

ANALISIS PERBANDINGAN ALGORITMA C4.5 DAN
REGRESSION TREE DALAM PENDETEKSIAN PENYAKIT
TUBERCULOSIS PARU

*Diajukan untuk Menyusun Tugas Akhir di Jurusan Teknik Informatika
Fakultas Ilmu Komputer Universitas Sriwijaya*



Oleh :

ACHMAD FEBRIAN
NIM : 09111002022

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2018**

LEMBAR PENGESAHAN TUGAS AKHIR

ANALISIS PERBANDINGAN ALGORITMA C4.5 DAN
REGRESSION TREE DALAM PENDETEKSIAN PENYAKIT
TUBERCULOSIS PARU

Oleh :

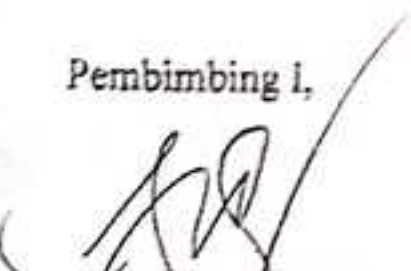
ACHMAD FEBRIAN

NIM : 09111002022

Palembang,

Juli 2018

Pembimbing I,


Rusdi Efendi M.Kom.

NIP-19820102 201102 1 201


Pembimbing II,


Denny Matthew Saputra M.SC.

NIP 1985051 201504 1 002

Menyetujui
Ketua Jurusan Teknik Informatika,




Rifkie Primartha, M.T.

NIP: 19770601 200912 1 004

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Kamis tanggal 24 Mei 2018 telah dilaksanakan Ujian Sidang Tugas Akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Achmad Febrian
NIM : 09111002022
Judul TA : Analisis Perbandingan Algoritma C4.5 dan *Regression Tree* Dalam Pendeteksian Penyakit Tuberculosis Paru

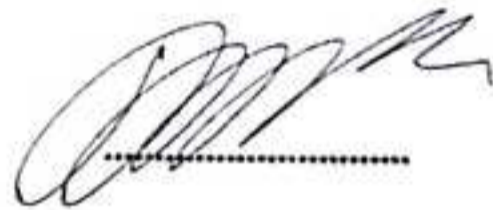
1. Ketua Penguji

Rusdi Efendi, M.Kom
NIP. 19820102 201102 1 201



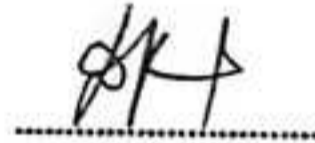
2. Sekretaris

Danny Matthew Saputra M.SC
NIP. 1985051 201504 1 002



3. Penguji I

Dian Palupi Rini, M.Kom, Ph.D
NIP. 19780223 200604 2 002



4. Penguji II

Kanda Januar Miraswan, M.T
NIK. 1671080901900006



Menyetujui,
Ketua Jurusan Teknik Informatika



Risde Primartha, M.T
NIP. 19770601 200912 1 004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Achmad Febrian
NIM : 0911002022
Program Studi : Teknik Informatika
Judul Skripsi : ANALISIS PERBANDINGAN
ALGORITMA C4.5 DAN
REGRESSION TREE DALAM
PENDETEKSIAN PENYAKIT
TUBERCULOSIS PARU
Hasil Pengecekan Software *iThenticate/Turnitin* : 2 %

Menyatakan bahwa Laporan Proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang,



Achmad Febrian
NIM. 0911002022

MOTTO DAN PERSEMBAHAN

MOTTO

- **Wisuda Setelah 14 Semester Adalah Kesuksesan Yang Tertunda**
- **Bersyukur Dengan Apa Yang Dimiliki Sekarang**

Kupersembahkan Karya Tulis Ini Kepada:

- **Kedua Orang Tua**
- **Keluarga Besarku**
- **INDEPENDENT**
- **Dosen Pembimbing dan Penguji**
- **Dosen Pengajar di Fasilkom**
- **Universitas Sriwijaya**

ANALISIS PERBANDINGAN ALGORITMA C4.5 DAN *REGRESSION TREE* DALAM PENDETEKSIAN PENYAKIT TUBERCULOSIS PARU

ACHMAD FEBRIAN

0911002022

ABSTRAK

Teknik pengklasifikasian telah banyak diterapkan dalam banyak area aplikasi untuk melakukan suatu prediksi. Salah satu teknik pengklasifikasian yang sering dipakai adalah pohon keputusan. Teknik ini banyak digunakan karena dapat menampilkan hasil pengklasifikasian yang mudah dimengerti manusia dalam bentuk mirip pohon. C4.5 dan CART merupakan algoritma dari pohon keputusan yang memiliki keunikan dalam pembentukan pohon. Dimana C4.5 akan menghasilkan pohon dengan *multi branch* dan CART akan menghasilkan pohon dengan *binary branch*. Kedua algoritma tersebut akan dibandingkan dalam pendeteksian penyakit Tuberculosis Paru, dimana menurut WHO penyakit ini merupakan salah satu dari sepuluh penyakit paling mematikan didunia pada tahun 2015, dimana Tuberculosis Paru bertanggung jawab terhadap penyebab kematian pada manusia lebih banyak daripada penyakit HIV dan malaria. Terdapat 102 data *training* yang digunakan dalam pembentukan pohon keputusan dan 30 data *testing* yang akan diuji pengklasifikasiannya menggunakan pohon yang telah dibentuk. Hasil dari pengujian menunjukkan, C4.5 memiliki tingkat akurasi dan waktu pemrosesan yang lebih cepat dari CART yaitu 86,66% dengan waktu yang diperlukan 31ms berbanding 73,33% dan 78ms saat tidak dilakukan pruning. Saat dilakukan pruning hasilnya 93,33% dengan waktu yang diperlukan 32ms untuk C4.5, sedangkan CART 80,00% dan 172ms. Pada pengujian menggunakan teknik *crossvalidate*, digunakan data *training* sebanyak 132 data dan jumlah *fold* sebanyak empat *fold*. Hasilnya menunjukkan tingkat akurasi dari C4.5 sebesar 93,93% dengan waktu 15ms, lebih baik dari hasil tingkat akurasi CART sebesar 90,15% dengan waktu 109ms.

Kata Kunci : C4.5, CART, Klasifikasi, Pohon Keputusan, Tuberculosis Paru.

COMPARATIVE ANALYSIS OF ALGORITHM C4.5 AND REGRESSION TREE IN DETECTING DISEASE OF LUNG TUBERCULOSIS

ACHMAD FEBRIAN

0911002022

ABSTRACT

The classification technique has been widely applied in many areas of the application to perform a prediction. One of the commonly used classification techniques is decision trees. This technique is widely used because it can display the results of the classification that is easily understood by humans in the form of a tree-like. C4.5 and CART are the algorithms of decision trees that are unique in the formation of trees. Where C4.5 will generate a multi-branch tree and CART will generate a tree with a binary branch. The two algorithms will be compared in the detection of Lung Tuberculosis, which according to the WHO is one of the ten most deadly diseases in the world by 2015, where Lung Tuberculosis is responsible for causing more human deaths than HIV and malaria diseases. There are 102 training data used in decision tree formation and 30 data testing to be tested for classification using established trees. The results of the test show that C4.5 has accuracy and processing time faster than CART that is 86,66% with time required 31ms compared to 73,33% and 78ms when not done pruning. When done pruning the result is 93.33% with time required 32ms for C4.5, while CART 80.00% and 172ms. In the test using crossvalidate technique, used data training as much as 132 data and the amount of fold as much as four fold. The results show the accuracy of C4.5 of 93.93% with a time of 15ms, better than the result of CART accuracy of 90.15% with time 109ms.

Keyword: C4.5, CART, Classification, Decision Tree, Lung Tuberculosis.

KATA PENGANTAR

Puji syukur kehadiran Allah SWT, atas limpahan Rahmat dan Karunia-Nya, sehingga penulis dapat merampungkan skripsi dengan judul “**Analisis Perbandingan Algoritma C4.5 dan Regression Tree Dalam Pendeteksian Penyakit Tuberculosis Paru**”. Tugas Akhir ini untuk memenuhi salah satu syarat menyelesaikan studi serta dalam rangka memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Dalam proses pengerjaan tugas kahir ini, penulis mengucapkan banyak terima kasih kepada pihak-pihak yang telah membantu dalam penyelesaian tugas akhir ini, diantaranya :

1. Ibuku Emi Artati dan Ayahku Hardi, yang telah hadir dalam setiap langkah penulis menyelesaikan pendidikan di Fakultas Ilmu Komputer Universitas Sriwijaya
2. Adik-adikku Mia Amalia dan Aaliyah Fahira yang senantiasa mengganggu dan rebut namun menghibur dalam setiap langkah yang dilampaui penulis
3. Si Sup, Gembul, dan Buntel kucing-kucing kesayangan yang selalu minta makan
4. Keluarga besar yang selalu mendoakan dan percaya dengan jalan yang ditempuh penulis
5. Bapak Jaidan Jauhari, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya
6. Bapak Rifkie Primartha, M.T. selaku Ketua Jurusan Fakultas Ilmu Komputer Universitas Sriwijaya
7. Bapak Rusdi Efendi M.Kom serta Bapak Danny Matthew Saputra M.SC selaku dosen pembimbing yang telah mengarahkan penulis dalam menyelesaikan tugas akhir ini.

8. Ibu Dian Palupi Rini, Ph.D serta Bapak Kanda Januar Miraswan, M.T selaku dosen penguji yang telah memberikan kritik dan saran terhadap penulis dalam proses menyelesaikan tugas akhir ini.
9. Rival Darmawan, Fitrah Rahmi Putri, Kurnia Andrea Kosim, Daniel dan Basni yang banyak membantu dalam proses penyelesaian tugas akhir yang ribet ini
10. Teman-teman seperjuanganku dari awal kuliah sampai akhir, walau ada yang mendahului, Ade, Hafis, Riyan, Imam, Eris, Erto, Devi, Las, Jujus serta teman-teman lain yang tidak bisa disebutkan satu persatu
11. Seluruh Pihak yang turut membantu baik secara langsung maupun tidak langsung dalam penyusunan dan penyempurnaan tugas akhir ini

Akhir kata penulis menyadari bahwa dalam penulisan skripsi ini masih jauh dari kesempurnaan. Karena itu, penulis memohon saran dan kritik yang sifatnya membangun demi kesempurnaannya dan semoga bermanfaat bagi kita semua.

Palembang,

2018

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	iv
DAFTAR TABEL	v
 BAB I PENDAHULUAN	
1.1 Latar Belakang	I-1
1.2 Rumusan Masalah	I-3
1.3 Tujuan Penelitian	I-3
1.4 Manfaat Penelitian.....	I-3
1.5 Batasan Masalah	I-4
1.6 Metodologi Penelitian.....	I-4
1.6.1 Jenis dan Sumber Data.....	I-4
1.6.2 Teknik Pengumpulan Data.....	I-4
1.7 Metode Pengembangan Perangkat Lunak.....	I-4
1.7.1 Fase Insepsi.....	I-5
1.7.2 Fase Elaborasi.....	I-5
1.7.3 Fase Kontruksi.....	I-5
1.7.4 Fase Transisi.....	I-6

BAB II TINJAUAN PUSTAKA

2.1	Penelitian Terkait	II-1
2.2	Landasan Teori	II-2
2.2.1.	<i>Data Mining</i>	II-2
2.2.2.	Klasifikasi	II-5
2.2.3.	Pohon Keputusan	II-6
2.2.4.	C4.5	II-8
2.2.5.	<i>Classification and Regression Tree</i>	II-9
2.2.6.	<i>Tree Pruning</i>	II-10
2.2.7	<i>Recall, Precision, dan f Measure</i>	II-11
2.2.8.	Tuberculosis Paru	II-12
2.3	Model Proses Pengembangan Perangkat Lunak.....	II-14
2.3.1.	<i>Rational Unified Process (RUP)</i>	II-14
2.3.2.	Struktur Proses RUP	II-15
2.3.3.	Fase RUP	II-16

BAB III METODOLOGI PENELITIAN

3.1	Analisis Masalah	III-1
3.1.1	Analisis Data.....	III-1
3.1.2	Analisis Proses Algoritma C4.5.....	III-4
3.1.3	Analisis Proses Algoritma <i>Classification and Regression Tree</i> ...	III-9
3.2	Analisis dan Perancangan Perangkat Lunak	III-16
3.2.1	Deskripsi Umum Perangkat Lunak	III-16
3.2.2	Spesifikasi Kebutuhan	III-16
3.2.2.1	Kebutuhan Fungsional.....	III-16
3.2.2.2	Kebutuhan Non-Fungsional.....	III-17
3.3	Pemodelan <i>Use Case</i>	III-17
3.3.1	Diagram <i>Use Case</i>	III-17
3.3.2	Definisi Aktor.....	III-18
3.3.3	Definisi <i>Use Case</i>	III-18
3.3.4	Skenario <i>Use Case</i>	III-20

3.3.5	Diagram Kelas Analisis.....	III-29
3.3.6	Diagram <i>Sequence</i>	III-34
3.4	Diagram Kelas	III-39
3.5	Perancangan Perangkat Lunak	III-40
3.5.1	Perancangan Antarmuka.....	III-40

BAB IV HASIL DAN PEMBAHASAN

4.1	Implementasi Perangkat Lunak	IV-1
4.1.1	Lingkungan Implementasi.....	IV-1
4.1.2	Implementasi Kelas.....	IV-2
4.1.3	Implementasi Antar Muka.....	IV-6
4.2	Pengujian Perangkat Lunak.....	IV-8
4.2.1	Lingkungan Pengujian.....	IV-8
4.2.2	Rencana Pengujian.....	IV-9
4.2.3	Kasus Uji.....	IV-11
4.3	Hasil Pengujian Perangkat Lunak.....	IV-16
4.3.1	Pengujian Terhadap Proses Algoritma dengan Data Test.....	IV-16
4.3.2	Pengujian Waktu Proses dengan Data Test.....	IV-21
4.3.3	Perhitungan <i>Precision</i> , <i>Recall</i> , dan <i>f Measure</i>	IV-19
4.4	Analisis Hasil	IV-24
4.4.1	Analisis Hasil Pengujian.....	IV-24
4.4.2	Perbandingan Hasil dengan Penelitian Sebelumnya.....	IV-26

BAB V KESIMPULAN DAN SARAN

5.1	Kesimpulan.....	V-1
5.2	Saran.....	V-1

DAFTAR PUSTAKA

DAFTAR GAMBAR

	Halaman
Gambar II-1 Proses Data Mining	II-3
Gambar II-2 Pohon Keputusan.....	II-7
Gambar II-3 Arsitektur <i>Rational Unified Process</i>	II-15
Gambar III-1 Flowchart C4.5.....	III-4
Gambar III-2 Flowchart <i>Classification and Regression Tree</i>	III-10
Gambar III-3 Diagram Use Case.....	III-18
Gambar III-4 Kelas Analisis Login.....	III-27
Gambar III-5 Kelas Analisis Diagnosa	III-27
Gambar III-6 Kelas Analisis Mencari Data Training.....	III-28
Gambar III-7 Kelas Analisis Mencari Data Testing	III-28
Gambar III-8 Kelas Analisis Uji Algoritma.....	III-29
Gambar III-9 Kelas Analisis Uji Algoritma <i>Pruned</i>	III-29
Gambar III-10 Kelas Analisis Menyimpan Hasil Uji Algoritma.....	III-30
Gambar III-11 Diagram <i>Sequence</i> Login.....	III-30
Gambar III-12 Diagram <i>Sequence</i> Diagnosa	III-31
Gambar III-13 Diagram <i>Sequence</i> Mencari Data Training.....	III-31
Gambar III-14 Diagram <i>Sequence</i> Mencari Data Testing.....	III-32
Gambar III-15 Diagram <i>Sequence</i> Uji Algoritma.....	III-32
Gambar III-16 Diagram <i>Sequence</i> Uji Algoritma <i>Pruned</i>	III-33
Gambar III-17 Diagram <i>Sequence</i> Menyimpan Hasil Uji Algoritma	III-33
Gambar III-18 Diagram Kelas Perangkat Lunak	III-34
Gambar III-19 Rancangan Antarmuka HalLogin	III-35
Gambar III-20 Rancangan Antarmuka HalUtama	III-36
Gambar III-21 Rancangan Antarmuka HalDiagnosa.....	III-37
Gambar III-22 Rancangan Antarmuka HalUjiAlgoritma	III-37
Gambar IV-1 Antarmuka HalLogin	IV-6
Gambar IV-2 Antarmuka HalUtama.....	IV-6
Gambar IV-3 Antarmuka HalDiagnosa	IV-7

Gambar IV-4 Antarmuka HalUjiAlgoritma	IV-7
Gambar IV-5 Hasil Uji Akurasi <i>Unpruned Tree</i>	IV-16
Gambar IV-6. Gambar Pohon Hasil Uji C4.5 <i>Unpruned Tree</i>	IV-17
Gambar IV-7. Gambar Pohon Hasil Uji CART <i>Unpruned Tree</i>	IV-17
Gambar IV-8 Hasil Uji Akurasi <i>Pruned Tree</i>	IV-18
Gambar IV-9. Gambar Pohon Hasil Uji C4.5 <i>Pruned Tree</i>	IV-19
Gambar IV-10. Gambar Pohon Hasil Uji CART <i>Pruned Tree</i>	IV-19
Gambar IV-11. Gambar Hasil Uji Akurasi <i>Crossvalidate</i>	IV-20

DAFTAR TABEL

	Halaman
Tabel III-1. Contoh Data	III-3
Tabel III-2. Nilai Entropy Keseluruhan	III-4
Tabel III-3. Nilai Entropie Masing- Masing Data.....	III-8
Tabel III-4. Nilai Gain.....	III-8
Tabel III-5. Kandidat Split	III-11
Tabel III-6. Nilai Optimal Split.....	III-13
Tabel III-7. Kebutuhan Fungsional	III-17
Tabel III-8. Kebutuhan Non-Fungsional.....	III-17
Tabel III-9. Definisi Aktor	III-18
Tabel III-10. Definisi Use Case	III-19
Tabel III-11. Skenario Use Case Login	III-20
Tabel III-12. Skenario Use Case Diagnosa	III-21
Tabel III-13. Skenario Use Case Memilih Data Training	III-21
Tabel III-14. Skenario Use Case Memilih Data Testing.....	III-22
Tabel III-15. Skenario Use Case Uji Algoritma	III-23
Tabel III-16. Skenario Use Case Uji Algoritma Pruned	III-24
Tabel III-17. Skenario Use Case Menyimpan Hasil Uji Algoritma.....	III-25
Tabel IV-1. Daftar Implementasi Kelas	IV-2
Tabel IV-2. Rencana Pengujian Use Case Login.....	IV-9
Tabel IV-3. Rencana Pengujian Use Case Diagnosa	IV-9
Tabel IV-4. Rencana Pengujian Use Case Memilih Data Training	IV-9
Tabel IV-5. Rencana Pengujian Use Case Memilih Data Testing	IV-9
Tabel IV-6. Rencana Pengujian Use Case Uji Algoritma.....	IV-10
Tabel IV-7. Rencana Pengujian Use Case Uji Algoritma Pruned	IV-10
Tabel IV-8. Rencana pengujian <i>use case Crossvalidate</i>	IV-10

Tabel IV-9. Rencana pengujian <i>use case</i> Simpan Hasil	IV-10
Tabel IV-10. Kasus Uji <i>Use Case</i> Diagnosa <i>Tuberculosis</i> Paru	IV-11
Tabel IV-11. Tabel Waktu Pemrosesan.....	IV-21
Tabel IV-12. Tabel <i>Confusion Matrix Unpruned Tree</i>	IV-21
Tabel IV-13. Tabel <i>Confusion Matrix Pruned Tree</i>	IV-22
Tabel IV-14. Tabel <i>Confusion Matrix Crossvalidate</i>	IV-23
Tabel IV-15. Tabel f Measure.....	IV-23
Tabel IV-16. Analisis Hasil Pengujian Data Test.....	IV-24
Tabel IV-17. Analisis Hasil Pengujian <i>Crossvalidate</i>	IV-25

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tugas klasifikasi dan regresi adalah memprediksi sebuah pekerjaan yang melibatkan pembuatan model untuk memprediksi suatu target dari variable dependen atau variabel independen (Sasikala, 2014). Dengan klasifikasi, model yang dihasilkan akan bisa memprediksi sebuah kelas untuk data yang diberikan tergantung pada informasi yang dipelajari dari data sebelumnya (Ramageri, 2010).

Menurut Han, Kamber, & Pei (2012), teknik pengklasifikasian telah digunakan di banyak area aplikasi seperti kedokteran, manufaktur dan produksi, analisis keuangan, astronomi, dan biologi molekular, yaitu dengan menggunakan teknik pohon keputusan. Teknik pohon keputusan telah banyak digunakan untuk membangun model klasifikasi karena model semacam itu sangat mirip dengan penalaran manusia dan mudah dipahami (Kotsiantis, 2013).

Sebelumnya teknik pohon keputusan sudah sering dipakai oleh peneliti, salah satunya oleh Sasikala (2014), dimana algoritma C4.5, dan CART dibandingkan untuk melihat tingkat akurasi dan kecepatan pada prediksi catatan medis beberapa penyakit, yaitu *arryhtmia*, diabetes, kanker payudara, jantung, dan *thyroid*.

Penelitian lainnya dilakukan oleh Shankar, Sugumaran, Karthikeyan, & Vijayaram (2016). Pada penelitian tersebut mereka menggunakan algoritma C4.5

yang merupakan bagian dari teknik pohon keputusan untuk mendiagnosis penyakit hepatitis.

Berdasarkan penelitian sebelumnya diatas, penulis melihat bahwa adanya perbedaan data yang digunakan pada algoritma pohon keputusan dapat menghasilkan tingkat akurasi dan kecepatan yang berbeda pada algoritma. Sehingga penulis berinisiatif menguji algoritma tersebut dengan data yang berbeda untuk melihat tingkat akurasi dan waktu pemrosesan dari kedua algoritma pada data uji penyakit yang lain.

Salah satu penyakit yang menjadi perhatian pemerintah dan ikatan dokter Indonesia adalah *Tuberculosis* (TB) paru. TB paru merupakan salah satu dari sepuluh penyakit mematikan didunia pada tahun 2015, dimana Tuberculosis Paru bertanggung jawab terhadap penyebab kematian pada manusia lebih banyak daripada penyakit HIV dan malaria (WHO, 2016). Tidak adanya alat diagnosa yang murah dan teknologi diagnostis yang akurat adalah masalah utama di daerah yang memiliki kekurangan ekonomi, di mana apusan dahak mikroskopi banyak digunakan untuk mendeteksi penyakit (Perkins, Roscigno and Zumla, 2006).

Sebelumnya beberapa peneliti telah menerapkan metode *data mining* dalam studi kasus TB paru, salah satunya Ali, Elfaki, & Jawawi (2012) menerapkan algoritma *Naïve Bayes* dan *Bayesian Network*. Selain itu Nagabhushanam, Naresh, Raghunath, & Praveen (2013) memprediksi penyakit TB paru menggunakan algoritma *Adaptive Neuro Fuzzy Inference System*, *Multilayer Perceptron*, dan *Partial Decision Trees*.

Pada penelitian kali ini, penulis akan membandingkan algoritma *C4.5* dan *CART* yang merupakan bagian dari teknik pohon keputusan dalam sistem pendeteksian penyakit TB paru. Perbedaan dalam cara menghasilkan pohon keputusan pada kedua algoritma diprediksi akan menghasilkan kecepatan dan tingkatan akurasi yang berbeda. Penelitian ini bertujuan untuk melihat dan membandingkan seberapa besar akurasi dan kecepatan yang dihasilkan dalam sistem.

1.2 Rumusan Masalah

Bagaimana membuat sebuah sistem pendeteksian TB paru yang dapat membandingkan algoritma *C4.5* dan *CART* untuk mengukur kecepatan dan akurasi dari algoritma, serta membantu pihak rumah sakit mendeteksi penyakit TB paru dengan cepat dan akurat

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah :

1. Menerapkan algoritma *C4.5* dan *CART* pada sistem pendeteksian TB paru.
2. Memperoleh informasi tingkatan akurasi dan kecepatan dari algoritma *C4.5* dan *CART* dalam sistem pendeteksian TB paru.

1.4 Manfaat Penelitian

Manfaat penelitian ini adalah :

1. Mempercepat proses pendiagnosaan pasien tersangka penyakit TB paru.
2. Mendapatkan model prediksi menggunakan algoritma *C4.5* dan *CART* dalam pendeteksian penyakit TB paru.

1.5 Batasan masalah

Dalam penelitian ini, diterapkan beberapa batasan sebagai berikut :

1. Masukan berupa data gejala penyakit yang dialami pasien TB paru.
2. Data yang dimasukkan harus lengkap diisi.
3. *Data Training* berjumlah 102 data dan *Data Testing* berjumlah 30 data dengan total keseluruhan data berjumlah 132.
4. Jenis File yang digunakan menggunakan file excel dengan format file .csv

1.6 Metodologi Penelitian

1.6.1 Jenis dan Sumber Data

Terdapat dua jenis data yang digunakan pada penelitian kali ini. Data primer didapatkan dari hasil wawancara langsung dengan dr. Mulyadi. Sedangkan data sekunder didapatkan dari Rumah Sakit Umum Daerah Sungai Lilin dalam bentuk arsip catatan medik pasien.

1.6.2 Teknik Pengumpulan Data

Data primer berisikan wawancara dengan dr. Mulyadi untuk menentukan variable gejala penyakit yang diperlukan. Data sekunder yang didapatkan masih dalam berbentuk tulisan manual di kertas dipindahkan kedalam file .csv untuk pemrosesan oleh perangkat lunak. Isi dari data yang didapatkan berisikan gejala-gejala penyakit dari tuberculosis paru dan hasil diagnosanya.

1.7 Metode Pengembangan Perangkat Lunak

Pada penelitian kali ini, metode yang akan digunakan pada pengembangan perangkat lunak adalah metode RUP (*Rational Unified Process*). Terdapat empat fase pada metode ini, yaitu:

1.7.1 Fase Insepsi

Pada fase ini hal-hal yang dilakukan adalah:

1. Memahami konsep dari algoritma C4.5 dan CART dalam sistem data mining untuk diterapkan pada perangkat lunak
2. Menentukan kebutuhan dari perangkat lunak dan keras yang digunakan
3. Membuat gambaran fitur-fitur yang diperlukan pada perangkat lunak untuk mendapatkan hasil yang diinginkan
4. Membuat gambaran diagram *use case* dari perangkat lunak

1.7.2 Fase Elaborasi

Pada fase ini hal-hal yang dilakukan adalah:

1. Melakukan perancangan terhadap fungsi-fungsi yang berkenaan terhadap perangkat lunak
2. Menggambarkan model kelas analisis, diagram kelas, dan sequence diagram dari perangkat lunak.

1.7.3 Fase Konstruksi

Pada fase ini hal-hal yang dilakukan adalah:

1. Melakukan perancangan antarmuka dari perangkat lunak.
2. Melakukan pengkodean secara keseluruhan dengan menerapkan algoritma yang dipakai.

1.7.4 Fase Transisi

Pada fase ini hal-hal yang dilakukan adalah melakukan pengujian dengan metodologi *black box*. Dan melakukan analisis terhadap hasil pengujian, serta melakukan perbaikan jika ditemukan kesalahan.

DAFTAR PUSTAKA

- Ali, A., Elfaki, M., & Jawawi, D. N. A. (2012). Using Naïve Bayes and Bayesian Network for Prediction of Potential Problematic Cases in Tuberculosis. *International Journal of Informatics and Communication Technology (IJ-ICT)*, 1(2), 63–71. Retrieved from <http://iaesjournal.com/online/index.php/IJICT>
- Candelieri, A. *et al.* (1996) ‘Ch-16 Data Mining in Neurology’, *Knowledge-Oriented Applications in Data Mining*.
- Campbell, I. A., & Bah-Sow, O. (2006). Pulmonary tuberculosis: diagnosis and treatment. *BMJ (Clinical Research Ed.)*, 332(7551), 1194–1197. <https://doi.org/10.1136/bmj.332.7551.1194>
- Delogu, G., Sali, M., & Fadda, G. (2013). The biology of mycobacterium tuberculosis infection. *Mediterranean Journal of Hematology and Infectious Diseases*, 5(1). <https://doi.org/10.4084/mjhid.2013.070>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37. <https://doi.org/10.1609/aimag.v17i3.1230>
- Ferdinand, S., Valétudie, G., Sola, C., & Rastogi, N. (2004). Data mining of Mycobacterium tuberculosis complex genotyping results using mycobacterial interspersed repetitive units validates the clonal structure of spoligotyping-defined families. *Research in Microbiology*, 155(8), 647–654. <https://doi.org/10.1016/j.resmic.2004.04.013>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. *Journal of Chemical Information and Modeling* (Vol. 3). <https://doi.org/10.1017/CBO9781107415324.004>
- Jahantigh, F. F., & Ameri, H. (2017). Evaluation of TB Patients Characteristics Based on Predictive Data Mining Approaches. *Journal of Tuberculosis Research*, 5(1), 13–22. <https://doi.org/10.4236/jtr.2017.51002>
- Kotsiantis, S. B. (2013) ‘Decision trees: A recent overview’, *Artificial Intelligence Review*, 39(4), pp. 261–283. doi: 10.1007/s10462-011-9272-4.
- Nagabhushanam, D., Naresh, N., Raghunath, A., & Praveen, K. K. (2013). Prediction of Tuberculosis Using Data Mining Techniques on Indian Patient ’

s Data. *International Journal of Computer Science and Technology*, 4(SPL-4), 262–265.

Patil, D. D., Wadhai, V., & Gokhale, J. (2010). Evaluation of decision tree pruning algorithms for complexity and classification accuracy. *Internation Journal of Computer Applications*, 11(2), 23–30. Retrieved from <http://core.kmi.open.ac.uk/download/pdf/870101.pdf>

Patil, T. R. (2013). Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification. *International Journal Of Computer Science And Applications*, ISSN: 0974-1011, 6(2), 256–261.

Perkins, M. D., Roscigno, G. and Zumla, A. (2006) ‘Progress towards improved tuberculosis diagnostics for developing countries’, *Lancet*, 367(9514), pp. 942–943. doi: 10.1016/S0140-6736(06)68386-4.

Ramageri, M. (2010) ‘Data Mining Techniques and Applications’, *Indian Journal of Computer Science and Engineering*, 1(4), pp. 301–305.

Sasikala, C. (2014). Comparison and Evaluation of Decision Tree Algorithms using Medical Records, 2(3), 286–290.

Shankar, V. *et al.* (2016) ‘Diagnosis of Hepatitis using Decision tree algorithm’, 8(3), pp. 1414–1419.

Singh, D., Choudhary, N., & Samota, J. (2013). Analysis of Data Mining Classification with Decision Tree Technique. *Global Journal of Computer Science and Technology*, 13(13).

Singh, S., & Giri, M. (2014). Comparative study ID3, cart and C4 . 5 Decision tree algorithm: a survey. *International Journal of Advanced Information Science and Technology (IJAIST)*, 27(27), 97–103. <https://doi.org/10.15693/ijaist/2014.v3i7.47-52>

Țăranu, I. (2015) ‘Data mining in healthcare: decision making and precision’, *Database Systems Journal*, 5(4), pp. 33–40.

WHO (2016) ‘Global Tuberculosis Report 2016’, *Cdc 2016*, (Global TB Report 2016), p. 214. doi: ISBN 978 92 4 156539 4.