

**Perancangan data mining berbasis web menggunakan algoritma
C 4,5 untuk klasifikasi kecanduan sosial media facebook
(study kasus PT JATIM TAMAN STEEL)**

Henny tri wulandari
Mahasiswa Teknik Informatika Universitas Islam Majapahit
Hennytriwulandari18@gmail.com

ABSTRAK

Pencarian data kecanduan sosial media facebook yang diketahui mempengaruhi kinerja karyawan PT JATIM TAMAN STEEL, perhitungan kecanduan menggunakan algoritma C 4,5 karena disinyalir dapat melakukan perhitungan dengan sederhana dan akurat. Algoritma C 4,5 berjalan dengan cara menghitung entorpy dari macam-macam atribut yang akan menghasilkan gain. Aplikasi ini mnghasilkan informasi berupa pohon keputusan dan aturan-aturan (rule) yang berguna saat melakukan analisa kecanduan pada sosial media yang dibagi menjadi empat kecanduan yaitu polos (tidak kecanduan), awam (batas aman), depresi (harus hati-hati) dan akut (batas bahaya). Setiap keputusan yang dihasilkan adalah hasil keseimbangan dari nilai input dan rule yang dihasilkan dari proses mining kecanduan sosial media. Sedangkan atribut yang digunakan ada 7 atribut yaitu jenis_kelamin, umur, bagian, jawaban A, jawaban B, jawaban C dan jawaban D. Bahasa yang digunakan dalam proses pencarian perhitungan kecanduan adalah bahasa PHP (Hypertext Preprocessor). perancangan menggunakan metode perancangan UML (Unified Modeling Language). Tingkat akurasi dalam perhitungan ini mencapai 77,5% dari 100 data kasus yang didapatkan. Data testing ada 75 dan data training ada 25.

Kata kunci: Algoritma C 4.5, PHP, MySQL, Data Mining

ABSTRACT

Searching for Facebook social media addiction data that is known to affect the performance of PT JATIM TAMAN STEEL employees, addiction calculation uses the C 4.5 algorithm because it is alleged that it can perform calculations simply and accurately. Logarithm C 4.5 runs by calculating the entorpy of the various attributes will generate gain. This application produces information in the form of decision trees and rules that are useful when doing addiction analysis on social media which is divided into four addictions, namely plain (not addicted), lay (safe limit), depression (must be careful) and acute (danger limit). Each decision produced is the result of the balance of the input and rule values generated from the mining process of social media addiction. While the attributes that are used there are 7 attributes, namely type-gender, age, section, answer A, answer B, answer C and answer D. Language used in the process of finding addiction calculations is the language PHP (Hypertext Preprocessor). the design uses the UML (Unified Modeling Language) design method. The accuracy rate in this calculation reaches 77.5% of the 100 case data obtained. There is 75 testing data and 25 training data.

Keywords: Algorithm C 4.5, PHP, MySQL, Data Mining.

PENDAHULUAN

Latar Belakang

Pada jaman kemajuan saat ini, kebutuhan akan informasi yang akurat dan cepat digunakan dalam kinerja saat ini, sehingga informasi akan menjadi suatu elemen penting dalam perkembangan perusahaan menuju ke arah yang lebih baik. Namun kebutuhan akan informasi yang akurat dan cepat sering tidak cocok dengan penyuguhan informasi yang tanggap dan akurat, sering kali perusahaan harus mencari data dari sumber yang harus digali dengan memakan waktu yang cukup lama. Kemampuan menganalisis pada saat ini lebih mudah dibandingkan kemampuan teknologi dalam mengumpulkan dan menyimpan data yang lebih besar. Metode yang awalnya menggunakan analisa manual tidak dapat menjangkau data yang lebih besar.

Pemanfaatan data yang ada dalam sistem perusahaan tidak dapat memberikan masukan yang akurat dalam mengambil keputusan yang jauh dari kata baik. Tidak dibutuhkan data yang hanya operasional yang ada diperusahaan karena tidak dapat menunjang keberadaan data yang makin lama makin besar jumlahnya, para analisa data pengambil keputusan sering harus menggali data untuk informasi dari gudang data hal ini sangat tidak efisien dan menguras tenaga serta banyak waktu yang akan terbuang maka dari itu para analisa menemukan cara baru untuk menganalisa dan menggali data yang berjumlah besar dengan menggunakan metode data mining. Data mining dipercaya dapat melakukan akurasi data yang lebih akurat dan efisien dalam menggali data dalam jumlah besar.

Kecanduan sosial media saat ini mempengaruhi kinerja karyawan dan kebanyakan dari karyawan menggunakan *facebook* dalam bekerja sehingga sangat rawan kecelakaan karena terlalu fokus dengan beranda di *facebook*, banyak karyawan yang menyalagunakan media social *facebook*.

Penggunaan sosial media saat ini sangatlah memprihatinkan banyak diantara pengguna sosial media yang menomor satukan sosial media dibandingkan kebutuhan lainnya. Seolah handphone dan kuota adalah kebutuhan primer saat ini.

Disini akan dibahas tentang cara menganalisis kecanduan penggunaan sosial media pada masa modern saat ini untuk sosial media *facebook* di perusahaan PT JATIM TAMAN STEEL.

Klasifikasi digunakan untuk menentukan tingkat kecanduan sosial media yang di bagi menjadi 4 tingkat kecanduan yaitu polos, awam, depresi dan akut. Untuk mendapatkan *rule* atau aturan tersebut, dibutuhkan *algoritma* metode *datamining* yang akurat dan efisien dengan kasus kecanduan sosial media *facebook*. Berdasarkan paper yang dipresentasikan pada *International Conference on Data Mining (ICDM)*, IEEE pada tanggal 22 desember 2006, mengidentifikasi 10 *algoritma data mining* terbaik yaitu :

1. *Alogaritma C 4,5*
2. *K-Means*
3. *Support Vector Machines*
4. *Apriori*
5. *Expectation Maximisation Algorithm*
6. *PageRank*
7. *K-Nearest Neighbors*
8. *Naive Bayes*
9. *Classification and Regression trees*
10. *Adaboost*

Kelebihan *algoritma C 4,5* dibandingkan yang lain adalah mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan serta aturan dapat dipahami dengan bahasa alami dan *Algoritma C4.5* merupakan *algoritma klasifikasi pohon keputusan yang banyak digunakan karena memiliki kelebihan utama dari algoritma yang lainnya. Kelebihan algoritma C4.5 dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, efisien dalam menangani atribut bertipe*

diskret dan dapat menangani atribut bertipe diskret dan numerik.

Dengan melihat keunggulan suatu *algoritma* dalam mengolah dan *mengkonver* data dalam tingkat akurasi yang tepat dan cukup sederhana, terpilihlah *Algoritma C4.5* sebagai *algoritma* terbaik diantara 10 nominasi *algoritma data mining*.

Berdasarkan masalah yang disampaikan diatas. Sehingga penulis menjadikannya sebagai topik. Tugas Akhir dengan judul "PERACANGAN DAN PEMBATAN APLIKASI DATA MINING BERBASIS WEB MENGGUNAKAN ALGORITMA C4.5 UNTUK MENGANALISIS KECANDUAN SOSIAL MEDIA

Rumusan Masalah

Berdasarkan latar belakang tersebut maka perumusan masalah yang dapat diambil dalam penelitian ini adalah "Bagaimana menerapkan *Algoritma C4.5* untuk menganalisis kecanduan penggunaan sosial media saat ini terutama pada study kasus *facebook*?"

Tujuan Masalah

Tujuan yang ingin dicapai dalam penelitian ini adalah untuk menganalisis kecanduan seseorang terhadap sosial media yang dia hadapi saat ini. Sebab sosial media saat ini mulai merenggut kebebasan seseorang.

Manfaat Penelitian

Penelitian ini dimaksudkan untuk dapat analisis kecanduan penggunaan sosial media dengan menerapkan *algoritma C4.5*, sehingga seseorang bisa mengetahui apakah dia kecanduan sosial media atau tidak.

Manfaat untuk perusahaan sendiri adalah agar perusahaan dapat mengetahui apakah karyawan PT JATIM TAMAN STEEL mengalami kecanduan sosial media *facebook*.

Batasan Masalah

Pada penulisan tugas akhir ini, penulis membatasi permasalahan dalam aplikasi data *mining* untuk menganalisa kecanduan terhadap sosial media yang meliputi :

1. Tidak membahas pada sistem pendukung keputusan maupun sistem informasi untuk pembahasan analisis.
2. Metode data *mining* yang digunakan adalah metode klasifikasi dengan *algoritma C4.5*.
3. Data yang digunakan dalam penelitian adalah data yang diambil dari *survei* yang dilakukan pada beberapa atau sekelompok orang.
4. Data yang digunakan hanya meliputi nama seseorang, alamat rumah, usia seseorang dan berapa lama dia memegang *handphone* sehari dan berapa banyak dia melakukan posting ke *facebook*.

TINJAUAN PUSTAKA

Bagian ini memuat dasar teori yang berfungsi sebagai sumber atau alat dalam memahami permasalahan yang berkaitan dengan aplikasi klasifikasi kecanduan sosial media menggunakan *algoritma C 4.5*.

PENGERTIAN DATA MINING

Kaplan, Andreas M.; Michael Haenlein (2010) Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.

PENGERTIAN ALGORITMA C 4,5

Waliyanto (2000) Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan (*Decision Tree*). Pohon keputusan merupakan metode klasifikasi

Konsep entropy

Mitchell, Tom M. (1997) Entropy (S) merupakan jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel S

Konsep Gain

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)*.

PENGERTIAN FACEBOOK

(Tabak, Alan J. February 9, 2004), Facebook adalah sebuah layanan jejaring sosial berkantor pusat di Menlo Park, California, Amerika Serikat yang diluncurkan pada bulan Februari 2004. Hingga September 2012,

PENGERTIAN WEB

Darma, Jarot S., Shenia A (2009), *Word Wide Web (www)* atau lebih dikenal dengan sebutan *web* merupakan dokumen yang ditulis dalam format HTML (*Hyper Text Markup Language*), yang hampir selalu bisa diakses melalui *http*, yaitu protokol yang menyampaikan informasi dari server untuk ditampilkan kepada para pemakai melalui *web browser*. Semua publikasi dari *web* tersebut dapat membentuk sebuah jaringan informasi yang sangat besar yang dapat diakses dengan menggunakan internet.

pengertian HTML

Wendy willard (2006), HTML adalah singkatan dari *HyperText Markup Language* yaitu bahasa pemrograman standar yang digunakan untuk membuat sebuah halaman web, yang kemudian dapat diakses untuk menampilkan berbagai informasi di dalam sebuah penjelajah *web Internet (Browser)*. HTML dapat juga digunakan sebagai link link antara file-file dalam situs atau dalam komputer dengan menggunakan *localhost*, atau *link* yang menghubungkan antar situs dalam dunia internet.

pengertian PHP

(Arief, 2011) PHP adalah *bahasa pemrograman script server-side yang didesain untuk pengembangan web*. Selain itu, *PHP juga bisa digunakan sebagai bahasa pemrograman umum (wikipedia)*. PHP di kembangkan pada tahun 1995 oleh Rasmus Lerdorf, dan sekarang dikelola oleh The PHP Group.

Pengertian database

Waliyanto (2000), *Basis data (database)* adalah kumpulan data yang disimpan secara sistematis di dalam komputer yang dapat diolah atau dimanipulasi menggunakan perangkat lunak (program aplikasi) untuk menghasilkan informasi. Pendefinisian basis data meliputi spesifikasi berupa tipe data, struktur data dan juga batasan-batasan pada data yang akan disimpan.

pengertian MYSQL

MySQL merupakan basis data sumber terbuka yang paling populer dan banyak digunakan untuk aplikasi berbasis web seperti website

XAMPP

(riyanto,2010). XAMPP yang merupakan singkatan dari *Apache, MySQL, PHP* dan *Perl* sedangkan huruf "X" dimaksudkan sebagai suatu *software* yang dapat dijalankan di empat OS utama seperti *Windows, Mac OS, Linux dan Solaris*. Istilah ini seringkali disebut dengan *cross platform (software multi OS)*.

KLASIFIKASI

(Sulistyo Basuki,1991) Mengetahui metode klasifikasi, secara sederhana algoritma klasifikasi pada data mining merupakan sebuah catatan record data hendak diklasifikasikan kedalam salah satu dari sekian klasifikasi data yang tersedia pada variabel tujuan berdasarkan nilai-nilai "variabel prediktor".

PENGERTIAN UML

UML itu singkatan dari *Unified Modelling Language*. Sesuai dengan kata terakhir dari kepanjangannya, UML merupakan salah satu bentuk language atau bahasa. Menurut pencetusnya, UML didefinisikan sebagai bahasa visual untuk menjelaskan, memberikan spesifikasi, merancang, membuat model, dan mendokumentasikan aspek-aspek dari sebuah sistem.

Use case

Use case diagram menggambarkan fungsionalitas yang diharapkan dari sebuah sistem.

Yang ditekankan adalah "apa" yang diperbuat sistem, dan bukan

“bagaimana”. Sebuah *use case* merepresentasikan sebuah interaksi antara aktor dengan sistem. *Use case* merupakan sebuah pekerjaan tertentu, misalnya login ke sistem, meng-*create* sebuah daftar belanja, dan sebagainya.

activity diagram

Menurut (Romi Satria Wahono, 2003) Pada dasarnya Diagram aktivitas adalah Diagram *flowchart* yang diperluas yang menunjukkan aliran kendali satu aktivitas ke aktivitas lain. Kegunaan diagram ini adalah untuk memodelkan workflow atau jalur kerja, memodelkan operasi, bagaimana

sequence diagram

Menurut Sri Dharwiyanti (2003) *Sequence Diagram* adalah suatu diagram yang digunakan untuk memodelkan skenario penggunaan. Skenario penggunaan adalah barisan kejadian yang terjadi selama satu eksekusi sistem. *Sequence* diagram digunakan untuk :

1. *Overview* perilaku sistem.
2. Menunjukkan objek-objek yang diperlukan.
3. Mendokumentasikan skenario dari suatu *use-case*
4. Memeriksa jalur-jalur pengaksesan.

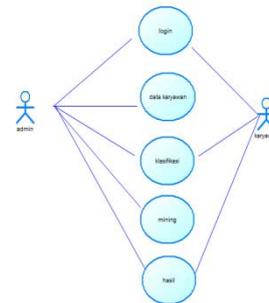
PERANCANGAN SISTEM

Pada Bab ini akan membahas perancangan sistem untuk penentuan kecanduan dengan menggunakan metode *algoritma C 4.5*. Dimana dalam bab ini akan menjelaskan tentang lokasi dan objek penelitian, kebutuhan sistem, perancangan *input* dan *output*, perancangan sistem, perancangan *database* dan perancangan *user interface*. Setiap penjelasan akan diberikan dalam sub bab.

LOKASI DAN OBJEK PENELITIAN

Penelitian berlokasi di PT.JATIM TAMAN STEEL yang berada di Jalan sepanjang km 36 Sidoarjo. Dan obyek penelitian adalah perilaku dari karyawan. proses penelitian ini dilakukan pada Pada tanggal 18 Juni 2018 jam 9 sampai jam 12

Usecase Diagram

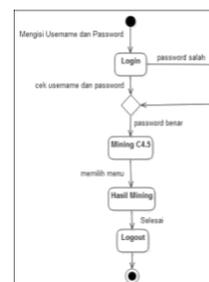


Gambar 1.1 *use case diagram*

Pada diagram gambar 3.1 dapat dilihat ada 2 aktor yang yang dapat berinteraksi dengan sistem, dimana aktor yang dimaksud adalah admin. Admin disini berperan dalam keseluruhan dari system mining dan proses pemasukan data sedangkan karyawan berperan sebagai user yang melakukan klasifikasi.

Activity Diagram

Setiap proses yang berjalan pada system digambarkan melalui rancangan Activity Diagram berikut untuk memberikan penjelasan alur sistem.



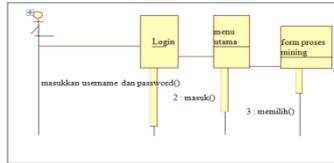
Gambar 1.2 melihat hasil mining Hasil dari proses mining ini hanya bisa dilihat oleh admin yang bersangkutan selalu *user* yang melakukan proses mining.

Sequence Diagram

Pada *sequence diagram* aplikasi ini, keseluruhan diagram menggunakan user admin karena memiliki interaksi terhadap semua *case / state*. Berikut

adalah gambar rancangan *sequence diagram* dari aplikasi yang diusulkan :

Sequence proses mining



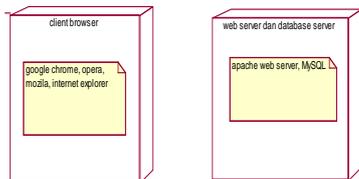
Gambar 1.3 proses mining

Deployment Diagram

Menggambarkan tentang detail hubungan infrastruktur yang terhubung dalam sistem, yaitu di mana komponen akan terletak (pada mesin, server atau piranti keras apa), bagaimana kemampuan jaringan pada lokasi tersebut, spesifikasi server, dan hal-hal lain yang bersifat fisik.

Adapun *Deployment Diagram*

pada sistem secara umum yaitu dapat dilihat pada gambar 1.4



Gambar 1.4 Deployment Diagram

RANCANG TAMPILAN APLIKASI

Desain user interface merupakan desain tampilan sistem yang akan dibuat, mulai dari desain halaman utama atau index sampai desain menu proses. Adapun desain yang dimaksud adalah sebagai berikut :

HALAMAN UTAMA

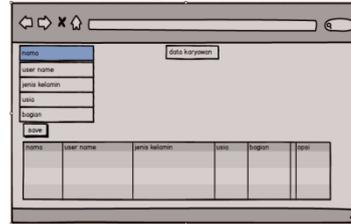
Perancangan halaman utama sistem adalah berisi home dan login untuk admin dan karyawan yang akan melakukan klasifikasi.



Gambar 1.5 rancangan halaman utama

HALAMAN DATA KARYAWAN

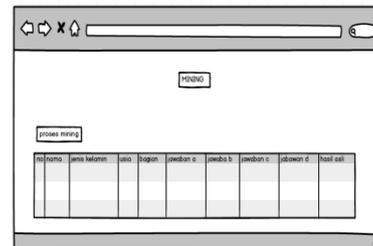
Halaman data karyawan yang berisikan nama, username, jenis kelamin, usia dan bagian yang harus diisi oleh admin



Gambar 1.6 rancangan data karyawan

HALAMAN DATA MINING

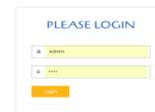
Halaman mining adalah halaman yang digunakan dalam proses mining dan membuat pohon keputusan.



Gambar 1.7 proses mining

PEMBAHASAN

FORM LOGIN



Gambar 1.8 form login

Form login ini berkoneksi dengan database karyawan. Pada form login terdapat username dan password

HALAMAN UTAMA



Gambar 1.9 halaman utama

Dalam halaman ini berisikan halaman utama dan daftar login untuk user dan karyawan yang melakukan klasifikasi kecanduan sosial media

HALAMAN UTAMA ADMIN



Gambar 1.10 halaman utama untuk admin

Form ini dikhususkan hanya untuk admin karena terdapat proses mining didalamnya yang akan diproses oleh admin yang bersangkutan.

HALAMAN MINING



Gambar 1.11 halaman mining

Halaman ini digunakan untuk melakukan proses mining yang dilakukan oleh admin.

Dalam halaman ini data mining bisa diubah dan dihapus oleh admin dapat juga ditambahkan namun *file* yang akan ditambahkan haruslah dalam bentuk *microsoft office axel*.

Kesimpulan

Pemilihan variabel (atribut kondisi dan atribut keputusan) karyawan dalam klasifikasi juga sangat mempengaruhi *knowledge* yang dihasilkan. Karena data atribut karyawan sangat mempengaruhi kalsifikasi kecanduan sosial media facebook

Metode *algoritma C 4.5* dianggap sebagai metode yang sangat membantu dalam melakukan klasifikasi data. Karena karakteristik data yang diklasifikasikan dapat diperoleh dengan jelas yaitu berupa *rule* baru sehingga memudahkan pengguna dalam melakukan penggalian informasi terhadap data yang bersangkutan.

Saran

Beberapa saran dari penulis untuk pengembangan penelitian dari Tugas Akhir ini adalah :

Penelitian dari Tugas Akhir ini dapat dilanjutkan dengan menambahkan atribut-atribut yang memiliki hubungan dengan masalah kecanduan karyawan agar nantinya tingkat pemberian keputusan yang lebih baik dapat tercapai.

Penelitian dari Tugas Akhir ini dapat dilanjutkan dengan menggunakan metode-metode Klasifikasi lainnya untuk mendapatkan perbandingan metode yang lebih baik.

DAFTAR PUSTAKA

David Hartanto Kamagi dan Seng Hansun.(2014).*Implementasi Data Mining*

Dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa.Ultimatics,6(1), 15-20.

Simon Simangunsong dan Dian RatnaSawitri (2017).*Hubungan Stes Dan Kecanduan SmartPhone pada siswa klas X SMA Negeri 5 Surakarta. Jurnal Empati,6(4),52-66.*

Teguh Budi Santoso.(2010).*Analisa dan Penerapan Metode C 4,5 untuk Prediksi Loyalitas pelanggan.Jurnal ImiahFakultasTeknik,10(1),12-20.*

Liliana Swastina.(2013).*Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa,2(1),93-98.*

Mohamed Abdel-Basset, Mai Mohamed 1, Florentin Smarandache,and Victor Chang(2018).*Neutrosophic Association Rule Mining Algorithm forBig Data Analysis,Symmetry,10(106),2-19.*

Darma, Jarot S., Shenia A.(2009). *Buku Pintar Menguasai Internet.jakarta: Media*

Waliyanto.(2000). *Sistem Basis Data AnalisisdanPemodelan Data. Yogyakarta: J & J Learning.*

Arief M Rudianto. (2011). *Pemrograman Web Dinamis menggunakan PHP dan MySQL*. Yogyakarta: C.V ANDI OFFSET

Riyanto.(2010).*Sistem Informasi Penjualan Dengan PHP Dan MySQL*. Yogyakarta:Gava Media.

Basuki, Sulistyو.(1991). *Pengantar Ilmu Perpustakaan*.Jakarta:PT Gramedia Pustaka Utama

Darwiyanti, Sri dan Romi Satria Wahono.(2003).*Pengantar Unified Modeling Language (UML)*. Jakarta: IlmuKomputer.com

Munawar.(2005). *Pemodelan Visual Dengan UML*. Yogyakarta:Grahallmu