

PENERAPAN ALGORITMA C4.5 UNTUK PREDIKSI PENYAKIT PARU-PARU MENGGUNAKAN RAPIDMINER

Fazrin Meila Azzahra Sofyan, Apriade Voutama, Yuyun Umaidah

Program Studi Sistem Informasi S1, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang
HS. Ronggo Waluyo, Karawang, Jawa Barat 41361, telp/fax (0267) 641177, Indonesia
2010631250044@student.unsika.ac.id

ABSTRAK

Paru-paru adalah sistem pernapasan yang menjadi tempat bertukarnya oksigen dan karbon dioksida. Namun, bukan hal yang mustahil jika paru-paru dapat mengalami gangguan sehingga dapat menyebabkan penyakit paru-paru. Merokok menjadi penyebab utama penyakit paru-paru, data dari World Health Organization menunjukkan 40% kematian akibat merokok dan setidaknya 8 juta orang terbunuh setiap tahunnya. Menurut *Lung Foundation Australia*, penyakit paru-paru dapat menyerang siapa saja entah itu perokok pasif atau aktif, pria, wanita dan anak-anak. Penelitian ini menggunakan metode KDD (*Knowledge Discovery in Database*) yang memiliki lima tahap yaitu *selection*, *preprocessing*, *transformation*, *data mining* dan *evaluation* serta menggunakan algoritma C4.5 untuk mendapatkan model prediksi yang dapat memprediksi data pasien penyakit paru-paru. Penelitian ini bertujuan untuk mendapatkan nilai akurasi, *recall* dan *precision* dengan menggunakan algoritma C4.5. Adapun data yang digunakan pada penelitian ini diperoleh dari Kaggle yang berisi 30.000 data dengan 11 atribut di dalamnya. RapidMiner digunakan sebagai *tools* untuk menguji dataset pasien yang digunakan sehingga menghasilkan sebuah pohon keputusan (*decision tree*) dengan nilai akurasi sebesar 89.77%, *recall* sebesar 78.61%, dan *precision* sebesar 100% yang diperoleh dari split data 90% (*data training*) – 10% (*data testing*).

Kata kunci: Paru-paru, Merokok, Algoritma C4.5, Kaggle, RapidMiner, KDD

1. PENDAHULUAN

Paru-paru merupakan sistem pernapasan pada manusia yang berperan penting untuk memenuhi kebutuhan oksigen dalam tubuh. Selain itu, paru-paru memiliki fungsi sebagai tempat bertukarnya oksigen dari udara dengan karbon dioksida dari darah. Pada beberapa kondisi, paru-paru dapat mengalami gangguan yang berakibat buruk pada kinerja sistem pernapasan [1], jika paru-paru tidak berfungsi dengan baik maka akan menyebabkan timbulnya suatu penyakit.

Penyakit paru-paru mempengaruhi sistem pernapasan yang dapat menyebabkan kerusakan jangka pendek maupun jangka panjang [2], faktor timbulnya penyakit paru-paru disebabkan oleh zat-zat berbahaya seperti rokok yang mengandung toksin dan karsinogen [3]. Dilansir dari *Lung Foundation Australia*, penyakit paru-paru dapat menyerang pria, wanita, anak-anak, perokok, mantan perokok, maupun yang tidak pernah merokok juga memiliki resiko yang sama akan terserang penyakit ini [4].

Di Indonesia, rokok semakin digemari dari kalangan muda maupun tua. Dari golongan yang kurang mampu sekalipun rokok semakin banyak penikmatnya, hal itu bermula dari pergaulan yang berkelanjutan menjadi candu terhadap nikotin ini. Dampak yang tidak dirasakan secara langsung membuat perokok mengabaikan kesehatan tentang bahaya rokok [5]. Mengutip *World Health Organization* (WHO), lebih dari 40% kematian akibat penggunaan dari tembakau yang menyerang kesehatan paru-paru. Tembakau membunuh

setidaknya 8 juta orang setiap tahun, data pada tahun 2017 melaporkan bahwa 3,3 juta pengguna tembakau dan orang yang terpapar asap rokok terbunuh karena penyakit paru-paru [6].

Penelitian sebelumnya mengenai “Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5” menggunakan 201 data lalu membaginya menjadi 40 data training dan 161 data testing dan menggunakan *tools RapidMiner*. Hasil dari penerapan algoritma C4.5 untuk mengklasifikasikan gangguan psikologis menghasilkan nilai akurasi untuk *data training* sebesar 57.50% untuk *mean precision* dan *classification error* sebesar 0%. Sedangkan data testing menghasilkan nilai akurasi sebesar 72.67%, yaitu *precision* sebesar 72.67% dan *recall* sebesar 100%. Hasil prediksi depresi mayor adalah 73%, masyarakat mengalami GAD 25% masyarakat, dan masyarakat yang memiliki gangguan konversi 2% [7].

Penelitian lainnya mengenai “Perbandingan Kinerja Algoritma C4.5 dan Naïve Bayes untuk Klasifikasi Penerima Beasiswa” menggunakan dataset sebanyak 164 *record* dengan atribut yang digunakan adalah semester, IPK, Prestasi ko/ekstra kurikuler, Penghasilan orang tua, Biaya listrik, dan Jumlah tanggungan. Model klasifikasi menggunakan *tools RapidManer* dengan *10-fold cross*. Perbandingan antara algoritma C4.5 dan Naïve Bayes menghasilkan tingkat akurasi sebesar 96.4% untuk algoritma C4.5 dan 95.11% untuk algoritma Naïve Bayes [8].

Berdasarkan penjelasan yang sudah diuraikan di atas, teknik data mining yang digunakan pada penelitian ini menggunakan metode *Decision Tree* melalui algoritma C4.5. Dengan menggunakan algoritma C4.5 ini, diharapkan akan menghasilkan akurasi yang tepat untuk memprediksi penyakit paru-paru.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining adalah teknik pengolahan suatu pengetahuan yang didasarkan pada sebuah *big data*, data yang biasa digunakan diambil dari *database*, *warehouse data*, *web* dan lain-lain untuk diproses menjadi informasi yang menarik. *Data mining* sendiri banyak digunakan di beberapa bidang untuk membantu peningkatan dalam hal pengetahuan, selain itu juga data mining digunakan untuk meningkatkan penjualan [7]. Cabang ilmu lain mendukung data mining yang mencakup statistik, teknologi basis data, *machine learning*, sistem pakar, algoritma paralel, algoritma genetika, pengenalan pola, visualisasi data, dan lain sebagainya [8].

2.2. Algoritma C4.5

Algoritma C4.5 merupakan *supervised learning* yang membutuhkan pra-klasifikasi variabel sasaran. Algoritma ini membuat struktur pohon keputusan (*Decision Tree*) yang setiap simpul internal (simpul bukan daun) menguji setiap atribut, label kelas diambil dari masing-masing simpul daun (atau simpul terminal), sedangkan nilai dihasilkan dari masing-masing cabang. Data training harus dipersiapkan terlebih dahulu sebelum di proses lalu atribut akan dievaluasi dengan menghitung nilai *Gain* dan *Entropy* [9]. Rumus 1 adalah rumus untuk menghitung nilai *Entropy*.

$$Entropy = \sum_{i=1}^n -p_i \log_2(p_i) \quad (1)$$

Keterangan:

S = Himpunan kasus

n = Jumlah partisi S

p_i = Proporsi S_i terhadap S

Untuk menghitung nilai *Gain* dapat menggunakan Rumus 2

$$Gain(S, A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} Entropy(S_i) \quad (2)$$

2.2. RapidMiner

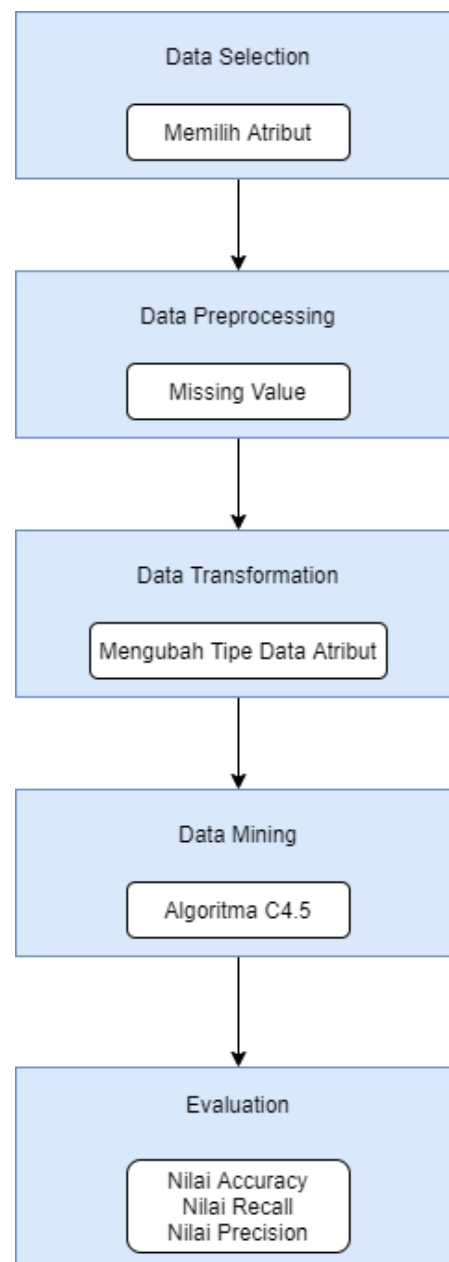
RapidMiner adalah perangkat lunak (*software*) yang bersifat *open source* dan dapat diakses oleh siapa saja. Menggunakan bahasa pemrograman Java sebagai proses pengoperasiannya, *RapidMiner* kerap kali menjadi solusi dalam hal menganalisa pada data *processing* dengan berbagai teknik yang dilakukan diantaranya teknik deskriptif dan prediksi [10].

3. METODE PENELITIAN

Penelitian ini menggunakan tahapan atau metode KDD (*Knowledge Discovery in Database*) yang memiliki lima tahapan diantaranya ada *Data Selection*, *Data Preparation*, *Data Transformation*, *Data Mining*, dan *Evaluation*.

3.1. KDD (*Knowledge Discovery in Database*)

KDD adalah proses untuk mencari informasi yang memiliki nilai, mudah dipahami dan baru dari penyimpanan data yang besar dan kompleks [9]. KDD sendiri merupakan metode yang digunakan untuk menghasilkan sebuah informasi melalui data yang sudah ada [11].



Gambar 1. Tahapan KDD

a. Data Selection

Tahapan pertama pada metode KDD ini adalah *data selection*, tahapan ini melakukan pengambilan dan pemilihan data serta melakukan pelabelan data [12] untuk diolah menjadi informasi yang berguna.

b. Data Preprocessing

Tahap ini dilakukan untuk mengurangi atribut yang tidak terlalu penting dalam melakukan proses klasifikasi [13].

c. Data Transformation

Tahap ini mengembangkan data dengan tujuan supaya data lebih baik lagi dan siap pada tahap pemodelan data mining [14].

d. Data Mining

Tahap ini menerapkan proses pengolahan data menggunakan metode atau algoritma tertentu [15], pada penelitian ini menggunakan algoritma C4.5.

e. Evaluation

Tahap yang terakhir yaitu *evaluation*, dimana *tahap* ini adalah keberhasilan dari model yang digunakan seperti menghasilkan nilai akurasi pada data diuji.

4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan metode algoritma C4.5 dengan alur proses seperti *data selection*, *data preprocessing*, *data transformation*, dan *evaluation*. Alur tersebut akan digunakan sebagai acuan dalam memprediksi gejala terserang penyakit paru-paru dengan hasil akhir yang didapat seperti nilai *accuracy*, *precision* dan *recall*.

4.1. Proses Pengumpulan Data

Dalam penelitian ini, proses pengambilan data menggunakan teknik pengambilan data sekunder yang sudah tersedia pada salah satu situs penyedia dataset yaitu *Kaggle*. Data sekunder digunakan karena sumber data sudah ada sehingga tidak perlu melakukan observasi, wawancara maupun *survey* untuk mendapatkan datanya [16]. Dataset yang digunakan dapat diakses melalui link <https://www.kaggle.com/datasets/akbarmuharram/dataset-predic-terkena-penyakit-paru-paru>. Tabel 1 di bawah merupakan atribut keseluruhan dari dataset yang diperoleh dari *Kaggle*.

Tabel 1. Dataset Pasien

Atribut	Tipe Data	Keterangan
Usia	Kategorik	Tua /Muda
Jenis_Kelamin	Kategorik	Pria / Wanita
Merokok	Kategorik	Pasif / Aktif
Bekerja	Kategorik	Tidak / Ya
Rumah_Tangga	Kategorik	Tidak / Ya
Aktivitas_Begadang	Kategorik	Tidak / Ya
Aktivitas_Olahraga	Kategorik	Sering / Jarang
Asuransi	Kategorik	Tidak / Ada
Penyakit_Bawaan	Kategorik	Tidak / Ada
Hasil	Kategorik	Ya / Tidak

Di bawah ini merupakan deskripsi pada Tabel 1.

a. Usia

Atribut “Usia” adalah data usia pasien yang ada pada dataset ini yang diberi nilai Tua dan Muda.

b. Jenis_Kelamin

“Jenis_Kelamin” adalah atribut yang mencakup jenis kelamin pasien dengan nilai Pria dan Wanita.

c. Merokok

“Merokok” merupakan atribut yang diberikan kepada pasien sebagai status apakah merupakan perokok aktif atau pasif, nilai yang diberikan adalah Pasif dan Aktif.

d. Bekerja

“Bekerja” merupakan atribut yang diberikan kepada pasien sebagai status apakah pasien bekerja atau tidak, nilai yang diberikan adalah Ya dan Tidak.

e. Rumah_Tangga

“Rumah_Tangga” merupakan atribut yang diberikan kepada pasien sebagai status apakah pasien sudah menikah atau belum, nilai yang diberikan adalah Ya dan Tidak.

f. Aktivitas_Begadang

“Aktivitas_Begadang” merupakan atribut yang diberikan kepada pasien apakah pasien sering begadang atau tidak, nilai yang diberikan adalah Ya dan Tidak.

g. Aktivitas_Olahraga

“Aktivitas_Olahraga” merupakan atribut yang diberikan kepada pasien apakah pasien sering melakukan olahraga atau tidak, nilai yang diberikan adalah Sering dan Jarang.

h. Asuransi

“Asuransi” merupakan atribut apakah pasien memiliki asuransi atau tidak, nilai yang diberikan adalah Ada dan Tidak

i. Penyakit_Bawaan

Penyakit_Bawaan” merupakan atribut yang diberikan kepada pasien apakah pasien memiliki riwayat penyakit bawaan atau tidak, nilai yang diberikan adalah Ada dan Tidak.

j. Hasil

Atribut “Hasil” merupakan label yang mencakup pasien yang terkena penyakit paru-paru dengan nilai Ya (terkena penyakit paru-paru) dan Tidak (Tidak terkena penyakit paru-paru).

3.2. Data Selection

Tahap ini dilakukan karena ada beberapa atribut yang tidak terlalu penting untuk digunakan, proses ini dilakukan untuk memilih atribut yang sesuai pada prediksi gejala penyakit paru-paru. Tabel 2 di bawah merupakan atribut yang akan digunakan dan diolah pada software RapidMiner.

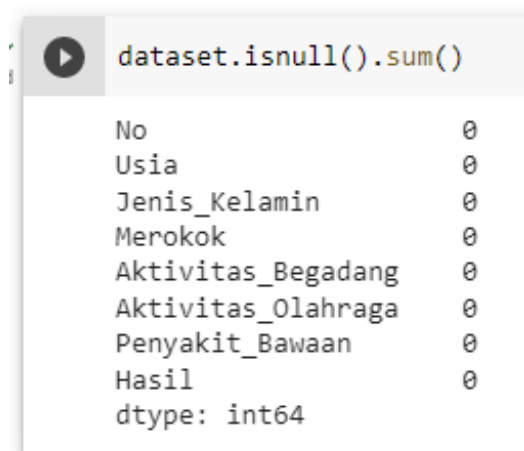
Tabel 2. Atribut Terpilih

No	Atribut Terpilih	Keterangan
1	Usia	Atribut Fitur
2	Jenis_Kelamin	Atribut Fitur
3	Merokok	Atribut Fitur
4	Aktivitas_Begadang	Atribut Fitur
5	Aktivitas_Olahraga	Atribut Fitur
6	Penyakit_Bawaan	Atribut Fitur
7	Hasil	Atribut Label

Atribut yang terpilih adalah Usia, Jenis_Kelamin, Merokok, Aktivitas_Begadang, Aktivitas_Olahraga, Penyakit_Bawaan, dan Hasil. Hal itu dikarenakan atribut tersebut sesuai dengan gejala terserang penyakit paru-paru. Atribut tersebut akan diolah menggunakan algoritma C4.5 dan software *RapidMiner* yang akan menghasilkan pohon keputusan serta akurasi untuk memprediksi penyakit paru-paru.

3.3. Data Preprocessing

Tahap ini dilakukan untuk memastikan pada dataset yang dipakai tidak ada *Missing Value*, hal ini bertujuan untuk mempermudah dalam pengolahan data pada software *RapidMiner*. Pengecekan *Missing Value* dapat dilakukan dengan menggunakan *Google Colaboratory*.



Gambar 2. Cek Missing Value

3.4. Data Transformation

Tahap pada *transformation* ini adalah mengubah data ke dalam format yang sesuai untuk selanjutnya dapat diolah pada *data mining*, proses pengubahan

pada dataset ini menggunakan *RapidMiner* dengan mengubah tipe data dari *polynomial* menjadi *binomial* pada atribut yang digunakan seperti yang ditunjukkan pada Gambar 3.

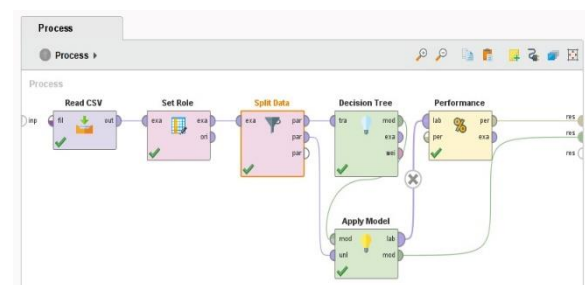
	Usia binomial	Jenis Kela... binomial	Merokok binomial	Aktivitas B... binomial	Aktivitas Ol... binomial	Penyakit B... binomial
1	Tua	Pria	Pasif	Ya	Sering	Tidak
2	Tua	Pria	Aktif	Ya	Jarang	Ada
3	Muda	Pria	Aktif	Ya	Jarang	Tidak
4	Tua	Pria	Aktif	Tidak	Jarang	Ada
5	Muda	Wanita	Pasif	Tidak	Sering	Ada
6	Muda	Wanita	Pasif	Tidak	Sering	Tidak
7	Tua	Wanita	Pasif	Tidak	Sering	Ada
8	Muda	Pria	Aktif	Ya	Sering	Tidak
9	Tua	Wanita	Aktif	Ya	Jarang	Ada
10	Muda	Wanita	Pasif	Ya	Jarang	Ada
11	Tua	Wanita	Pasif	Tidak	Sering	Ada

Gambar 3. Transformation pada Dataset

3.5. Data Mining

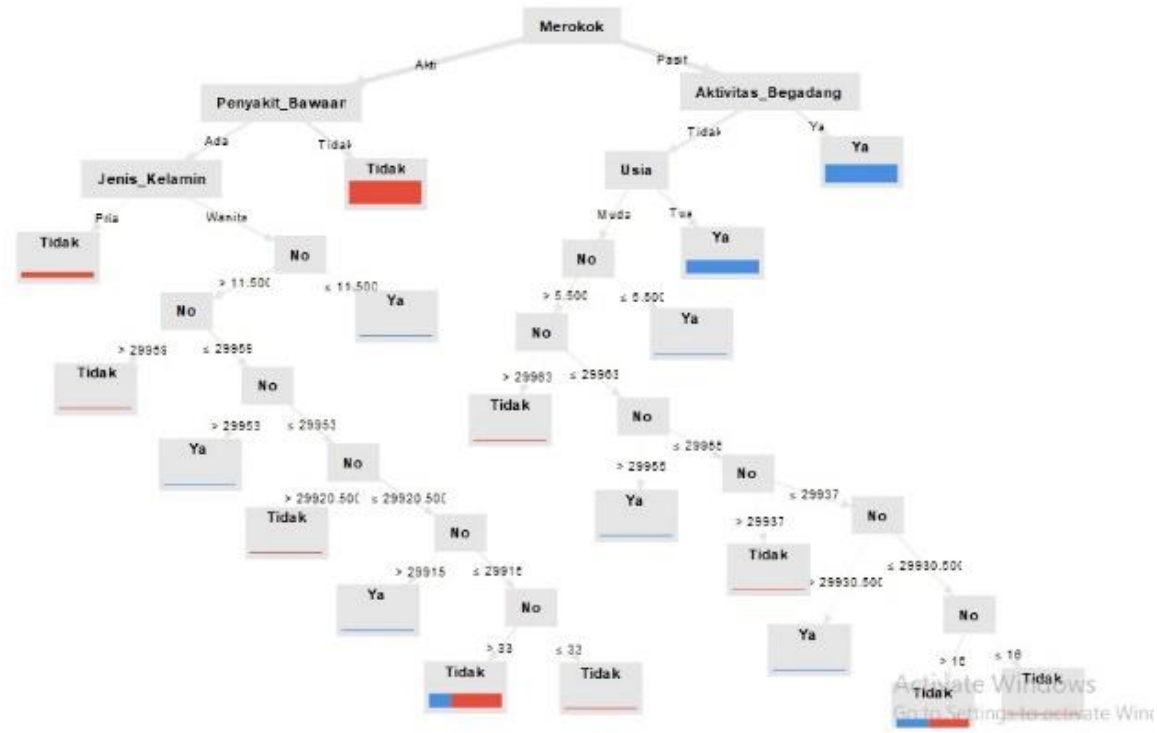
Algoritma C4.5 digunakan dalam penerapan data mining ini dengan tujuan untuk memprediksi pasien penyakit paru-paru berdasarkan dataset yang sudah melewati tahap *selection* dan *transformation*.

3.6. Pengujian Menggunakan RapidMiner



Gambar 4. Pengujian Menggunakan RapidMiner

Gambar 4 menjelaskan tentang proses pengujian menggunakan *RapidMiner* dimulai dengan melakukan import dataset kemudian *Set Role* digunakan dalam pemilihan label pada atribut “Hasil” untuk diprediksi, *Split Data* dibutuhkan untuk membagi data training dan data testing, kemudian memilih *Apply Model* yang dimana model *Decision Tree* telah dipilih sebagai pengujian ini. *Performance* digunakan untuk menghitung akurasi yang diperoleh dari hasil prediksi, semua *socket* tersebut dihubungkan sehingga menghasilkan pohon keputusan (*Decision Tree*) pada prediksi pasien penyakit paru-paru yang ditunjukkan pada Gambar 5.



Gambar 5. Model Algoritma C4.5 Prediksi Pasien Penyakit Paru-Paru

Model Algoritma C4.5 ini menggunakan 7 atribut terpilih yang dinilai sesuai dengan gejala pada prediksi penyakit paru-paru seperti Usia, Jenis_Kelamin, Merokok, Aktivitas_Begadang, Aktivitas_Olahraga, Penyakit_Bawaan dan Hasil.

3.7. Nilai Accuracy

Nilai *accuracy* diperoleh dari hasil membagi data (*split data*) menjadi *data training* dan *data testing*, berikut Tabel 3 yang merupakan perbandingan dari beberapa *split data* yang diuji dari hasil nilai akurasi algoritma C4.5.

Tabel 3. Nilai Accuracy

Split Data (Training – Testing)	Accuracy Algoritma C4.5
90 – 10	89.77%
80 – 20	89.30%
70 – 30	89.27%
60 – 40	89.41%
50 – 50	89.25%

3.8. Nilai Recall

Nilai *Recall* yang diperoleh pada pengolahan data melalui proses *split data* menggunakan algoritma C4.5 pada dataset prediksi penyakit paru dengan perbandingan beberapa *split data* yang diuji yang ditunjukkan pada Tabel 4.

Tabel 4. Nilai Recall

Split Data (Training – Testing)	Recall
90 – 10	78.61%
80 – 20	77.63%
70 – 30	77.59%
60 – 40	77.93%
50 – 50	77.69%

3.9. Nilai Precision

Nilai *Precision* yang diperoleh pada pengolahan data melalui proses *split data* menggunakan algoritma C4.5 pada dataset prediksi terkena penyakit paru dengan beberapa *split data* yang diuji yang ditunjukkan pada Tabel 5.

Tabel 5. Nilai Precision

Split Data (Training – Testing)	Precision
90 – 10	100%
80 – 20	100%
70 – 30	99.97%
60 – 40	99.91%
50 – 50	99.79%

Berdasarkan Tabel 3, 4 dan 5 di atas pembagian *split data* 90% - 10% menghasilkan nilai yang tinggi dari nilai *accuracy*, *recall*, dan *precision*. Dapat dilihat pada Gambar 6 yang merupakan hasil dari *split data* melalui *RapidMiner*.

accuracy: 89.77%

	true Ya	true Tidak	class precision
pred. Ya	1128	0	100.00%
pred. Tidak	307	1565	83.60%
class recall	78.61%	100.00%	

Gambar 6. Hasil Nilai Accuracy, Recall dan Precision

5. KESIMPULAN DAN SARAN

Paru-paru merupakan salah satu organ yang memiliki peran sebagai tempat bertukarnya oksigen dan karbon dioksida, namun ada kalanya paru-paru dapat mengalami kerusakan akibat gaya hidup tidak sehat. Merokok menjadi penyebab utama penyakit paru-paru, *World Health Organization* mengatakan lebih dari 40% kematian akibat merokok dan setidaknya 8 juta orang terbunuh setiap tahunnya. Penelitian ini mengambil data melalui situs penyedia dataset yaitu *Kaggle* sebanyak 30.000 data yang memiliki 10 atribut, tetapi hanya 7 atribut saja yang terpilih karena dinilai sesuai dengan prediksi penyakit paru-paru. Metode *Decision Tree* dengan algoritma C4.5 dipilih sebagai proses pengujian pada penelitian ini, hasil yang diperoleh mulai dari nilai *accuracy*, *precision*, dan *recall* dengan membandingkan *split data* pada *data training* dan *data testing* yang terbesar adalah *split data* 90 – 10 yaitu *accuracy* = 89.77%, *recall* = 78.61%, dan *precision* = 100%. Algoritma C4.5 yang diterapkan pada penelitian menghasilkan akurasi yang baik untuk memprediksi penyakit paru-paru, untuk penelitian selanjutnya dapat dilakukan dengan mencoba algoritma lain dari teknik data mining agar mengetahui hasil akurasi yang di peroleh melalui algoritma lain sebagai perbandingan dengan penelitian ini.

Saran untuk penelitian selanjutnya dapat dilakukan dengan mencoba algoritma lain dari teknik data mining agar mengetahui hasil akurasi yang di peroleh melalui algoritma lain sebagai perbandingan dengan penelitian yang sudah dilakukan ini.

DAFTAR PUSTAKA

- [1] D. Kurniawan, "Rancang Bangun Alat Deteksi Suara Paru-Paru Untuk Menganalisa Kelainan Paru-Paru Berbasis Android," *Elinvo (Electronics, Informatics, Vocat. Educ.*, vol. 2, no. 2, pp. 156–168, 2017, doi: 10.21831/elinvo.v2i2.17309.
- [2] M. Mustika, "ANALISIS DETEKSI KELAINAN SUARA PARU-PARU BERBASIS PENGOLAHAN SINYAL DIGITAL DSP TMS320C6416T," 2019, [Online]. Available: <http://eprints.polsri.ac.id/7934/>
- [3] Ciputra Hospital, "Bagaimana Cara Mencegah Penyakit Paru-Paru?," 2022. [https://ciputrahospital.com/bagaimana-cara-](https://ciputrahospital.com/bagaimana-cara-mencegah-penyakit-paru-paru/#:~:text=Penyakit paru-paru disebabkan oleh,akan dilepaskan dan nantinya mencemari udara. (accessed Apr. 29, 2023).)
- [4] Lung Foundation Australia, "A lung disease diagnosis can be shocking. But you are not alone.," 2022. <https://lungfoundation.com.au/lung-health/lung-disease/what-is-lung-disease/> (accessed Apr. 29, 2023).
- [5] E. W. Tipa, P. A. Kawatu, and A. F. C. Kalesaran, "Hubungan Kebiasaan Merokok Dengan Kapasitas Vital Paru Pada Penambang Emas Di Desa Tatelu Kabupaten Minahasa Utara," *J. KESMAS*, vol. 10, no. 3, pp. 140–146, 2021.
- [6] World Health Organization, "TBC," 2023. <https://www.who.int/news-room/fact-sheets/detail/tuberculosis> (accessed Apr. 28, 2023).
- [7] S. Febriani and H. Sulistiani, "Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5," *89Jurnal Teknol. dan Sist. Inf.*, vol. 2, no. 4, pp. 89–95, 2021.
- [8] C. Anam and H. B. Santoso, "Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa," *Energy - J. Ilm. Ilmu-Ilmu Tek.*, vol. 8, no. 1, pp. 13–19, 2018, [Online]. Available: <https://ejournal.upm.ac.id/index.php/energy/article/view/111>
- [9] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naive Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.
- [10] A. Natuzzuhriyyah and N. Nafisah, "Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Secara Daring Menggunakan Algoritma Naive Bayes," *Techno Xplore J. Ilmu Komput. dan Teknol. Inf.*, vol. 6, no. 2, pp. 61–67, 2021, doi: 10.36805/technoxplore.v6i2.1377.
- [11] A. Yoga Pratama *et al.*, "Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja)," *J. Sains*

- Komput. Inform. (J-SAKTI*, vol. 5, no. 2, pp. 897–910, 2021.
- [12] D. L. Rianti, Y. Umaidah, and A. Voutama, “Tren Marketplace Berdasarkan Klasifikasi Ulasan Pelanggan Menggunakan Perbandingan Kernel Support Vector Machine,” *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 6, no. 1, p. 98, 2021, doi: 10.30998/string.v6i1.9993.
- [13] Euis Saraswati, Yuyun Umaidah, and Apriade Voutama, “Penerapan Algoritma Artificial Neural Network untuk Klasifikasi Opini Publik Terhadap Covid-19,” *Gener. J.*, vol. 5, no. 2, pp. 109–118, 2021, doi: 10.29407/gj.v5i2.16125.
- [14] F. M. Almufqi, A. Voutama, S. Informasi, I. Komputer, U. S. Karawang, and T. Timur, “PERBANDINGAN METODE DATA MINING UNTUK MEMPREDIKSI PRESTASI,” vol. 15, no. 1, pp. 61–66, 2023, doi: 10.30736/jt.v15i1.929.
- [15] Abdussalam Amrullah, Intam Purnamasari, Betha Nurina Sari, Garno, and Apriade Voutama, “Analisis Cluster Faktor Penunjang Pendidikan Menggunakan Algoritma K-Means (Studi Kasus: Kabupaten Karawang),” *J. Inform. dan Rekayasa Elektron.*, vol. 5, no. 2, pp. 244–252, 2022, doi: 10.36595/jire.v5i2.701.
- [16] DQLab, “Pentingnya Implementasi Data Sekunder dalam Sebuah Penelitian,” 2022. [https://dqlab.id/pentingnya-implementasi-data-sekunder-dalam-sebuah-penelitian#:~:text=pembahasannya dibawah ini!-,1.,teknik pengumpulan data tertentu lainnya.](https://dqlab.id/pentingnya-implementasi-data-sekunder-dalam-sebuah-penelitian#:~:text=pembahasannya%20dibawah%20ini!-,1.,teknik%20pengumpulan%20data%20tertentu%20lainnya.) (accessed Apr. 30, 2023).