

ANALISA METODE *CLASSIFICATION-DECISSION TREE* DAN ALGORITMA C.45 UNTUK MEMPREDIKSI PENYAKIT DIABETES DENGAN MENGGUNAKAN APLIKASI *RAPID MINER*

Febie Elfaladonna¹, Ayu Rahmadani²

¹, Teknik Komputer, Politeknik META Industri
 Cikarang, Indonesia

² Teknik Informatika, Universitas Putra Indonesia-YPTK
 Padang, Indonesia

e-mail: febie.elfakhrul@gmail.com¹, ayurahmadani@gmail.com²

Received : Juli, 2018	Accepted : April, 2019	Published : April, 2019
-----------------------	------------------------	-------------------------

Abstract

Diabetes disease is a degenerative disease that each year the presentation of its victims are always increasing. Ignorance of lay people to predict the likelihood of the disease either from a derivative or derivatives is still not a bit. These things affect the level of vigilance sufferers against things that can trigger diabetes getting worse. Classification of research aims to form model decision tree in order for handling derivative-based diabetes disease are increasingly easy to do. To generate new information then used calculation algorithm c. 45 and testing algorithms that use application rapid miner would further reinforce the decision. The research on testing using multiple attribute classification i.e. the attribute weight, gender, blood pressure, blood sugar levels, and a history of diabetes. All of these attributes will be used as reference in search results so that sufferers can predict whether diabetes is the diabetes disease suffered a derivative or derivatives not

Keywords: *Data mining, Classification Methods, Algorithm C. 45, Diabetes*

Abstrak

Penyakit diabetes merupakan penyakit degeneratif yang setiap tahun presentasi penderitanya selalu meningkat. Ketidaktahuan orang awam untuk memprediksi kemungkinan penyakit ini baik dari turunan maupun bukan turunan masih sedikit. Hal tersebut berpengaruh kepada tingkat kewaspadaan penderita terhadap hal-hal yang dapat memicu penyakit diabetes yang semakin parah. Klasifikasi pada penelitian ini bertujuan untuk membentuk model pohon keputusan agar penanganan penyakit diabetes berdasarkan riwayat turunan semakin mudah dilakukan. Untuk menghasilkan informasi baru maka digunakan perhitungan algoritma C.45 dan pengujian algoritma yang menggunakan aplikasi rapid miner akan semakin memperkuat keputusan. Pada pengujian penelitian ini menggunakan beberapa atribut klasifikasi yakni atribut berat badan, jenis kelamin (atribut bantuan), tekanan darah, kadar gula darah, dan riwayat penyakit diabetes. Semua atribut tersebut akan dijadikan acuan dalam penelusuran hasil sehingga penderita dapat memprediksi sendiri apakah penyakit diabetes yang diderita adalah penyakit diabetes turunan atau tidak turunan.

Kata Kunci: *Data mining, Metode Klasifikasi, Algoritma C.45, Penyakit Diabetes*

1. PENDAHULUAN

Data mining merupakan gabungan sejumlah disiplin ilmu komputer yang didefinisikan sebagai penemuan pola-pola baru dari kumpulan-kumpulan data sangat besar, meliputi metode-metode yang merupakan irisan dari *artificial intelligence*, *machine learning*, *statistics*, dan *database systems*. Data mining memiliki banyak sekali manfaat dalam pengolahan data sehingga data-data yang ada pada masa lampau dapat di jadikan kembali sebagai sumber pengetahuan dan informasi yang baru. Pemanfaatan data mining tidak hanya sebatas ilmu teknologi, tetapi juga pada bidang kesehatan. Data mining dapat dijadikan sebagai acuan untuk memprediksi dan mendiagnosa suatu jenis penyakit dengan menggunakan metode-metode yang dapat diterapkan. Dari sekian banyak penyakit, salah satu penyakit degeneratif yang dapat diprediksi dengan menggunakan metode data mining adalah penyakit diabetes. Penyakit diabetes atau yang sering dikenal dengan sebutan kencing manis ini merupakan penyakit di mana kadar glukosa (gula sederhana) di dalam darah menjadi tinggi karena tubuh tidak dapat memproduksi atau mengeluarkan insulin secara cukup. Penyakit diabetes dapat disebabkan oleh penderita yang memiliki riwayat penyakit diabetes turunan dan karena faktor pemicu lainnya yang tidak berkaitan dengan faktor turunan.

Penyakit diabetes turunan dapat diprediksi sejak dini dengan melihat riwayat turunan penderita. Maka dari itu peneliti tertarik untuk membuat sebuah analisa resiko seseorang yang berpotensi memiliki penyakit diabetes berdasarkan riwayat turunan dan bukan turunan dengan menggunakan metode klasifikasi algoritma C.45 dan memetakannya ke dalam sebuah bentuk pohon keputusan (teknik *decission tree*). Selain untuk menjadikan suatu keputusan kompleks menjadi lebih mudah, kelebihan lain yang didapatkan dari penelitian ini adalah pengujian sampel masalah tersebut hanya berdasarkan pada kriteria atau kelas tertentu. Peneliti juga ingin mengimplementasikan metode tersebut dengan menggunakan aplikasi rapid miner agar dapat menghasilkan keputusan prediksi yang lebih akurat.

1.1 Data Mining

Data mining merupakan disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data (Han and Kamber, 2006) [4]. Data mining sering juga disebut *knowledge discovery in database (KDD)*, adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam *set* data berukuran besar. Keluaran dari *Data mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan (Budi Santosa, 2007) [1]. Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database. *Data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat dan tersimpan di dalam database besar (Turban et al, 2005) [8]. Secara sederhana, data mining atau penambangan data dapat didefinisikan sebagai proses seleksi, eksplorasi, dan pemodelan dari sejumlah besar data untuk menemukan pola atau kecenderungan yang biasanya tidak disadari keberadaannya. Data mining dapat dikatakan sebagai proses mengekstrak pengetahuan dari sejumlah besar data yang tersedia. Pengetahuan yang dihasilkan dari proses data mining harus baru, mudah dimengerti, dan bermanfaat. Dalam data mining, data disimpan secara elektronik dan diproses secara otomatis oleh komputer menggunakan teknik dan perhitungan tertentu (Gambbeta, Windy, 2012) [3].

Selain definisi di atas beberapa definisi juga diberikan seperti, "*Data mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual." (Pramudiono, 2006). "*Data mining* adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya (Pramudiono, 2006) [7]. *Data mining* merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. *Data mining*

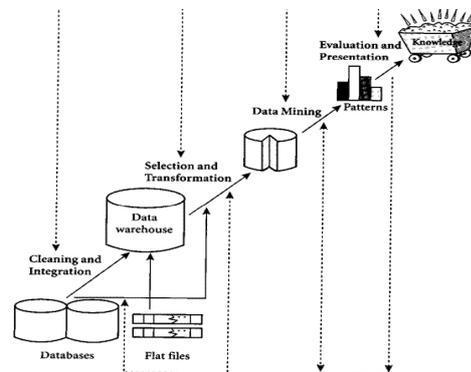
merupakan bidang dari beberapa keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar (Larose, 2006) [5].

1.2 Proses Data Mining

Data mining sesungguhnya merupakan salah satu rangkaian dari proses pencarian pengetahuan pada database (*Knowledge Discovery in Database*). KDD adalah keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti. Serangkaian proses tersebut yang memiliki tahap sebagai berikut (Han and Kamber, 2006) [4]:

- a. Pembersihan data dan integrasi data (*cleaning and integration*)
Proses ini digunakan untuk membuang data yang tidak konsisten dan bersifat *noise* dari data yang terdapat di berbagai basis data yang mungkin berbeda format maupun *platform* yang kemudian diintegrasikan dalam satu database data *warehouse*.
- b. Seleksi dan transformasi data (*selection and transformation*)
Data yang terdapat dalam *database* data *warehouse* kemudian direduksi dengan berbagai teknik. Proses reduksi diperlukan untuk mendapatkan hasil yang lebih akurat dan mengurangi waktu komputasi terutama untuk masalah dengan skala besar (*large scale problem*)
- c. Penambangan data (*data mining*)
Data-data yang telah diseleksi dan ditransformasi ditambah dengan berbagai teknik. Proses *data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan fungsi-fungsi tertentu. Fungsi atau algoritma dalam data mining sangat bervariasi. Pemilihan fungsi atau algoritma yang tepat sangat bergantung pada tujuan dan proses pencarian pengetahuan secara keseluruhan.
- d. Evaluasi pola dan presentasi pengetahuan
Tahap ini merupakan bagian dari proses pencarian pengetahuan yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya. Langkah terakhir KDD adalah

mempresentasikan pengetahuan dalam bentuk yang mudah dipahami oleh pengguna.



Gambar 1. Tahap *Knowledge Discovery in Database*

1.3 Metode Classification

Classification adalah Sebuah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Hal ini juga dapat dikatakan sebagai pembelajaran (klasifikasi) yang memetakan sebuah unsur (item) data kedalam salah satu dari beberapa kelas yang sudah didefinisikan (Bustami, 2013) [2].

Klasifikasi banyak digunakan dalam berbagai aplikasi, diantaranya adalah untuk mendeteksi kecurangan (*fraud detection*), pengelolaan pelanggan, diagnosis medis, prediksi penjualann dan lain sebagainya.

1.4 Decission Tree

Decission tree adalah salah satu metode klasifikasi yang paling populer dan banyak digunakan karena bersifat praktis. salah satu metode pohon keputusan yang paling banyak digunakan adalah *Iterative Dychotomizer Version 3*. Metode ID3 berusaha membangun model klasifikasi yang berupa pohon keputusan secara *top-down*. Caranya adalah dengan mengevaluasi semua atribut menggunakan suatu ukuran statistik, biasanya berupa *information gain*, untuk mengukur efektifitas suatu atribut dalam mengklasifikasikan himpunan sampel data.

Algoritma ID3 atau Iterative Dichotomiser 3 (ID3) merupakan sebuah metode yang digunakan untuk membangkitkan pohon keputusan. Algoritma pada metode ini menggunakan konsep dari entropy informasi. Pemilihan atribut dengan menggunakan

Information Gain. Pemilihan atribut pada ID3 dilakukan dengan properti statistik, yang disebut dengan information gain. Gain mengukur seberapa baik suatu atribut memisahkan training example ke dalam kelas target. Atribut dengan informasi tertinggi akan dipilih. Dengan tujuan untuk mendefinisikan gain, pertama-tama digunakanlah ide dari teori informasi yang disebut entropy. Entropy mengukur jumlah dari informasi yang ada pada atribut dengan rumus :

$$Entropy(S) = -P_+ \log_2 P_+ - P_- \log_2 P_-$$

S adalah ruang (data) sample yang digunakan untuk training. P₊ adalah jumlah yang bersolusi positif (mendukung) pada data sample untuk kriteria tertentu. P₋ adalah jumlah yang bersolusi negative (tidak mendukung) pada data sample untuk kriteria tertentu. Dari rumus entropy diatas dapat disimpulkan bahwa definisi entropy (S) adalah jumlah bit yang

1.5 Algoritma C.45

Algoritma C45 merupakan metode yang menjadi pilihan pertama dan sering digunakan dalam pengembangan Data Mining karena kecepatan dalam pengklasifikasian pohon keputusan disamping dapat mengkonstruksi pengklasifikasian dengan aturan-aturan yang lain (Wu Xindong et al, 2008) [9].

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

1.6 Rapid Miner

Rapid miner merupakan platform perangkat lunak ilmu data yang dikembangkan oleh perusahaan yang bernama sama dengan pihak yang menyediakan lingkungan terintegrasi untuk persiapan data, penambangan teks, dan analisis prediktif. Rapid miner dilengkapi dengan satu prosesor logika dan 10.000 baris data. Rapid miner yang tersedia berada di bawah lisensi AGPL.

diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau-) dari sejumlah data acak pada suatu ruang sampel S. Entropy bisa dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. Semakin kecil nilai entropy maka semakin baik digunakan dalam mengekstraksi suatu kelas. Pada algoritma ID3 pengurangan entropy disebut dengan informasi gain. Pembagian sample S terhadap atribut A dapat dihitung information gain dengan rumus:

$$Gain(S,A) = Entropy(S) - \sum_{v \in nilai(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Pilih atribut yang memiliki nilai information gain terbesar, ulangi proses perhitungan information gain akan terus dilaksanakan sampai semua data telah masuk dalam kelas yang sama. Atribut yang telah dipilih tidak diikuti lagi dalam perhitungan nilai information gain (Kristanto Obbie, 2011) [6].

2. METODE PENELITIAN

Berisi mengenai tahapan-tahapan penelitian dimulai dari penelitian awal berupa pengumpulan data sampai kepada penerapan algoritma kepada aplikasi *rapid miner* sehingga ditemukan kecocokan antara hasil akhir terhadap masing-masing perhitungan. Output dari penelitian ini adalah analisa terhadap hasil prediksi faktor seseorang bisa terkena penyakit diabetes turunan dan bukan turunan.

Pada penelitian ini, peneliti menemukan beberapa keadaan yang membuat peneliti tertarik untuk melakukan suatu penganalisaan terhadap kemungkinan – kemungkinan yang akan membuat seseorang terkena penyakit diabetes. Seperti yang peneliti ketahui, penyakit diabetes tidak hanya dapat di jumpai pada orang yang memiliki keturunan penyakit diabetes, namun juga berpotensi pada seseorang yang memiliki gaya hidup tidak sehat serta beberapa factor-faktor pemicu lainnya.

Tahapan dari penelitian ini dilakukan dengan beberapa metode tahapan yang saling berhubungan antara satu dengan yang lainnya. Metode yang dilakukan berupa:

1. Mengumpulkan data pasien penderita diabetes. Pada tahapan ini dilakukan pengumpulan data dari pasien penderita diabetes. Pengumpulan data dilakukan melalui wawancara serta meminta sebagian data pasien dari petugas *Medical Record* RSUP. DR.M. Djamil Padang.

2. Pengklasifikasian data – data pasien penderita diabetes. Setelah data–data dikumpulkan, maka dilanjutkan dengan proses mengelompokkan data ke dalam MS.Excel.
3. Melakukan pengolahan data dengan menggunakan *software rapid miner*.
4. Setelah data diklasifikasikan ke dalam *Ms.Excel*, maka data tersebut kemudian akan diproses dengan menggunakan *software rapid miner*. Dengan menggunakan software tersebut maka data akan dapat dianalisa untuk mendapatkan sebuah keputusan.

3. HASIL DAN PEMBAHASAN

Berikut merupakan data-data pasien penderita diabetes yang telah diklasifikasikan berdasarkan beberapa atribut yang sudah ditetapkan. Atribut-atribut tersebut terdiri dari berat badan, jenis kelamin, tekanan darah, kadar gula darah dan riwayat penyakit (turunan dan bukan turunan).

Tabel 1. Atribut Berat Badan Penyakit Diabetes

Atribut	Instance			Variabel
	Under Weight	Average	Over Weight	
Berat Badan	≤ 45Kg	46-60 Kg	≥ 61 Kg	Turunan dan Tidak Turunan

Tabel 2. Atribut Tekanan Darah Penyakit Diabetes

Atribut	Instance			Variabel
	Rendah	Normal	Tinggi	
Tekanan Darah	<100/90 mmHg	120/80 mmHg	>140/90 mmHg	Turunan dan Tidak Turunan

Tabel 3. Atribut Kadar gula darah Penyakit Diabetes

Atribut	Instance			Variabel
	Rendah	Normal	Tinggi	
Kadar gula darah	>70 mg/Dl	145 mg/Dl	<200 mg/Dl	Turunan dan Tidak Turunan

Berikut ini merupakan tabel data-data yang akan diolah dengan menggunakan perhitungan manual algoritma C.45.

Tabel 4. Data Pasien Penyakit Diabetes (1)

Berat Badan	Jenis Kelamin	Tekanan Darah	Kadar gula darah	Riwayat Diabetes
Under Weight	Laki-laki	Normal	Tinggi	Tidak Turunan
Under Weight	Perempuan	Normal	Tinggi	Tidak Turunan
Average	Laki-laki	Normal	Tinggi	Turunan
Over Weight	Laki-laki	Tinggi	Tinggi	Tidak Turunan
Over Weight	Laki-laki	Rendah	Normal	Turunan
Over Weight	Perempuan	Rendah	Normal	Turunan
Average	Perempuan	Rendah	Normal	Turunan

Under Weight	Laki-laki	Tinggi	Tinggi	Tidak Turunan
Under Weight	Laki-laki	Rendah	Normal	Turunan
Over Weight	Laki-laki	Tinggi	Normal	Turunan
Under Weight	Perempuan	Tinggi	Normal	Turunan
Average	Perempuan	Tinggi	Tinggi	Turunan
Average	Laki-laki	Normal	Normal	Turunan
Over Weight	Perempuan	Tinggi	Tinggi	Tidak Turunan
Under Weight	Laki-laki	Normal	Rendah	Turunan
Average	Laki-laki	Tinggi	Rendah	Turunan
Under Weight	Perempuan	Tinggi	Rendah	Turunan
Over Weight	Laki-laki	Rendah	Tinggi	Turunan
Average	Perempuan	Normal	Rendah	Turunan
Over Weight	Perempuan	Rendah	Tinggi	Turunan

Setelah ditetapkan atribut dalam prediksi penyakit diabetes, langkah selanjutnya yang dilakukan adalah melakukan perhitungan manual dengan menggunakan algoritma C.45.

- Hitung jumlah kasus
- Hitung jumlah keputusan **Turunan**
- Hitung jumlah keputusan **Tidak Turunan**

Hitung :

- Entropy (Total)
- Entropy (Berat Badan)
- Entropy (Jenis Kelamin)
- Entropy (Tekanan Darah)
- Entropy (Kadar gula darah)

$$\sum_{j=1}^k - p_j \log_2 p_j$$

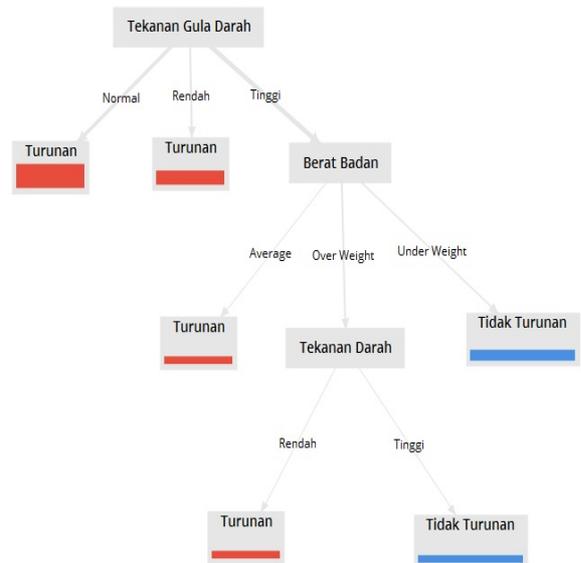
Gambar 2. Rumus Mencari Nilai Entropy

Setelah melakukan perhitungan dengan menggunakan Algoritma C.45, maka didapatkan hasil akhir yang telah disesuaikan dengan menggunakan pohon keputusan seperti yang terlihat pada gambar berikut.

Gambar 3. Pohon Keputusan Prediksi Penyakit Diabetes

4. Pengujian Hasil

Sistem aplikasi yang akan peneliti gunakan untuk pengolahan data pasien dengan resiko diabetes adalah *rapid miner 8.2*. Berikut adalah ilustrasi atau gambar hasil pohon keputusan yang didapatkan setelah data dikelola dengan menggunakan *rapid miner 8.2*.



Gambar 4. Pohon Keputusan Rapid Miner 8.2

4. KESIMPULAN

- a. *Data mining* bersifat independen terhadap aplikasi terkait, ini berarti untuk aplikasi basis data yang berbeda, maka teknik *data mining* yang digunakan mungkin juga akan berbeda.
- b. Dalam perkembangan teknologi *data mining*, terdapat model atau node yang digunakan untuk melakukan proses penggalian informasi terhadap data-data yang ada.
- c. Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki.
- d. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:
 - 1) Pilih atribut sebagai akar.
 - 2) Buat cabang untuk masing-masing nilai.
 - 3) Bagi kasus dalam cabang.
 - 4) Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.
- e. Dalam melakukan perbandingan perhitungan secara manual algoritma C.45 dan menggunakan aplikasi *rapid miner* 8.2, ternyata peneliti menemukan pohon keputusan yang di hasilkan adalah sama.
- f. Kesimpulan hasil dari pohon keputusan berisi tentang analisa terhadap faktor – faktor seseorang berpotensi terkena penyakit diabetes dengan melihat kepada atribut – atribut seperti jenis kelamin (jenis kelamin hanya sebagai atribut bantuan yang tidak dapat dijadikan perhitungan dalam memprediksi riwayat penyakit diabetes), berat badan, tekanan darah, dan kadar gula darah dan variabel riwayat penyakit turunan dan tidak turunan. Ternyata, setelah di teliti dengan membuat perhitungan secara manual, dan membandingkannya dengan penggunaan sistem aplikasi, ada dua kemungkinan seseorang menderita penyakit diabetes. Dua kemungkinan tersebut antara lain:
 - 1) “Seseorang yang menderita penyakit diabetes adalah seseorang dengan kadar gula darah yang tinggi, berat badan tinggi, dan tekanan darah rendah, tetapi memiliki riwayat penyakit diabetes secara turunan”.
 - 2) “Seseorang yang menderita penyakit diabetes adalah seseorang dengan kadar

gula darah tinggi, berat badan yang tinggi, dan tekanan darah tinggi, tetapi tidak memiliki riwayat penyakit diabetes (tidak turunan)”.

DAFTAR PUSTAKA

- [1] Budi Santosa, 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta. Graha Ilmu.
- [2] Bustami, 2013. Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *Jurnal Penelitian IJNS – Indonesian Journal on Networking and Security - Volume 5 No 1 – 2016 – ijns.org* ISSN : 1979-9330 (Print) - 2088-0154 (Online) 54 *Teknik Informatika (TECHSI)* Vol. 2: 2, 127-146, 2013.
- [3] Gambbetta, Windy. 2012. *Pohon Keputusan (Decision Tree)*. Departemen Teknik Informatika. Institute Teknologi Bandung. Bandung.
- [4] Iwan Santosa, dkk. 2018. Implementasi Algoritma Decision Tree C.45 Untuk Diagnosa penyakit Tuberculosis (TB). *Jurnal Ilmiah Nero Vol.3, No 3. Program Studi Teknik Informatika. Universitas Trunojoyo Madura*.
- [5] Jiawei Han and Micheline Kamber Simon Fraser, 2000. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers. All rights reserved.
- [6] Larose D, T., 2006, *Data Mining Methods and Models*, Jhon Wiley & Sons, Inc. Hoboken New Jersey.
- [7] Obbie Kristanto, 2015. Penerapan Algoritma Klasifikasi Data Mining Id3 Untuk Menentukan Penjurusan Siswa SMAN 6 Semarang. Jurusan Teknik Informatika. Universitas Dian Nuswantoro. Semarang.
- [8] Pramudiono, I. 2006. Apa itu Data Mining. Dalam <http://datamining.japati.net/cgibin>, diakses tanggal 16 september 2017. ISSN : 2548-964X. <http://jptiik.ub.ac.id>. Program Studi Teknik Informatika. Fakultas Ilmu Komputer. Universitas Brawijaya

- [9] Rizky Haqmanullah Pambudi, dkk. Penerapan Algoritma C.45 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Vol.2, No. 7, Juli 2018, hlm 2637-2643.
- [10] Turban, E. et al, 2005. Decision Support Systems and Intelligent Systems (Sistem Pendukung Keputusan dan Sistem Cerdas). Andi Offset. Yogyakarta
- [11] Wu Xindong et al. Top 10 Algorithms in Data Mining. Di dalam: Knowledge Information System. Vol. 14. London: Springer; 2008. hlm. 1-37