk memprediksi kelas dari kasus baru untuk membantu proses pengambilan keputusan.

2.4. Pohon Keputusan (*Decision Tree*)

Metode ini merupakan salah satu metode yang ada pada teknik klasifikasi dalam *Data Mining*. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti *Structured Query Language* untuk mencari *record* pada kategori tertentu [9]. Pohon keputusan menggunakan representasi struktur pohon (*tree*) di mana setiap *node* merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut dan daun merepresentasikan kelas. *Node* yang paling atas dari pohon keputusan disebut sebagai *root*. Pohon keputusan merupakan metode klasifikasi yang paling populer digunakan. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun mudah untuk dipahami[5].

Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan *record*. Atribut menyatakan suatu parameter yang disebut sebagai kriteria dalam pembentukan pohon. Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk mem-*break down* proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan[8]. Contoh dari pohon keputusan dapat dilihat di Gambar 3 yang merupakan identifikasi pembeli komputer, dari pohon keputusan tersebut diketahui bahwa salah satu kelompok yang potensial membeli komputer adalah orang yang berusia di bawah 30 tahun dan juga pelajar [5].



Gambar 3. Pohon Keputusan Identifikasi

Dengan pohon keputusan, dapat dengan mudah mengidentifikasi dan melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah dan dapat mencari penyelesaian terbaik dengan memperhitungkan faktor-faktor tersebut. Pohon keputusan ini juga dapat menganalisa nilai resiko dan nilai suatu informasi yang terdapat dalam suatu alternatif pemecahan masalah. Peranan pohon keputusan sebagai alat bantu dalam mengambil keputusan (*decision support tool*) telah dikembangkan oleh manusia sejak perkembangan teori pohon yang dilandaskan pada teori graf. Kegunaan pohon keputusan yang sangat banyak ini membuatnya telah dimanfaatkan oleh manusia dalam berbagai macam sistem pengambilan keputusan. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, C4.5, CART.

2.5. Algoritma C4.5

Algoritma C4.5 yaitu sebuah algoritma yang digunakan untuk membangun *decision tree* (pengambilan keputusan). Algoritma C.45 adalah salah satu algoritma induksi pohon keputusan yaitu ID3 (*Iterative Dichotomiser 3*). ID3 dikembangkan oleh J. Ross Quinlan. Dalam prosedur algoritma ID3, input berupa sampel *training*, label *training* dan atribut. Algoritma C4.5 merupakan pengembangan dari ID3. Beberapa pengembangan yang dilakukan pada C4.5 adalah sebagai antara lain bisa mengatasi *missing value*, bisa mengatasi *continu data*, dan *pruning* [6]. Menurut [9] tahap pembelajaran algoritma C4.5 memiliki 2 prinsip kerja yaitu:

1. Pembuatan pohon keputusan. Tujuan dari algoritma penginduksi pohon keputusan adalah mengkontruksi struktur data pohon yang dapat digunakan untuk memprediksi kelas dari sebuah kasus atau *record* baru yang belum memiliki kelas. C4.5 melakukan konstruksi pohon keputusan dengan metode *divide and conquer*. Pada awalnya hanya dibuat *node* akar dengan menerapkan algoritma *divide and conquer*. Algoritma ini memilih pemecahan kasus-kasus yang terbaik dengan menghitung dan membandingkan *gain ratio*, kemudian *node-node* yang terbentuk di *level* berikutnya, algoritma *divide and conquer* akan diterapkan lagi sampai terbentuk daun-daun.
2. Pembuatan aturan-aturan (*rule set*). Aturan-aturan yang terbentuk dari pohon keputusan akan membentuk suatu kondisi dalam bentuk *if-then*. Aturan aturan ini didapat dengan cara menelusuri pohon keputusan dari akar sampai daun. Setiap *node* dan syarat percabangan akan membentuk suatu kondisi atau suatu *if*, sedangkan untuk nilai-nilai yang terdapat pada daun akan membentuk suatu hasil atau suatu *then*.

2.6. Pembentukan Pohon Keputusan Algoritma C4.5

Menurut[7] dalam [10] algoritma pembentukan pohon keputusan menggunakan sampel pelatihan yang terdiri atas kumpulan kasus di mana setiap kasus memiliki atribut dan kelas. Atribut-atribut dalam kasus dapat bertipe kontinyu maupun diskret tetapi pada kelas setiap kasus harus bertipe diskret yang dapat dinotasikan *CI*.,*C*. Pohon keputusan adalah pohon data yang terdiri atas simpul keputusan dan simpul daun yang terstruktur. Simpul daun tersebut memiliki kelas-kelas dan simpul keputusan menguji beberapa atribut sampai diperoleh atribut yang terpilih sebagai pemilah. Setiap cabang pada pengujian tersebut menghasilkan simpul anak (*Child Node*) di bawah simpul induknya (*Parent Node*).

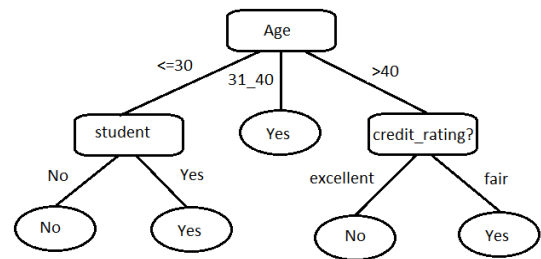
Metode ini merupakan salah satu metode yang ada pada teknik klasifikasi dalam *Data Mining*. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti *Structured Query Language* untuk mencari *record* pada kategori tertentu [9] Pohon keputusan menggunakan representasi struktur pohon (*tree*) di mana setiap *node* merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut dan daun merepresentasikan kelas. *Node* yang paling atas dari pohon keputusan disebut sebagai *root*. Pohon keputusan merupakan metode klasifikasi yang paling populer digunakan. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun mudah untuk dipahami.

Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan *record*. Atribut menyatakan suatu parameter yang disebut sebagai kriteria dalam pembentukan pohon. Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk *mem-break down* proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga

pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan. Misalkan untuk menentukan main tenis, kriteria yang diperhatikan adalah cuaca, angin, dan suhu. Salah satu atribut merupakan atribut yang menyatakan data solusi per *item* data yang disebut atribut hasil. Pada pohon keputusan terdapat 3 jenis *node*, yaitu:

- a. *Root Node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.
- b. *Internal Node*, merupakan *node* percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.
- c. *Leaf node atau terminal node*, merupakan *node* akhir, pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*

Contoh dari pohon keputusan dapat dilihat di Gambar 2.4 yang merupakan identifikasi pembeli komputer, dari pohon keputusan tersebut diketahui bahwa salah satu kelompok yang potensial membeli komputer adalah orang yang berusia di bawah 30 tahun dan juga pelajar[2]



Gambar 4. Pohon Keputusan Identifikasi Pembeli Komputer

Dengan pohon keputusan, dapat dengan mudah mengidentifikasi dan melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah dan dapat mencari penyelesaian terbaik dengan memperhitungkan faktor-faktor tersebut. Pohon keputusan ini juga dapat menganalisa nilai resiko dan nilai suatu informasi yang terdapat dalam suatu alternatif pemecahan masalah. Peranan pohon keputusan sebagai alat bantu dalam mengambil keputusan (*decision support tool*) telah dikembangkan oleh manusia sejak perkembangan teori pohon yang dilandaskan pada teori graf. Kegunaan pohon keputusan yang sangat banyak ini membuatnya telah dimanfaatkan oleh manusia dalam berbagai macam sistem pengambilan keputusan. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, C4.5, CART.

2.7. Algoritma C4.5

Algoritma C4.5 yaitu sebuah algoritma yang digunakan untuk membangun *decision tree* (pengambilan keputusan). Algoritma C.45 adalah salah satu algoritma induksi pohon keputusan yaitu ID3 (*Iterative Dichotomiser 3*). ID3 dikembangkan oleh J. Ross Quinlan. Dalam prosedur algoritma ID3, input berupa sampel *training*, label *training* dan atribut. Algoritma C4.5 merupakan pengembangan dari ID3. Beberapa pengembangan yang dilakukan pada C4.5 adalah sebagai antara lain bisa mengatasi *missing value*, bisa mengatasi *continu data*, dan *pruning* Menurut [8] tahap pembelajaran algoritma C4.5 memiliki 2 prinsip kerja yaitu:

1. Pembuatan pohon keputusan. Tujuan dari algoritma penginduksi pohon keputusan adalah mengkontruksi struktur data pohon yang dapat digunakan untuk memprediksi kelas dari sebuah kasus atau *record* baru yang belum memiliki kelas. C4.5 melakukan konstruksi pohon keputusan dengan metode *divide and conquer*. Pada

awalnya hanya dibuat *node* akar dengan menerapkan algoritma *divide and conquer*. Algoritma ini memilih pemecahan kasus-kasus yang terbaik dengan menghitung dan membandingkan *gain ratio*, kemudian *node-node* yang terbentuk di *level* berikutnya, algoritma *divide and conquer* akan diterapkan lagi sampai terbentuk daun-daun.

2. Pembuatan aturan-aturan (*rule set*). Aturan-aturan yang terbentuk dari pohon keputusan akan membentuk suatu kondisi dalam bentuk *if-then*. Aturan aturan ini didapat dengan cara menelusuri pohon keputusan dari akar sampai daun. Setiap *node* dan syarat percabangan akan membentuk suatu kondisi atau suatu *if*, sedangkan untuk nilai-nilai yang terdapat pada daun akan membentuk suatu hasil atau suatu *then*.

2.8. Pembentukan Pohon Keputusan Algoritma C4.5

Menurut [3] algoritma pembentukan pohon keputusan menggunakan sampel pelatihan yang terdiri atas kumpulan kasus di mana setiap kasus memiliki atribut dan kelas. Atribut-atribut dalam kasus dapat bertipe kontinu maupun diskret tetapi pada kelas setiap kasus harus bertipe diskret yang dapat dinotasikan C_1, \dots, C_n . Pohon keputusan adalah pohon data yang terdiri atas simpul keputusan dan simpul daun yang terstruktur. Simpul daun tersebut memiliki kelas-kelas dan simpul keputusan menguji beberapa atribut sampai diperoleh atribut yang terpilih sebagai pemilah. Setiap cabang pada pengujian tersebut menghasilkan simpul anak (*Child Node*) di bawah simpul induknya (*Parent Node*). Pembentukan pohon keputusan menggunakan prinsip membagi sampel pelatihan menjadi beberapa sub-himpunan yang berbeda. Langkah awal dari konstruksi pohon keputusan algoritma C4.5 adalah mencari simpul akar. Berikut ini langkah-langkah konstruksi pohon keputusan menggunakan Algoritma C4.5 berdasarkan [3] dalam [6]:

1. Misalkan T adalah himpunan kasus-kasus yang akan dibuat simpul di mana kasus-kasus tersebut memiliki kelas dan atribut-atribut. Frekuensi terboboti $freq(C_j, T)$ diperoleh dari perhitungan T dan kelas yang dihasilkan adalah C_j , untuk setiap $j \in \{1, 2, 3, \dots, n\}$
2. Jika semua kasus berada dalam kelas C yang sama maka simpul yang dihasilkan adalah simpul daun yang diberi label dengan kelas C_j sebagai kelas terbanyak. Kesalahan klasifikasi pada simpul daun merupakan kasus-kasus dalam T yang berbeda kelas dengan kelas C_j .
3. Jika T memiliki kasus yang memiliki dua atau lebih kelas maka dapat dihitung j *information gain* dari setiap atribut tersebut. Untuk atribut diskret, *information gain* disesuaikan dengan pembagi dalam T dengan nilai atribut yang sudah diketahui sebelumnya. Untuk atribut kontinu, *information gain* disesuaikan dengan pembagi T ke dalam dua irisan (biner) yang dilabeli kasus dengan nilai atribut kurang dari atau sama dengan nilai ambang batas ($A \leq v$) dan nilai atribut dengan nilai atribut lebih besar dari nilai ambang batas ($A > v$).
4. Atribut dengan nilai *information gain* tertinggi terpilih sebagai pemilah dalam simpul tersebut.
5. Simpul keputusan memiliki cabang sebanyak s yaitu T_1, \dots, T_s di mana $s = 2$ untuk atribut kontinu dan $s = h$ untuk atribut diskret dengan nilai h yang sudah diketahui.
6. Untuk setiap $i = \{1, 2, \dots, s\}$, jika T_i tidak memiliki cabang lagi maka simpul tersebut secara langsung menjadi simpul daun yang diberi label kelas terbanyak di bawah simpul induknya dan kesalahan klasifikasi bernilai 0.
7. Apabila T_i memiliki cabang lagi maka pemilahan diproses kembali menggunakan kasus-kasus dalam T_i . Catatan khusus untuk kasus-kasus dengan nilai yang hilang pada atribut terpilih tersebut dilakukan proses pemilihan pemilah pada setiap simpul anaknya dengan pembobotan banyak

kasus yang diketahui dibagi dengan banyak kasus pada simpul tersebut.

8. Terakhir, kesalahan klasifikasi simpul dihitung dari penjumlahan dari kesalahan-kesalahan simpul anak yang dibandingkan dengan simpul induknya.

2.9. Entropy dan Information Gain

Sebuah obyek yang diklasifikasikan dalam pohon harus dities nilai *Entropy* -nya. *Entropy* adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari *impurity* dan *homogeneity* dari kumpulan data. Dari nilai *Entropy* tersebut kemudian dihitung nilai *information gain* (IG) masing-masing atribut. *Entropy* (S) merupakan jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel S . *Entropy* dapat dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. Semakin kecil nilai *Entropy* maka akan semakin *Entropy* digunakan dalam mengekstrak suatu kelas. *Entropy* digunakan untuk mengukur ketidakefektifan S . Menurut [6] besarnya *Entropy* pada ruang sampel S didefinisikan dengan:

$$Entropy(S) \equiv -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}$$

Di mana: S ruang (data) sampel yang digunakan untuk pelatihan

p_{\oplus} : jumlah yang bersolusi positif atau mendukung pada data sampel untuk kriteria tertentu.

p_{\ominus} : jumlah yang bersolusi negatif atau tidak mendukung pada data sampel untuk kriteria tertentu.

Information gain adalah salah satu *attribute selection measure* yang digunakan untuk memilih *test attribute* tiap *node* pada *tree*. Atribut dengan informasi *gain* tertinggi dipilih sebagai *test attribute* dari suatu *node* [2].

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

(1)Gain (S,A)

merupakan perolehan informasi dari atribut A *relative* terhadap *output* data S . Perolehan informasi didapat dari *output* data atau *variable dependent* S yang dikelompokkan berdasarkan atribut A , dinotasikan dengan $gain(S, A)$. Di mana:

S : himpunan kasus

A : atribut

N : jumlah partisi atribut A

$[S_i]$: jumlah kasus pada partisi ke- i

$[S]$: jumlah kasus pada S

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut[2]:

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

3. METODOLOGI

Metode penelitian dilakukan dengan studi literature terhadap sumber-sumber yang relevan dalam hal ini berkaitan dengan buku-buku dan jurnal yang relevan dan juga termasuk kepada pelanggan dari PT BRIGHT itu sendiri yang sering menggunakan website untuk keperluannya. Analisis pengetahuan terhadap indikator-indikator untuk menilai sebuah website menggunakan algoritma C4.5. pada dasarnya banyak lagi algoritma yang dapat dipakai dalam pembentukan pohon keputusan seperti ID3 yang ditemukan oleh J. Ross Quinlan tahun 1979 yang dikenal dengan algoritma yang digunakan untuk membangun sebuah keputusan (decision tree atau pohon

keputusan), Chart dan Agoritma C4.5. Untuk algoritma C4.5 ini sendiri adalah sebuah algoritma yang didapat dengan pengembangan dari algoritma ID3. Pada dasarnya algoritma ini menggunakan struktur hirarki untuk pembelajaran supervised, dimana proses akan dimulai dari root node hingga leaf node yang dilakukan secara rekursif, dimana setiap percabangan menyatakan suatu kondisi yang harus dipenuhi dan pada setiap ujung pohon menyatakan kelas dari suatu data. Lebih lanjut model Algoritma ID3 atau C4.5 mengubah bentuk data dalam bentuk tabel menjadi pohon (tree) kemudian mengubah model pohon tersebut menjadi aturan (rule).

Pohon keputusan merupakan metode dan prediksi yang sangat kuat dan terkenal, dimana metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang memprediksi aturan. Aturan dapat dengan mudah dipahami dengan alami dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti structured query language untuk mencari record pada kategori tertentu. Pohon keputusan tersebut juga memperlihatkan factor-faktor kemungkinan atau probabilitas yang akan mempengaruhi alternatif-alternatif keputusan tersebut, disertai dengan estimasi hasil akhir yang akan didapat bila kita mengambil alternatif keputusan tersebut. proses pada pohon keputusan tersebut adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi rule dan menyederhanakan rule. Secara umum algoritma C4.5 untuk membangun keputusan adalah sebagai berikut:

1. Pilih atribut sebagai akar
2. Buat cabang untuk tiap nilai
3. Bagi kasus dalam cabang
4. Ulang proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama

Untuk memilih atribut sebagai akar, didasarkan kepada nilai gain tertinggi dari atribut-atribut yang ada, untuk menghitung nilai gain tersebut dapat dirumuskan seperti dibawah ini:

$$= Entropy(S) - \sum_{i=1}^n \frac{|Status_i|}{|Total|} * Entropy(Status_i)$$

Dimana:

S : himpunan kasus

A : atribut

N : jumlah partisi atribut A

$[S_i]$: jumlah kasus pada partisi ke- i

$[S]$: jumlah kasus pada S

Selanjutnya untuk perhitungan nilai entropy Menurut Aprilla (2013) besarnya $Entropy$ pada ruang sampel S didefinisikan dengan adalah seperti 2 (dua) persamaan dibawah ini:

$$Entropy(S) \equiv -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus}$$

4. HASIL DAN PEMBAHASAN

Berikut adalah uraian tentang hasil dan pembahasan yang antara lain adalah sebagai berikut:

4.1. Analisis Data

Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal dimana metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang sangat merepresentasikan aturan. Aturan dengan mudah dapat dipahami dengan Bahasa alami. Dan mereka juga dapat

diekspresikan dalam bentuk bahasa basis data seperti *structured query language* (SQL) untuk mencari *record* pada kategori tertentu. Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan *record*. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria pembentukan pohon, salah satu atribut merupakan atribut yang menyatakan solusi per item data yang disebut target atribut. Atribut memiliki nilai-nilai yang dinamakan dengan instance.

4.2. Pemilihan Variabel

Dari data-data yang telah diperoleh, maka akan ditentukan suatu variabel yang menjadi variabel keputusan dalam pemilihan indikator website yang baik. Artinya dalam sebuah website yang baik akan didukung oleh beberapa indikator dalam ha ini adalah tujuh indikator, dari tujuh indikator ini akan dicari lagi mana indikator yang paling berpengaruh dominan terhadap website tersebut. Untuk selanjutnya indikator disini akan disebut variabel maka diketahui beberapa fariabel yang menjadi penentu dalam pemilihan indikator website yang baik diantaranya adalah fariabel usability (mudah untuk dipelajari dan digunakan), sistem navigasi (mudah dipahami oleh pengunjung secara keseluruhan), graphic design (pemilihan grafis, layout, warna yang menarik, isi konten (manfaat isi dari website), kompatibilitas (perangkat pendukung website), loading (waktu panggil), functionality (dinamis dan interaktif sebuah website).

Tabel 1. Data Hasil Pra-Proses

Usability	Navigasi	Graphic design	Content	Kompatibilitas	Loading	Functionality	Keputusan
Sm	M	Ss	Sb	Sl	Sc	Sb	B
Sm	M	S	Bm	L	Cp	B	B
Tm	Tm	S	Bm	Tl	Cp	B	J
Sm	Sm	S	Bm	L	Cp	B	J
Sm	Sm	Ts	Sb	Tl	Sc	B	B
Tm	Sm	Ss	Sb	Tl	Sc	Jl	B
M	Tm	Ss	Bm	Sl	Sc	Jl	B
M	Tm	Ss	Bm	L	Lb	Sb	J
Tm	M	Ss	Sb	Sl	Lb	Sb	B
Tm	M	Ts	Sb	Sl	Lb	Sb	J
M	M	Ts	Bm	L	Sc	Sb	J
Tm	M	Ts	Bm	Tl	Sc	Sb	J
Sm	Tm	Ss	Bm	L	Sc	Sb	J
Sm	Tm	Ss	Tb	L	Cp	Jl	B
Sm	Tm	Ss	Tb	Sl	Cp	B	J
M	Sm	S	Bm	Sl	Sc	B	B
M	Sm	S	Bm	L	Cp	B	B
Tm	M	S	Sb	L	Cp	B	B
Tm	Tm	Ts	Sb	Sl	Cp	Jl	J
M	Sm	S	Tb	Sl	Lb	B	B

Sumber: Data Penulis 2018.

Adapun pra-proses yang dilakukan dalam mempertimbangkan faktor diatas diambil berdasarkan sampel data pengamatan dilapangan.

1. Pengelompokkan Usability

S : Ruang (data) sampel yang digunakan untuk pelatihan

p_{\oplus} : Jumlah yang bersolusi negatif atau tidak mendukung pada data sampel untuk kriteria tertentu.

p_{\ominus} : Jumlah yang bersolusi positif atau mendukung pada data sampel untuk.

Tabel 2. Klasifikasi Usability

Usability	Klasifikasi
Sangat Mudah	SM
Mudah	M
Tidak Mudah	TM

2. Pengelompokkan Navigasi

Tabel 3. Klasifikasi Navigasi

Navigasi	Klasifikasi
Sangat Menarik	SM
Menarik	M
Tidak Menarik	TM

3. Pengelompokkan Graphic Design

Tabel 4. Klasifikasi Graphic design

Graphic Design	Klasifikasi
Sangat Sesuai	SS
Sesuai	S
Tidak Sesuai	TS

4. Pengelompokkan Isi Atau Konten

Tabel 5. Klasifikasi Isi atau Konten

Isi atau Konten	Klasifikasi
Sangat Bermamfaat	SB
Bermamfaat	BM
Tidak Bermamfaat	TB

5. Pengelompokkan Kompatibilitas

Tabel 6. Klasifikasi Kompatibilitas

Navigasi	Klasifikasi
Sangat Luas	SL
Luas	L
Tidak Luas	TL

6. Pengelompokkan Loading

Tabel 7. Klasifikasi Loading

Loading	Klasifikasi
Sangat Cepat	Sc
Cepat	C
Lambat	LB

7. Pengelompokkan Functionality

Tabel 8. Klasifikasi Functionality

Functionality	Klasifikasi
Sangat Baik	SB
Baik	B
Jelek	JL

4.3. Perhitungan Entropy dan Gain.

$$Entropy(total) = \left(-\frac{11}{20} * \log_2\left(\frac{11}{20}\right)\right) + \left(-\frac{9}{20} * \log_2\left(\frac{9}{20}\right)\right) = 0,99277$$

$$Entropy Usability SM = \left(-\frac{4}{7} * \log_2\left(\frac{4}{7}\right)\right) + \left(-\frac{3}{7} * \log_2\left(\frac{3}{7}\right)\right) = 0,98522$$

$$Entropy Usability M = \left(-\frac{4}{6} * \log_2\left(\frac{4}{6}\right)\right) + \left(-\frac{2}{6} * \log_2\left(\frac{2}{6}\right)\right) = 0,91829$$

$$Entropy Usability TM = \left(-\frac{3}{7} * \log_2\left(\frac{3}{7}\right)\right) + \left(-\frac{4}{7} * \log_2\left(\frac{4}{7}\right)\right) = 0,98522$$

Gain Total Usability

$$0,99277 - \left(\left(\frac{7}{20} * 0,9852281\right) + \left(\frac{6}{20} * 0,9182\right) + \left(\frac{7}{20} * 1\right)\right) = 0,02762$$

$$Entropy Navigasi SM = \left(-\frac{5}{6} * \log_2\left(\frac{5}{6}\right)\right) + \left(-\frac{1}{6} * \log_2\left(\frac{1}{6}\right)\right) = 0,65002$$

$$Entropy Navigasi M = \left(-\frac{4}{7} * \log_2\left(\frac{4}{7}\right)\right) + \left(-\frac{3}{7} * \log_2\left(\frac{3}{7}\right)\right) = 0,98522$$

$$Entropy Navigasi TM = \left(-\frac{2}{7} * \log_2\left(\frac{2}{7}\right)\right) + \left(-\frac{5}{7} * \log_2\left(\frac{5}{7}\right)\right) = 0,86312$$

Gain Total Navigasi

$$0,99277 - \left(\left(\frac{6}{20} * 0,65002\right) + \left(\frac{7}{20} * 0,98522\right) + \left(\frac{7}{20} * 0,86312\right)\right) = 0,15084$$

$$Entropy Graphic SS = \left(-\frac{6}{8} * \log_2\left(\frac{6}{8}\right)\right) + \left(-\frac{2}{8} * \log_2\left(\frac{2}{8}\right)\right) = 0,954434$$

$$Entropy Graphic S = \left(-\frac{5}{7} * \log_2\left(\frac{5}{7}\right)\right) + \left(-\frac{2}{7} * \log_2\left(\frac{2}{7}\right)\right) = 0,86312$$

$$Entropy graphic TS = \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right) + \left(-\frac{4}{5} * \log_2\left(\frac{4}{5}\right)\right) = 0,72192$$

Gain Total Graphic

$$0,99277 - \left(\left(\frac{8}{20} * 0,81127\right) + \left(\frac{7}{20} * 0,86312\right) + \left(\frac{5}{20} * 0,72193\right)\right) = 0,128426$$

$$Entropy Conten SB = \left(-\frac{5}{7} * \log_2\left(\frac{5}{7}\right)\right) + \left(-\frac{2}{7} * \log_2\left(\frac{2}{7}\right)\right) = 0,86312$$

$$Entropy Comtems BM = \left(-\frac{4}{10} * \log_2\left(\frac{4}{10}\right)\right) + \left(-\frac{6}{10} * \log_2\left(\frac{6}{10}\right)\right) = 0,97095$$

$$Entropy Conten TB = \left(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right)\right) + \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) = 0,91829$$

Gain Total Contents

$$0,99277 - \left(\left(\frac{7}{20} * 0,86312\right) + \left(\frac{10}{20} * 0,97095\right) + \left(\frac{5}{20} * 0,91829\right)\right) = 0,06746.$$

$$Entropy Kompatibilitas SL = \left(-\frac{5}{8} * \log_2\left(\frac{5}{8}\right)\right) + \left(-\frac{3}{8} * \log_2\left(\frac{3}{8}\right)\right) = 0,9544$$

$$Entropy Kompatibilitas L = \left(-\frac{4}{8} * \log_2\left(\frac{4}{8}\right)\right) + \left(-\frac{4}{8} * \log_2\left(\frac{4}{8}\right)\right) = 1$$

$$Entropy Kompatibilitas TL = \left(-\frac{2}{4} * \log_2\left(\frac{2}{4}\right)\right) + \left(-\frac{2}{4} * \log_2\left(\frac{2}{4}\right)\right) = 1$$

Gain Total Kompatibilitas

$$0,99277 - \left(\left(\frac{8}{20} * 0,86312\right) + \left(\frac{8}{20} * 1\right) + \left(\frac{4}{20} * 1\right)\right) = 0,011008$$

$$Entropy Loading SC = \left(-\frac{5}{8} * \log_2\left(\frac{5}{8}\right)\right) + \left(-\frac{3}{8} * \log_2\left(\frac{3}{8}\right)\right) = 0,9544$$

$$Entropy Loading L = \left(-\frac{4}{8} * \log_2\left(\frac{4}{8}\right)\right) + \left(-\frac{4}{8} * \log_2\left(\frac{4}{8}\right)\right) = 1$$

$$Entropy Loading L = \left(-\frac{2}{4} * \log_2\left(\frac{2}{4}\right)\right) + \left(-\frac{2}{4} * \log_2\left(\frac{2}{4}\right)\right) = 1$$

Gain Total Kompatibilitas

$$0,99277 - \left(\left(\frac{8}{20} * 0,86312\right) + \left(\frac{8}{20} * 1\right) + \left(\frac{4}{20} * 1\right)\right) = 0,011008$$

$$Entropy Functinal SB = \left(-\frac{2}{7} * \log_2\left(\frac{2}{7}\right)\right) + \left(-\frac{5}{7} * \log_2\left(\frac{5}{7}\right)\right) = 0,86312$$

$$Entropy Functional B = \left(-\frac{6}{9} * \log_2\left(\frac{6}{9}\right)\right) + \left(-\frac{3}{9} * \log_2\left(\frac{3}{9}\right)\right) = 0,918292$$

$$Entropy Functional JL = \left(-\frac{3}{4} * \log_2\left(\frac{3}{4}\right)\right) + \left(-\frac{1}{4} * \log_2\left(\frac{1}{4}\right)\right) = 0,81137$$

Gain Total Kompatibilitas

$$0,99277 - \left(\left(\frac{7}{20} * 0,86312\right) + \left(\frac{9}{20} * 0,918292\right) + \left(\frac{4}{20} * 0,81137\right)\right) = 0,11519$$

Tabel 9. Perhitungan Node.1

No	de	Jml Kasus	T	Y	Entropy	Gain
1	Total Usability	20	11	9	0,99277	0.02762
	sm	7	4	3	0.98522	0.15084
	m	6	4	2	0.91829	
	tm	7	3	4	0.98522	
	Navigation					0.128426
	sm	6	5	1	0.65002	
	m	7	4	3	0.98522	
	tm	7	2	5	0.86312	0.06746
	Graphic					
	ss	8	6	2	0.95443	
	s	7	5	2	0.86312	0.011008
	ts	5	1	4	0.72192	
	Content					
	sb	7	5	2	0.86312	0.011008
	bm	10	4	6	0.99277	
	tb	3	2	1	0.06746	
	Kompas					0.11519
	sl	8	5	3	0.95443	
	l	8	4	4	1	
	tl	4	2	2	1	0.011008
	Loading					
	sl	8	5	3	0.95443	
	l	8	4	4	1	0.11519
	tl	4	2	2	1	
	Functionality					
	sb	7	2	5	0.86312	0.11519
	b	9	6	3	0.91829	
	jl	4	3	1	0.81137	

Sumber: Data Penulis 20018

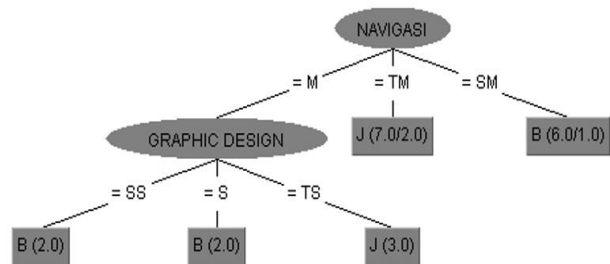
Berdasarkan hasil pada tabel 2.9 dapat diketahui bahwa atribut dengan Gain tertinggi adalah navigasi yaitu sebesar 0.15084 hal ini berarti bahwa Navigasi dapat menjadi node Akar. Selanjutnya jika dilihat dari nilai Navigasi masih ada tiga nilai lagi yaitu Navigasi sangat menarik (SM), menarik (M) dan tidak menarik (TM). Jadi masih perlu dilakukan proses perhitungan lagi.

4.4. Pembahasan

Pengujian terhadap Analisa, sangat penting dilakukan untuk menentukan dan memastikan apakah hasil Analisa tersebut sesuai dengan keputusan yang diharapkan. Untuk menguji kebenaran dari pengolahan data yang dilakukan secara manual, maka dapat menggunakan salah satu aplikasi WEKA 44.0.2 decision tree J.45 sebagai pengaplikasian dari Algoritma C.45.

4.5. Langkah-langkah Implementasi

Seluruh variabel yang terdiri dari atribut kondisi dan atribut keputusan yang digunakan untuk menentukan indikator-indikator yang tepat dalam penentuan sebuah website yang baik disimpan pada *Microsoft Excel* dengan nama file data web.xls kemudian disimpan dengan extension CSV atau *comma delimited* (CSV atau *.csv). kemudian data disesuaikan dengan menambahkan informasi awal dan data tersebut dapat dipergunakan sebagai inputan dalam WEKA 44.0.2. selanjutnya klik buttom choose, pilih J.48 dalam pembentukan pohon keputusan dan klik kana pada menu start, pada tahap ini proses data mining dilakukan dengan memilih algoritma yang dipakai dalam menghasilkan sebuah pohon keputusan Algoritma C.45. hasil pengujian dapat dilihat pada gambar dibawah ini:



Gambar 5. Hasil pengujian dengan Weka

5. SIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang dilakukan pada PT BRIGHT Batam tentang indikator dari *website* yang paling baik (mendukung) maka penulis dapat menarik kesimpulan bahwa hal ini jelas memberikan mamfaat bagi PT BRIGHT itu sendiri dalam proses pengembangan *Website* nya, karena dari 7 indikator yang dibahas untuk yang mempunyai kekurangan bisa diperbaiki atau diperbarui lagi, untuk lebih jelasnya dapat dijelaskan seperti dibawah ini:

1. Yang menjadi faktor tertinggi dari tujuh indikator yang dibahas dalam hal ini yang mempengaruhi baik atau tidak sebuah *website* adalah Navigasi. Hal ini dapat diartikan pelanggan merasa puas dengan dengan sistem navigasi karena mudah menemukan halaman-halaman *website* yang dibutuhkan dan semua informasi yang dibutuhkan mudah icari dan didapatkan dengan cepat dan tepat.
2. Faktor selanjutnya atau yang kedua dalam hal yang mempengaruhi dalam menilai sebuah indikator *website* baik atau tidak adalah *Graphic Design*, hal ini berarti pelanggan merasa senang dan puas dengan tampilan dari *website* yang ada baik segi unsur pewarnaan, grafis, layout maupun dari segi typografi yang ada.
3. Sedangkan kelima faktor yang lain tidak mempengaruhi baik atau buruknya sebuah *website*, jadi keliam faktor ini bisa diperahankan. karena pelanggan tidak terlalu mempermasalahkannya.

Selanjutnya penulis menyarankan agar hasil ini dapat dipertimbangkan dalam proses perbaikan *website* kedepannya dan juga hal-hal lain yang seyogyanya disesuaikan dengan kebijakan dari perusahaan yang berlaku.

DAFTAR PUSTAKA

- [1] Benni R Suburian. (2014). Aplikasi Data Mining Untuk Menampilkan Tingkat Kelulusan Mahasiswa Dengan Algoritma Apriori. *Teknik Informatika STMIK Budi Dharma Medan*.
- [2] Eki Roziqa Maris. (2017). Analisis Kepuasan Pelanggan Menggunakan Algoritma C4, Semarang. *Program Studi Teknik Informatika Universitas Nurswantoro Semarang*.
- [3] Haryanto, F. F. (2017). Penerapan Algoritma C4.5 Untuk Memprediksi Penerimaan Calon Pegawai Pada PT WISE. *Program Studi Teknik Informatika Universitas Multi Media Nusantara, Tangerang Banten*.
- [4] Iwandari. (2015). Model Keputusan Untuk Klasifikasi Persetujuan Kredit Menggunakan Algoritma C4.5. *Jurnal Litbang Kota Pekalongan*.
- [5] Jumanto. (2014). Implementasi Data Mining Algoritma C.45 Untuk Memprediksi Perilaku Mahasiswa Diploma 3 Melanjutkan Strata 1 Di STMIK AMIKOM Yogyakarta. *STMIK AMIKOM Yogyakarta*.
- [6] Larissa Maharani. (2015). Klasifikasi Nasabah Bank Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit. *Padang Fakultas Ilmu Komputer Universitas Putera Indonesia Padang*.
- [7] Liliana Swastina. (2013). Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa, Banjarmasin. *Program Teknik Informatika STMIK Indonesia Banjarmasin*.
- [8] Nurcahyo, G. W. (2008). Penerapan Data Mining dengan Algoritma Apriori untuk Mendukung Strategi Promosi Pendidikan. *Universitas Putra Indonesia "YPTK" Padang Sumatera Barat*.
- [9] Sukma Putri Utari. (2015). Implementasi Metode C.45 Untuk Menentukan Guru Terbaik Pada SMK 1 Percut Sei Tuan Medan. *Teknik Informatika STMIK Budi Dharma Medan*.

BIODATA PENULIS



Rika Harman

Dosen tetap Universitas Putera Batam aktif pada Fakultas Teknik jurusan sistem Informasi, untuk semester genap tahun ajaran 2017 - 2018 mengampu matakuliah Proyek Sistem Informasi