

**KLASIFIKASI PENYAKIT DIABETES MELLITUS TIPE 2
DENGAN METODE Algoritma C4.5
(Studi Kasus: Rumah Sakit Waluyojati Kraksan Probolinggo)**

Moh. Jasri¹

Program Studi Teknik Informatika dan Teknik Elektronika, STT Nurul Jadid
JL. KH Zaini Mun'im, Kecamatan Paiton, Karanganyar, Paiton, Probolinggo, Jawa Timur 67291

*Email: jasri@sttnj.ac.id

Abstrak

Angka penderita diabetes selama 50 tahun terakhir meningkat pesat seiring dengan meningkatnya angka kegemukan. Pada tahun 2010, diperkirakan ada 285 juta orang mengalami penyakit ini, dibandingkan ada 30 juta pasien pada tahun 1985. Komplikasi jangka panjang yang mungkin terjadi akibat kadar glukosa darah tinggi antara lain penyakit jantung, stroke, retinopati diabetes yang mempengaruhi penglihatan mata, gagal ginjal yang memerlukan dialisis, dan kurangnya sirkulasi darah di bagian tungkai yang mengharuskan dilakukan amputasi. Dari pemaparan tersebut jumlah penderita DM2 rentan dengan komplikasi dan bisa berakibat fatal bagi penderita. Dengan mengklasifikasikan jenis-jenis penyakit lain yang menyertai DM2 dengan tingkat akurasi yang tinggi, maka diharapkan sangat membantu dalam proses penanganan penyakit, sehingga bisa diatasi dengan langkah yang sangat tepat. Salah satu langkah yang sangat tepat ialah klasifikasi.

Klasifikasi adalah salah satu solusi dalam data mining untuk mengurai permasalahan tersebut, dengan mengklasifikasikan jenis-jenis penyakit yang berhubungan dengan DM2 dengan menggunakan metode C4.5 dengan tujuan agar diketahui pola hubungan DM2 dengan penyakit lainnya, sehingga kita bisa memprediksi penyakit apa saja yang akan diderita dari pasien selain DM2.

Hasil dari penelitian ini adalah 10 tree penyakit yang dihasilkan dari data pola hidup dan data fisik pasien, dengan mengabungkan 10 tree adalah salah satu solusi agar penyakit yang lain tampil dalam klasifikasi penyakitnya. Bukan hanya fokus pada satu penyakit saja, dan akurasi dari 10 tree setelah dilakukan pemangkasan atau pruning menjadi rata-rata meningkat sebesar 1 %.

Kata kunci: Klasifikasi, Data Mining, DM2, dan C4.5.

1. PENDAHULUAN

Diabetes Melitus Tipe 2 atau DM2 adalah penyakit gangguan metabolik yang ditandai oleh kenaikan gula darah akibat penurunan sekresi insulin oleh sel beta pankreas dan atau gangguan fungsi insulin (resistensi insulin). Hasil Riset Kesehatan Dasar pada tahun 2008, menunjukkan angka kejadian Diabetes Melitus atau DM di Indonesia mencapai 57%, sedangkan kejadian di dunia DM2 adalah 95%. Faktor resiko dari DM2 yaitu usia, jenis kelamin, obesitas, hipertensi, genetik, makanan, merokok, alkohol, kurang aktivitas, lingkaran perut. Penatalaksanaan dilakukan dengan cara penggunaan obat oral hiperglikemi dan insulin serta modifikasi gaya hidup untuk mengurangi kejadian dan komplikasi mikrovaskular maupun makrovaskular dari DM2.

Data mining adalah sinonim untuk penemuan pengetahuan istilah lain yang populer dalam database (KDD). Ada tiga proses dalam KDD (preprocessing, data mining, pengolahan pasca). Preprocessing dieksekusi sebelum teknik data mining yang diterapkan pada data. Proses preprocessing meliputi data cleaning, integrasi data, seleksi data dan data transformasi. Proses utama KDD adalah proses data mining. Dalam proses ini berbeda algoritma yang diterapkan untuk menghasilkan pengetahuan yang tersembunyi. Setelah proses lain disebut post pengolahan, mengevaluasi hasil mining sesuai dengan kebutuhan pengguna dan pengetahuan domain. Mengenai hasil evaluasi, pengetahuan dapat disajikan jika hasilnya memuaskan, sebaliknya kita harus menjalankan beberapa atau semua proses-proses lagi sampai kita mendapatkan hasil yang memuaskan. Oleh karena itu maka dipandang perlu untuk melakukan sebuah penelitian dalam mengklasifikasikan jenis-jenis penyakit yang berhubungan dengan DM2 dengan menggunakan metode C4.5 dengan tujuan agar diketahui pola hubungan DM2 dengan penyakit lainnya, sehingga

kita bisa memprediksi penyakit apa saja yang akan diderita dari pasien selain DM2. Dengan begitu bisa dilakukan pencegahan lebih dini agar tidak berakibat fatal dikemudian hari.

2. METODOLOGI

Dalam sub bab ini akan membahas tentang analisa permasalahan, analisa kebutuhan sistem arsitektur sistem, beserta spesifikasi dan kemampuan sistem.

2.1 ANALISIS PERMASALAHAN

Data dalam penelitian ini didapat dari seorang petugas rumah sakit, setelah mendapat surat keputusan penelitian dari kampus data bisa didapatkan secara lengkap dari pihak Rumah Sakit Waluyoajati Kraksan Probolinggo (RSML), sesungguhnya data pada penelitian ini memiliki banyak sekali penyakit yang terhubung dengan DM2, namun 10 penyakit yang sering muncul dalam data sehingga hanya 10 penyakit yang digunakan dalam klasifikasi penyakit yang terkait dengan DM2, Fitur-fitur yang terdapat dalam data penelitian ini terdiri dari nama_pasien, alamat, tanggal periksa, diagnosa, Merokok, Olah Raga, Diet, Rutin Kontrol, Suntik Insulin, Berat, Tinggi, Kadar Gula Puasa - 2 jam, Tekanan Darah, Jenis Kelamin, Obat dan Umur.

Tabel1 Attribute Klasifikasi Penyakit DM2

Attribute	Values
Merokok	Boolean
Olah Raga	Boolean
Diet	Boolean
Rutin Kontrol	Boolean
Insulin	Boolean
Berat	Number
Tinggi	Number
Kadar Gula Puasa - 2 jam	Number
Tekanan Darah	Number
Jenis Kelamin	Boolean
Umur	Number

Namun karena nama_pasien, alamat, tanggal periksa dan obat tidak berpengaruh terhadap klasifikasi penyakit terhadap penderita DM2, maka fitur yang digunakan hanya 11 fitur, sedangkan fitur yang lain tidak digunakan dan nampak pada tabel1

Setelah itu dilakukan klasifikasi, maka penyakit yang mempunyai frekuensi kemunculan diambil 10 kategori. Penyakit yang menyertai DM2 dalam penelitian ini adalah sebanyak 10 penyakit. Jenis-jenis penyakit tersebut bisa dilihat pada tabel2.

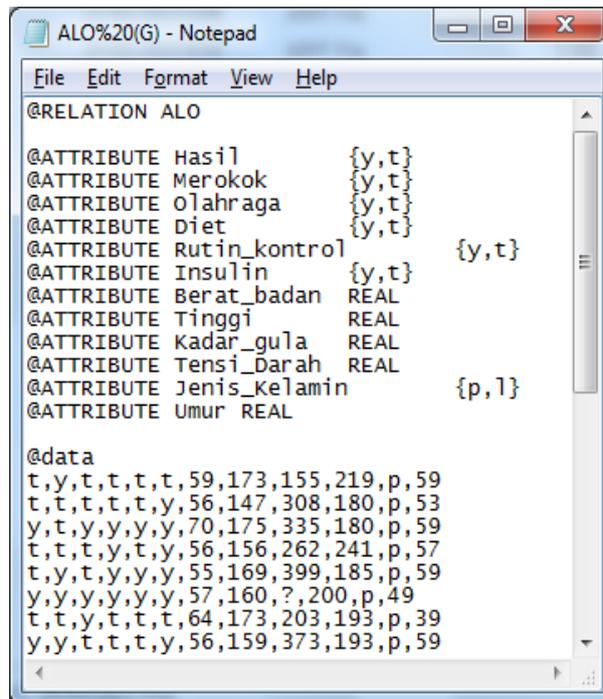
Dataset ini mengandung data *Numeric Value* dan *Missing Value* yang sering dijumpai pada sebuah data yang akan di-mining, kemudian akan menghasilkan *Decision Tree*, namun *Decision Tree* yang dihasilkan kurang efisien sehingga perlu dilakukan *pruning* untuk mendapatkan *Decision Tree* yang lebih pendek namun hasil keputusannya sama. Pruning dapat dilakukan dengan cara menghitung pengaruh tiap simpul pada keakuratan pohon kemudian pangkas simpul yang kurang berpengaruh.

Tabel2 Dataset Klasifikasi Penyakit DM2

No.	Penyakit Yang Menyertai DM2	Kode
1	ALO	A
2	Cardiomyopathy	H
3	CVA Infark	M
4	Decompensatio Cordis	L
5	Dispepsia	P
6	Dislipidemia	Q
7	HHF	S
8	Hypertensi Essensial	T
9	OMI	V
10	PJK	A5

Dalam penelitian ini awalnya berbentuk microsoft excel dalam format CSV, dan software data mining yang digunakan dalam penelitian ini hanya bisa membaca data yang berformat Arff, sehingga proses klasifikasi bisa dilakukan oleh software data mining weka. Arff adalah format file yang digunakan dalam weka dan dibuat oleh Andrew Donkin. Dalam file .arff, terdiri dari 2 tab, yaitu, header dan data. Dalam file di atas, kotak pertama adalah header, dan kotak ke dua adalah data.

Pada bagian kedua, yakni data, cukup dituliskan **@data** dan selanjutnya, menulis tiap baris dengan isi data. Tiap atribut dipisahkan dengan tanda koma, Untuk atribut yang belum diketahui, cukup diganti dengan tanda tanya (?). Sedangkan untuk tipe string, gunakan tanda petik tunggal untuk mengapitnya. Kelas penyakit yang ada di Rumah Sakit Waluyojadi Kraksan Probolinggo ada 4 kelas, diantaranya: penyakit ALO, penyakit Cardiomyopathy, penyakit CVA Infark, penyakit Decompensatio Cordis, penyakit Dislipidemia, penyakit Decompensatio Cordis, penyakit HHF, penyakit Hypertensi Essensial, penyakit OMI dan penyakit PJK. Sebenarnya, file csv ini bisa diterima oleh WEKA sebagai input. Tapi ada 2 standar .csv yang menggunakan (,) dan (;). Dan WEKA menggunakan (,), sehingga ketika disimpan ke .csv, dengan menggunakan (;) tidak akan terbaca. Oleh, karena itu harus secara manual diubah (;) menjadi (,) dengan cara: buka di notepad, dan gunakan fungsi replace, file ber-ekstensi *.csv. Jika file tersebut diopen menggunakan notepad. Format inilah yang nantinya dirubah secara manual, kemudian disave dengan nama yang sama, tapi ekstensinya disimpan dalam format *.arff.



```

ALO%20(G) - Notepad
File Edit Format View Help
@RELATION ALO
@ATTRIBUTE Hasil {y,t}
@ATTRIBUTE Merokok {y,t}
@ATTRIBUTE Olahraga {y,t}
@ATTRIBUTE Diet {y,t}
@ATTRIBUTE Rutin_kontrol {y,t}
@ATTRIBUTE Insulin {y,t}
@ATTRIBUTE Berat_badan REAL
@ATTRIBUTE Tinggi REAL
@ATTRIBUTE Kadar_gula REAL
@ATTRIBUTE Tensi_Darah REAL
@ATTRIBUTE Jenis_Kelamin {p,l}
@ATTRIBUTE Umur REAL

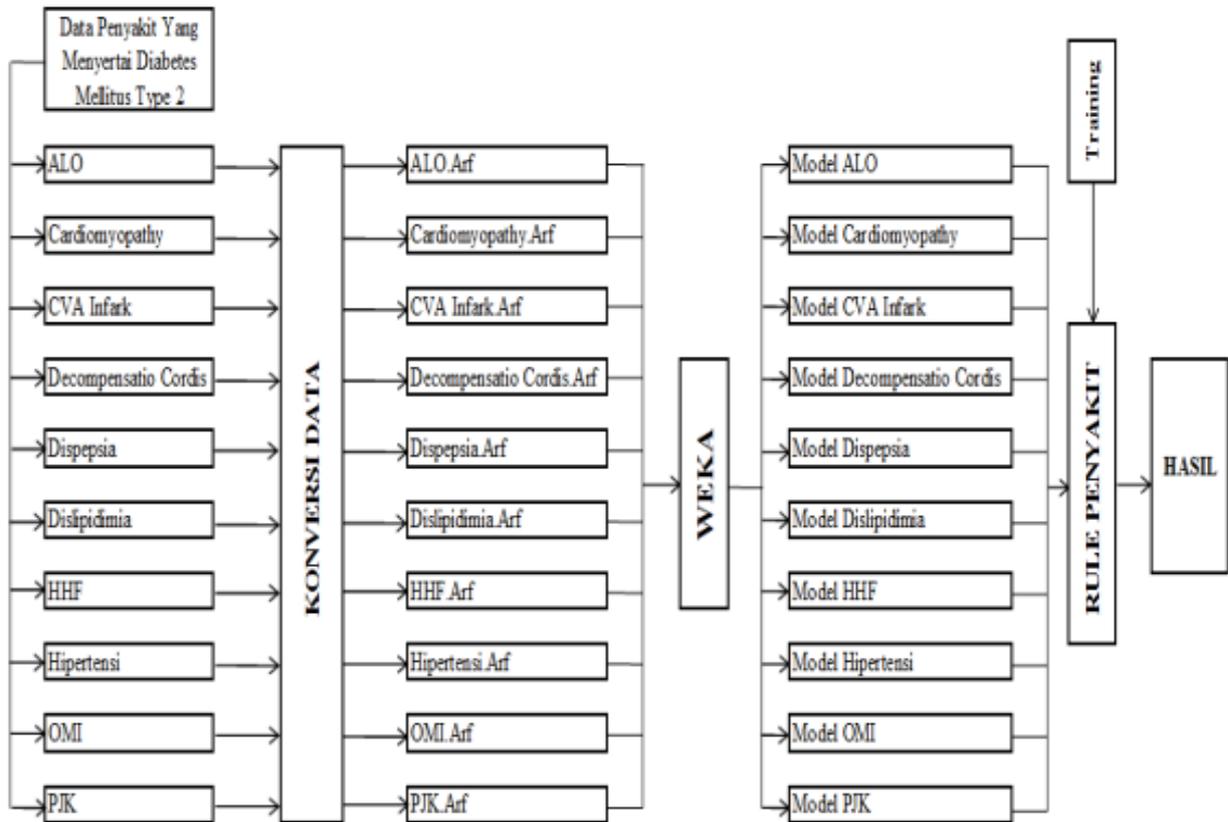
@data
t,y,t,t,t,t,59,173,155,219,p,59
t,t,t,t,t,y,56,147,308,180,p,53
y,t,y,y,y,y,70,175,335,180,p,59
t,t,t,y,t,y,56,156,262,241,p,57
t,y,t,y,y,y,55,169,399,185,p,59
y,y,y,y,y,y,57,160,?,200,p,49
t,t,y,t,t,t,64,173,203,193,p,39
y,y,t,t,t,y,56,159,373,193,p,59

```

Gambar1 Data File .Arff Open Notepad

2.2 Rancangan algoritma

Pada bab ini akan dijelaskan mengenai desain sistem yang didasarkan pada analisa sistem, yang meliputi perancangan algoritma, dan penerapannya dalam bahasa pemrograman.



Gambar2 Blok Diagram Metode C4.5 (C4.5)

Tahap pertama yang dilakukan adalah mengumpulkan data penyakit yang menyebabkan terjadinya DM2 di Rumah Sakit Waluyojati Kraksaan Probolinggo (RSML) dalam program excel yang dikonversi ke dalam bentuk Arff. Arff merupakan inputan data dari software *Weka*.

Selanjutnya, konversi data dari excel ke Arff yang bisa dilakukan dengan dua cara, yaitu secara otomatis menggunakan software *Weka* atau secara manual dengan cara membuat sendiri data arff. Jika dilakukan secara manual maka excel kita save as kembali ke dalam format CSV dan dirubah secara manual dengan menggunakan notepad.

Tahap kedua adalah melakukan proses *C4.5* dengan input data penyakit yang sudah terkonsversi ke bentuk Arff. Dasar dari penggunaan metode ini adalah karena *C4.5* terbukti mampu memperbaiki data yang memiliki banyak noise atau dalam kata lain memiliki banyak missing value. Selain itu *C4.5* juga bisa mengatasi data yang berupa numeric. Oleh karena itu metode ini sangat tepat jika digunakan untuk mengklasifikasi jenis-jenis penyakit yang menyertai DM2 yang memiliki nilai *noise* dan nilai *numerik*.

Tahap ketiga adalah melakukan pembuatan *rule* yang dibantu dengan software *Weka*, sehingga proses pembuatan rulenya bisa lebih cepat dan efisien. *Weka* adalah library yang sangat tepat dalam penelitian ini karena library *Weka* bisa diimport dibahasa pemograman, salah satunya adalah java.

Tahap keempat adalah membuat aplikasi dengan menggunakan java dengan menggunakan library yang ada di *Weka*, sehingga proses pembuatan program atau aplikasi data mining bisa lebih cepat dari pada membuat secara manual dengan meng-*Algoritma* seperti biasa.

Tahap kelima adalah melakukan uji coba pada program java yang sudah terkoneksi pada *Weka*. Uji coba ini dilakukan untuk menghasilkan prediksi klasifikasi penyakit DM2 sesuai dengan inputan data yang ada.

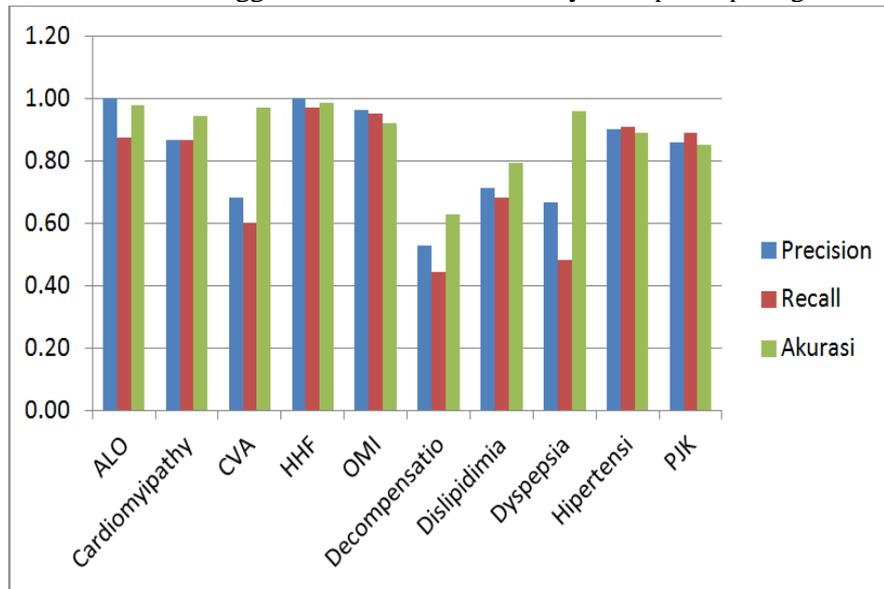
3. HASIL DAN PEMBAHASAN

Hasil dari prototype interface perangkat lunak yang sudah diujikan ke anak-anak dikembangkan lagi ke tahapan yang lebih lanjut yaitu ke perancangan interface perangkat lunak

yang baru. Evaluasi hasil pengujian dari sisi efektif, efisien dan kepuasan pengguna dijadikan pertimbangan oleh peneliti dalam merancang interface yang baru.

Hasil Perhitungan Sebelum Pruning

Tujuan dari penelitian ini menguji keakuratan klasifikasi jenis-jenis penyakit yang berkaitan dengan DM2 di Rumah Sakit Waluyoajati Kraksan Probolinggo (RSML). pada perhitungan Precision, Recall, dan akurasi menggunakan weka maka hasilnya nampak seperti gambar3.



Gambar3 Grafik Precision Sebelum Pruning

Sedangkan untuk nilai Precision, Recall, dan Akurasi untuk masing-masing penyakit bisa dilihat pada tabel3.

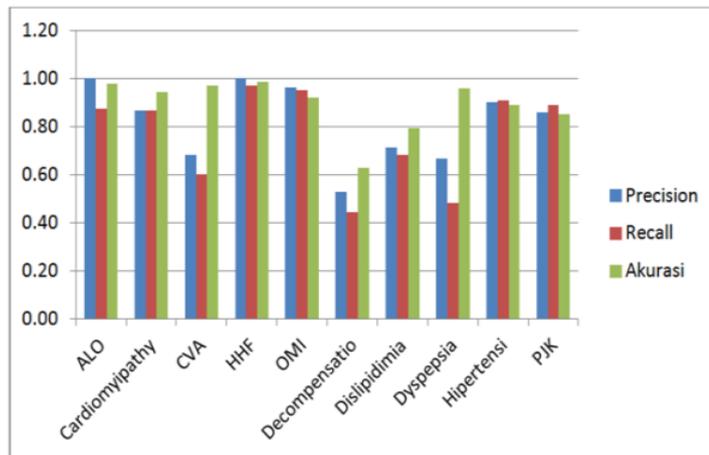
Tabel3 Nilai Sebelum Pruning

Penyakit	Precision	Recall	Akurasi
ALO	100 %	88 %	98 %
Cardiomyipathy	87 %	86 %	94 %
CVA	68 %	60 %	97 %
HHF	100 %	97 %	99 %
OMI	96 %	95 %	92 %
Decompensatio	53 %	44 %	63 %
Dislipidimia	71 %	68 %	79 %
Dyspepsia	67 %	48 %	96 %
Hipertensi	90 %	91 %	89 %
PJK	86 %	89 %	85 %
AVG	82 %	77 %	89 %

Hasil Perhitungan Sesudah Pruning

Tujuan dari penelitian ini menguji keakuratan klasifikasi jenis-jenis penyakit yang berkaitan dengan DM2 di Rumah Sakit Waluyoajati Kraksan Probolinggo (RSML).

Adapun hasilnya nampak seperti Gambar4.



Gambar4 Grafik Precision Sesudah Pruning

Sedangkan untuk nilai Precision, Recall, dan akurasi untuk masing-masing penyakit setelah pruning bisa dilihat pada tabel 4.

Tabel4 Nilai Precision, Recall dan akurasi penyakit

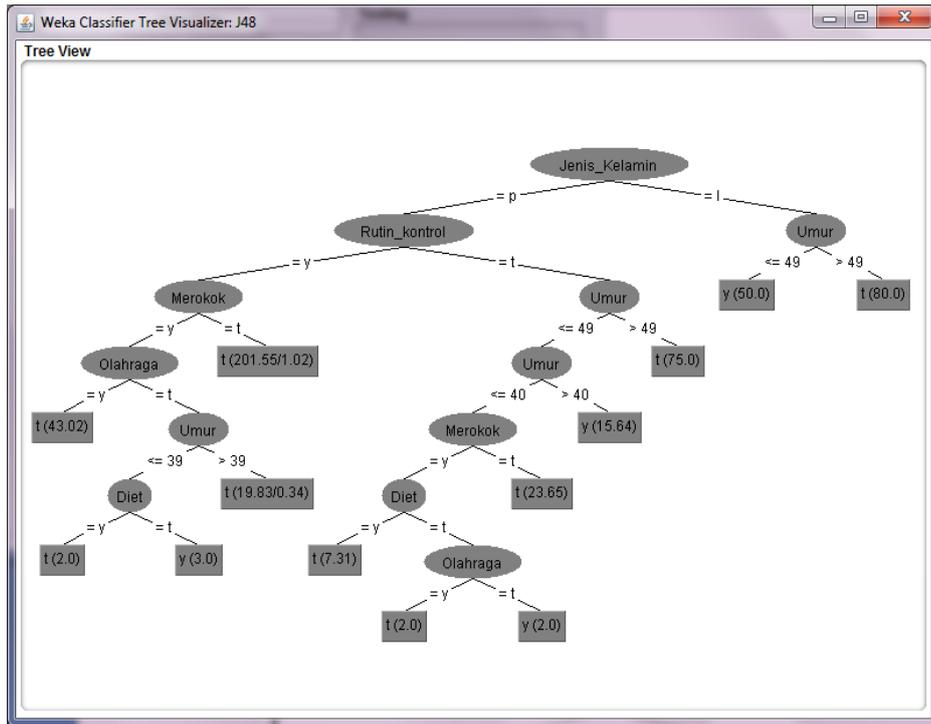
Penyakit	Precision	Recall	Akurasi
ALO	100 %	88 %	98 %
Cardiomyopathy	87 %	86 %	95 %
CVA	68 %	60 %	97 %
HHF	100 %	97 %	99 %
OMI	96 %	95 %	93 %
Decompensatio	53 %	44 %	66 %
Dislipidimia	71 %	68 %	79 %
Dyspepsia	67 %	48 %	96 %
Hipertensi	90 %	91 %	91 %
PJK	86 %	89 %	88 %
AVG	82 %	77 %	90 %

Membandingkan No Pruning dan Pruning

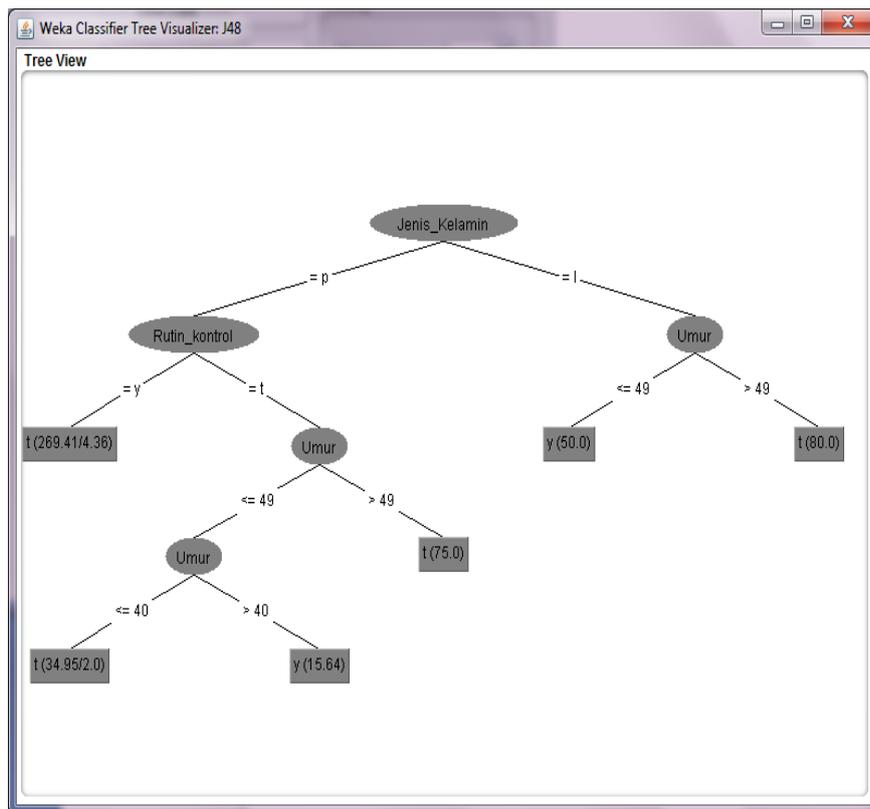
Berdasarkan uji coba yang telah dilakukan, diperoleh bahwa *Decision Tree C4.5* memberikan hasil akurasi rata-rata sebesar 96 %. Sedangkan setelah dilakukan proses missing value tingkat akurasinya meningkat sebesar 2,058 % ini bisa dilihat pada tabel5

Tabel5 Nilai Precision, Recall dan akurasi penyakit

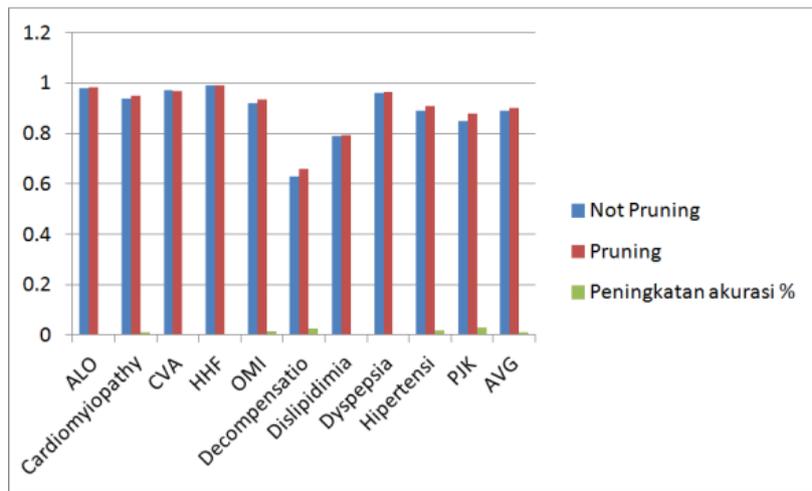
Penyakit	Not Pruning	Pruning	Peningkatan Akurasi
ALO	98 %	98.286 %	0.286 %
Cardiomyopathy	94 %	95.048 %	0.105 %
CVA	97 %	96.762 %	0.238 %
HHF	99 %	98.857 %	0.143 %
OMI	92 %	93.333 %	1.333 %
Decompensatio	63 %	65.714 %	2.714 %
Dislipidimia	79 %	79.191 %	0.191 %
Dyspepsia	96 %	96.381 %	0.381 %
Hipertensi	89 %	90.857 %	1.857 %
PJK	85 %	88 %	3 %
AVG	89 %	90 %	1 %



Gambar5 Tree ALO Sebelum Pruning



Gambar6 Tree ALO Sebelum Pruning



Gambar7 Grafik Batang Perbandingan Not Pruning Dan Pruning

4. KESIMPULAN

Dari hasil penelitian ini dapat disimpulkan hal-hal sebagai berikut :

- Tingkat akurasi rata-rata pada klasifikasi penyakit yang berkaitan dengan DM2 sebesar 90 %.
- Precision dan recall pada 10 penyakit memiliki nilai yang berdekatan, hal ini membuktikan bahwa 10 penyakit tersebut memiliki keterkaitan dengan DM2.
- Pruning memberikan tingkat akurasi pada 10 penyakit.
- Nilai akurasi terkecil ditempati oleh Decompensatio sebesar 65% dan Dislipidimia sebesar 79 %. Sedangkan untuk nilai akurasi tertinggi ditempati oleh HHF sebesar 99% dan ALO 98%
- Untuk memunculkan klasifikasi penyakit lebih 10 kelas, maka diperlukan lebih dari 1 tree.
- Pola hidup yang salah dapat berakibat untuk bertumpuknya penyakit yang berkaitan dengan DM 2.

Menjalankan pola hidup sehat seperti rajin olahraga, tidak merokok, diet agar tidak terjadi obesitas adalah langkah yang tepat untuk mengurangi bertumpuknya penyakit yang berkaitan dengan DM 2.

DAFTAR PUSTAKA

- Dr. V. Karthikeyani, 2013. *Comparison A Performance of Data Mining Algorithms (CPDMA) In Prediction of Diabetes Disease*. IJCSE.
- S. Santhos Kumar, E. Ramaraj, 2012. *Analysis of Sequence Based Classifier Prediction for HIV Subtypes*. IJSE.
- Bendi Venkata Ramana, Prof. M. Surendra Prasad Babu, Prof. N. B. Venkatewarlu, 2011. *A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis*. IJDMS.
- Made Sumarwati, Waluyo Sejati, Roisca Dyah Pramitasari, 2008. *Eksplorasi Persepsi Penderita Tentang Faktor-Faktor Penyebab dan Dampak Penyakit Diabetes Melitus di Wilayah Puskesmas Purwokerto Barat, Kecamatan Purwokerto Barat, Kabupaten Banyuwang*. Jurnal Keperawatan Soedirman (The Soedirman Journal of Nursing).
- Handayani, 2012. *Modifikasi Gaya Hidup Dan Intervensi Farmakologis Dini Untuk Pencegahan Penyakit Diabetes Mellitus Tipe II*, Media Gizi Masyarakat Indonesia.
- Restyana Noor Fatimah. *Diabetes Mellitus Tipe 2*. Medical Faculty Lampung University. 2015.
- Jnanamurthy HK, Vishesh HV, Vishruth Jain, Preetham Kumar, Radhika M. Pai. *Discovery of Maximal Frequent Item Sets Using Subset Creation*. IJDKP. 2013 Vol.3 No.1.
- Liliana Swastina, "Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa", Jurnal Gema
- Sheu, Jyh-Jian. *An Efficient Two-phase Spam Filtering Methode Based on E-mails Categorization*. International Journal of Network Security, Vol. 8, No. 3, PP.334-343, Taiwan, May 2008.