Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi
Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru
Loktabat – Banjarbaru (Tlp. 0511 4782881), e-mail: puslit.stmikbjb@gmail.com
e-ISSN: 2685-0893
p-ISSN: 2089-3787

Penerapan Algoritme *Decision Tree* Pada Klasifikasi Penyakit Kanker Paru-Paru

Hadi Widya Nugraha Surya Putra[1*], Vihi Atina[2], Joni Maulindar[3]
[1,3]Teknik Informatika, Universitas Duta Bangsa, Surakarta, Indonesia
[2]Manajemen Informatika, Universitas Duta Bangsa, Surakarta, Indonesia
*e-mail *Corresponding Author: 190103105@fikom.udb.ac.id*

Abstract

One of the deadliest diseases in the world is lung cancer. It is the biggest trigger of cancer-related deaths than any other type of cancer. The cause of this disease is due to uncontrolled cell growth in the body. Through classification, lung cancer patterns can be found. One of the widely used classificastion methods is the Decision Tree techniques. The Decision Tree C4.5 technique is a simple classification algorithm but has high performance and accuracy. This algorithm can generate decisions by forming a decision tree and can be used to predict lung cancer. From the result of the test carried out, a considerable accuracy was obtained, namely 89% with a Precission of 70% and Recall of 74.5%.

Keywords: *Data Mining; Classification, Decision Tree C4.5, Lung Cancer*

Abstrak

Salah satu penyakit yang paling mematikan didunia adalah penyakit kanker paru-paru. Penyakit ini adalah pemicu terbesar kematian terkait kanker dibanding setiap jenis macam penyakit kanker lainnya. Penyebab tingginya angka kematian pasien kanker paru-paru disebabkan karena terlambat dideteksi. Klasifikasi memungkinkan untuk menemukan pola pada kanker paru-paru, yang memungkinkan pendeteksian awal penyakit kanker paru-paru. Salah satu teknik klasifikasi yang banyak digunakan adalah Decision Tree. Metode *Decision Tree* C4.5 merupakan teknik klasifikasi sederhana dengan performa dan akurasi tinggi. Algoritme ini dapat digunakan untuk memprediksi kanker paru-paru. Dari hasil pengujian yang dilakukan didapatkan akurasi yang cukup besar yaitu 89%, dengan *Precission* sebesar 70% dan *Recall* sebesar 74.5 %.

Kata kunci: *Data Mining; Klasifikasi; Decision Tree C4.5; Kanker Paru-paru*

1. Pendahuluan

Penyakit kanker dapat terjadi karena adanya perkembangan sel yang tak terkontrol serta mempunyai kemampuan untuk menyebar keluar jaringan [1]. Penyebab terjadinya kondisi ini disebabkan oleh terjadinya perubahan pada *Deoxyribonucleic* atau yang biasa dikenal dengan DNA, yang menyebabkan sel kehilangan fungsi normalnya [2]. Penyakit kanker dapat berubah bentuk dan menyebar ke organ lainnya atau dapat disebut dengan metastase [3]. Penyakit kanker bisa tumbuh di beragam organ tubuh manusia, contohnya seperti payudara, prostat, ginjal, paru-paru, dan masih banyak lagi [4]. Sel kanker bisa menimbulkan tumor, gangguan pada sistem kekebalan tubuh, dan gangguan lainnya yang dapat mempengaruhi fungsi tubuh [5]. Kanker paru-paru adalah penyakit kanker yang memiliki tingkat kematian tertinggi mencapai hingga 13% dari semua jenis diagnosis kanker lainnya [6]. Kanker paru-paru bisa mengenai siapa pun tanpa melihat jenis kelamin [7]. Kanker paru-paru memiliki beberapa gejala yang sebagian besar terjadi di dalam organ paru-paru, tapi bisa juga dialami di beberapa bagian tubuh yang lain jika sel abnormal sudah menyebar.

Penyakit kanker terus mengalami peningkatan sepanjang tahun 2022. Berdasarkan informasi keterangan yang ada pada *American Cancer Society*, ditemukan masalah baru pada pasien penyakit kanker dengan jumlah 1.918.030 kasus, dimana jumlah tersebut didapat hanya pada tahun 2022 saja. Dalam kejadian tersebut terdapat 236.740 kasus yang merupakan pasien yang mengidap kanker paru-paru dan 130.180 dari mereka meninggal [8]. Setelah dokter mendiagnosa bahwa pasien terkena penyakit kanker paru-paru, hanya 17% pasien yang dapat bertahan hidup selama lima tahun [9]. Berdasarkan informasi yang ada di Pusat Kajian Jaminan

Sosial Universitas Indonesia (PKJS-UI). Di Indonesia, penyakit kanker paru-paru adalah satu diantara penyakit yang ada dengan paling banyak pengidapnya serta mempunyai angka kematian tertinggi dibandingkan dengan jenis kanker lainnya. Dengan prosentase lebih dari 70%, usia pasien kanker paru-paru di Indonesia tergolong muda dengan usia kurang dari 60 tahun [10].

Banyak kasus kanker paru-paru yang tidak dapat ditemukan lebih awal karena susah terdeteksi sehingga mengakibatkan tingginya tingkat kematian yang disebabkan karena kanker paru-paru. Melalui teknologi pengenalan pola, statistik dan matematika, data mining bertujuan untuk mendapatkan informasi baru yang signifikan dari kumpulan data besar yang tersimpan didalam suatu repositori. Proses ini melibatkan penyaringan data yang berguna untuk menemukan pola, korelasi, dan tren baru [11]. Data mining juga memiliki fungsi untuk mengidentifikasi dan mengungkap informasi yang relevan dan berharga dari berbagai sumber data [12]. Guna membantu diagnosis awal pada penyakit kanker paru-paru, proses klasifikasi dapat membantu pola kanker paru-paru ditemukan.

Klasifikasi merupakan proses untuk menemukan pola yang memilah kelas data sehingga bisa dimanfaatkan untuk memprediksi data yang masih belum mempunyai kelas data yang spesifik [13]. Klasifikasi yang banyak digunakan yaitu dengan metode *Decision Tree* yang merupakan teknik klasifikasi sederhana dengan performa dan akurasi tinggi. Decision tree atau yang biasa disebut *top-down induction of decision trees* (TIDIDT), merupakan teknik supervised learning dengan menggunakan partisi gabungan data train secara rekursif yang berguna untuk membuat perwakilan aturan klasifikasi yang memiliki struktur sekuensial [14].

Penelitian ini dilakukan dengan tujuan menerapkan algoritme *Decision tree* pada klasifikasi penyakit kanker paru-paru yang bermanfaat untuk pendeteksian awal penyakit kanker paru-paru.

2. Tinjauan Pustaka

Sebuah penelitian yang dilaksanakan oleh Eva Wulandari, serta Andreas Perdana di tahun 2022 dengan judul "Klasifikasi Kanker Paru-paru Menggunakan Metode *Naïve Bayes*", penelitian tersebut memiliki tujuan agar dapat mengklasifikasikan sebuah data secara presisi serta memiliki akurasi yang baik. Hasil yang diperoleh pada penelitian ini adalah implementasi data mining dengan memanfaatkan algoritme *Naïve bayes* bisa mendapatkan sebuah model klasifikasi penyakit kanker paru-paru [15].

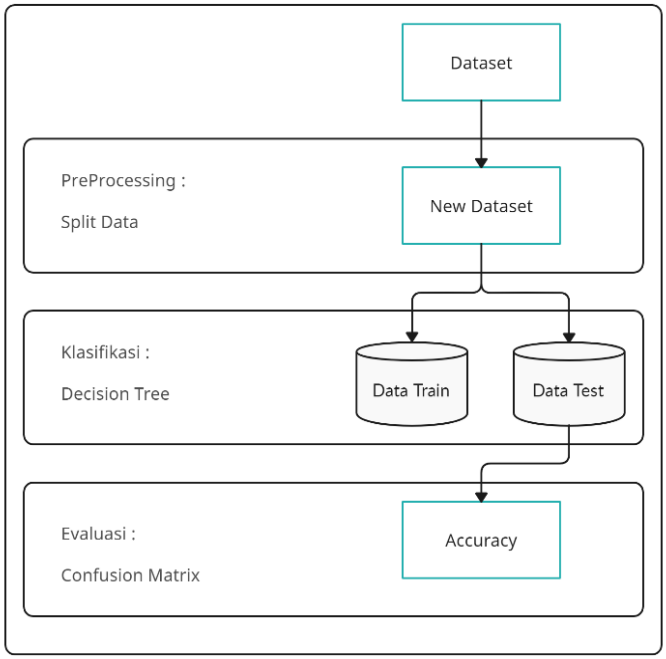
Anwar Rifai, dan Yani Prabowo melakukan penelitian pada tahun 2022 dengan judul "Diagnosis Kanker Paru-paru Dengan Sistem Fuzzy", penelitian tersebut bertujuan untuk mendiagnosa penyakit kanker paru-paru dan menolong masyarakat untuk melakukan diagnosa lebih awal penyakit kanker paru-paru. Penelitian yang dilakukan menggunakan sistem yang berlogika *fuzzy* dengan menggunakan metode mamdani. Untuk sebuah sistem fuzzy, tingkat akurasi yang didapatkan pada penelitian kali ini termasuk cukup baik, baik untuk data yang sudah diperbarui ataupun yang belum diperbarui. Penelitian ini membuktikan dengan menggunakan inferensi sistem *fuzzy* tipe mamdani dapat menghasilkan diagnose dengan cukup baik [16].

Pada tahun 2021 Yunianto, Dkk melakukan penelitian yang berjudul "Klasifikasi Kanker Paru-paru menggunakan naïve bayes dengan variasi filter dan ekstraksi ciri *gray level co-occurance matrix* (GLCM)", pengkajian tersebut bertujuan agar dapat mengelompokan citra kanker paru-paru ataupun paru-paru yang sehat dengan melewati fase *preprocessing*, proses segmentasi dengan tahapan ini berhasil mendapatkan citra biner dengan batas tepi yang jelas dan berdasarkan klasifikasi dengan memakai metode *Naïve bayes*, didapatkan nilai akurasi sejumlah 88,33% yang berarti teknik tersebut dapat diimplementasikan pada pengelompokan citra kanker paru-paru dan paru-paru sehat dari performa citra CTScan [17].

Penelitian kali ini memiliki fungsi yang sama dengan ketiga penelitian sebelumnya yaitu diagnose penyakit kanker paru-paru, hanya saja penelitian ini memiliki perbedaan. Pada penelitian yang telah dilakukan Eva Wulandari dan Andreas Perdana [15] pada penelitian tersebut memiliki perbedaan penggunaan metode *Naïve bayes* sedangkan pada pengkajian ini menggunakan metode *Decision tree* C4.5. Pada riset [16] memiliki perbedaan yaitu menggunakan sistem fuzzy, sedangkan pada penelitian ini menggunakan data mining klasifikasi. Pada penelitian [17] memiliki perbedaan menggunakan metode *Naïve bayes* dan esktraksi ciri *gray level occurance matrix* sedangkan riset ini memanfaatkan teknik *Decision tree* C4.5. Berdasarkan perbandingan dengan beberapa penelitian diatas, peneliti mendapatkan sebuah kesimpulan bahwa penelitian yang akan dibuat ini memiliki perbedaan metode dengan penelitian diatas.

3. Metodologi

Pada penelitian ini terdapat Langkah-langkah penelitian yang digunakan pada klasifikasi kanker paru-paru yang bisa diamati pada gambar dibawah:



Gambar 1. Tahapan Penelitian

Tahap pertama yaitu menyiapkan dataset yang selanjutnya data tersebut akan di proses dengan cara split data, data dibagi kedalam dua bagian, data dipisah menjadi data train dan data test dengan rasio pembagian 70% pada data train dan 30% pada data test. Data training merupakan gabungan data yang mempunyai atribut label dengan fungsi untuk mengidentifikasi karakteristik rangkaian data dan kemudian bisa mendapatkan suatu pola atau model data, sedangkan data testing merupakan kumpulan data dengan label yang berfungsi sebagai penguji kepresisian model pada pengklasifikasian data testing [18]. Model yang digunakan dalam pengujian yaitu klasifikasi dengan algoritme C4.5. dan akan dievaluasi dengan memanfaatkan confusion matrix dengan tujuan menghasilkan akurasi. Akurasi merupakan hasil dari membandingkan jumlah data dokumen yang sesuai dan jumlah total data yang ada [19].

3.1 Sumber Data

Penulis menggunakan data sekunder yang berarti sumber data didapatkan dengan cara tidak langsung dengan sumber yang didapat melalui internet [20]. Penulis menggunakan dataset lung cancer, dataset diperoleh dari website Kaggle yang bisa ditemukan pada alamat web https://www.kaggle.com/datasets/mysarahmadbhat/lung-cancer, dimana pengumpulan data dilakukan dengan menggunakan kuesioner. Terdapat 16 atribut dan 309 total dataset yang digunakan dalam penelitian.

3.2 Split Data

Pada tahapan split data, sebanyak 309 data dipisah menjadi dua bagian, data dipisah menjadi data train serta data test dengan presentasi 70% data train lalu 30% data test. Pembagian data ini diproses secara otomatis dengan memanfaatkan sebuah modul yang ada pada pemrograman python yang bernama Scikit-learn menggunakan perintah *train_test_split*.

3.3 Pemodelan

Pada penelitian ini menggunakan teknik klasifikasi menggunakan algoritme C4.5 yang nantinya akan di implementasikan menggunakan bahasa pemrograman Python dan modul pustaka Scikit. Algoritme C4.5 memiliki rumus yang terbagi menjadi 2, untuk menghitung sebuah gain diperlukan rumus persamaan pertama pada gambar 2 berikut:

$$Gain(S,A) = Entrophy(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} x\ Entophy(S_i)$$

Penjelasan:
S : Kumpulan kasus
A : Atribut
N : Banyaknya partisi atribut A
|Si| : Jumlah kasus pada partisi ke-i
|S| : Jumlah kasus dalam S

Gambar 2. Rumus Gain

Sedangkan untuk menghitung nilai entropi diperlukan rumus persamaan kedua pada gambar 3 berikut:

$$Entrophy(S) - \sum_{i=1}^{n} - pi\ x\ log_2 pi$$

Keterangan:
S : Himpunan kasus
n : Jumlah partisi S
pi : Proporsi Si terhadap S

Gambar 3. Rumus Entrophy

3.4 Evaluasi

Pada tahapan evaluasi menggunakan *Precission, Recall, F1-score,* dan *Confussion Matrix.*

4. Hasil dan Pembahasan

Variable data penelitian yang dipakai ditunjukan pada table di bawah ini:

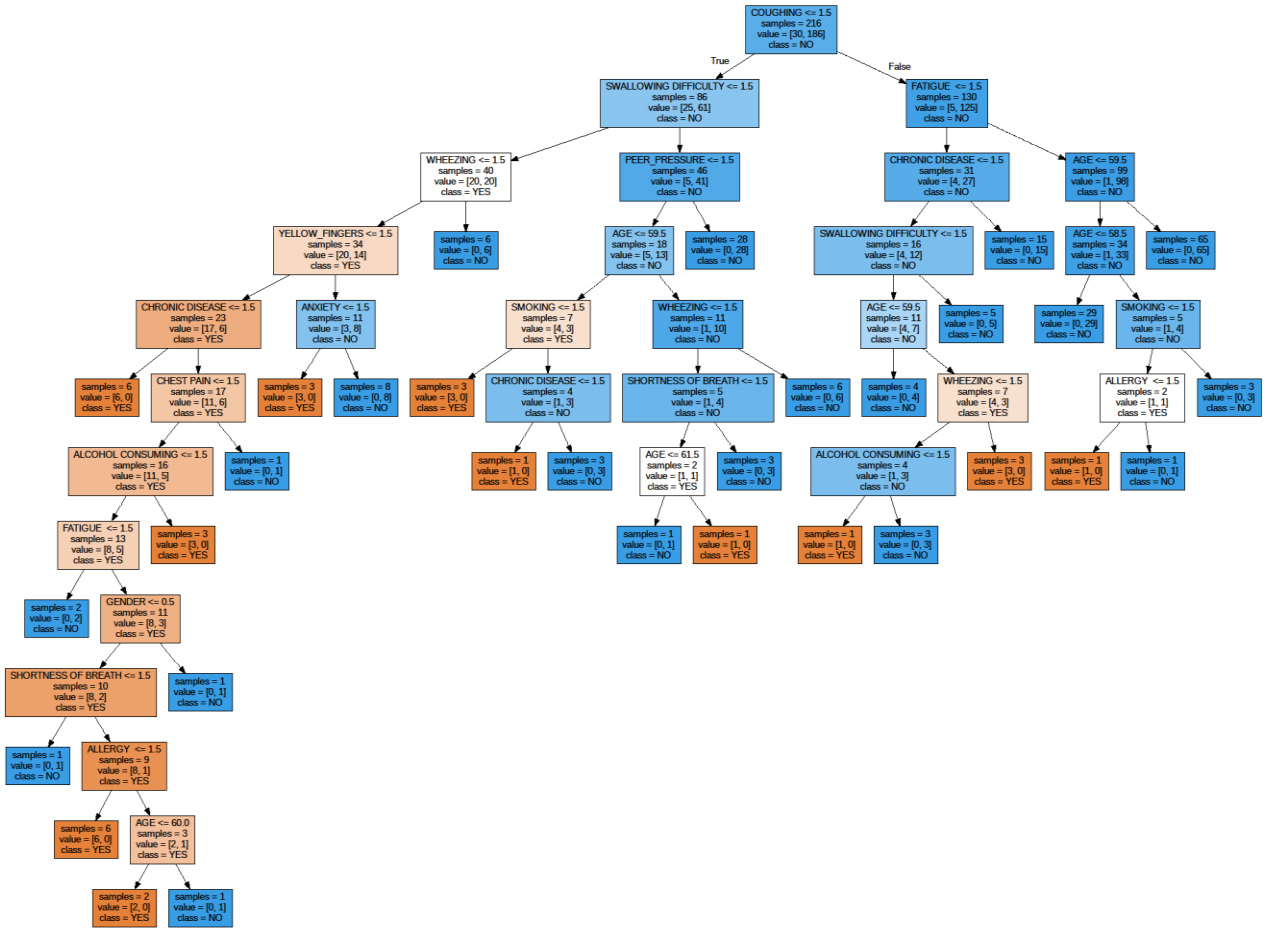
Tabel 1. Variabel data

No.	Atribut	Tipe data	Value	Keterangan
1	Gender	text	M F	Male Female
2	Age	Numerik	21, 38, 44, 47, 48, 49, 51 – 81, 87	Mulai dari usia 21 hingga 87
3	Smoking	Numerik	1, 2	1 = No, 2 = Yes
4	Yellow fingers	Numerik	1,2	1 = No, 2 = Yes
5	Anxiety	Numerik	1,2	1 = No, 2 = Yes
6	Peer pressure	Numerik	1,2	1 = No, 2 = Yes
7	Chronic Disease	Numerik	1,2	1 = No, 2 = Yes
8	Fatigue	Numerik	1,2	1 = No, 2 = Yes
9	Allergy	Numerik	1,2	1 = No, 2 = Yes
10	Wheezing	Numerik	1,2	1 = No, 2 = Yes
11	Alcohol	Numerik	1,2	1 = No, 2 = Yes
12	Coughing	Numerik	1,2	1 = No, 2 = Yes
13	Shortness of breath	Numerik	1,2	1 = No, 2 = Yes
14	Swallowing difficulty	Numerik	1,2	1 = No, 2 = Yes
15	Chest pain	Numerik	1,2	1 = No, 2 = Yes
16	Lung cancer	Text	Yes, No	Yes, No

Pada tabel diatas terdapat 16 variable dataset gejala sebagai atribut dan 1 variable sebagai target atribut penentu proses klasifikasi. Dataset Gender “Male” diberi label angka 1

sedangkan Gender “Female” diberi label dengan angka 0. Dan dataset Lung cancer “YES” diberi label dengan angka 1 sedangkan Lung cancer “NO” diberi label dengan angka 0.

Dari total 309 data, dibagi menjadi dua dengan presentasi 70% data train yaitu sebanyak 216 data, dan 30% data test yaitu sebanyak 93 data. Setelah data split dilakukan, selanjutnya dibentuk pohon keputusan dengan menggunakan algoritme C4.5 agar menghasilkan sebuah alur keputusan. Berikut merupakan pohon keputusan yang didapat.



Gambar 2. Pohon Keputusan

Berikut ini merupakan hasil pengujian dari 216 data training dan 93 data testing dengan setiap data memiliki variable dataset gejala berjumlah 15 dan 1 variabel sebagai target atribut.

Table 2. *Confusion Matrix*

	True yes	True no	Class precission
Predic. yes	5	4	95%
Predic. no	6	78	45%
Class recall	93%	56%	

Untuk mendapatkan hasil *precision* dan r*ecall* dapat diketahui dengan rumus sebagai berikut:

$$precision = \frac{1}{2} x \left(\frac{negatif - F.N}{negatif - F.N + F.P} + \frac{positif - F.P}{positif - F.P + F.N} \right) x100\%$$

$$recall = \frac{1}{2} x \left(\frac{negatif - F.N}{negatif} + \frac{positif - F.P}{positif} \right) x100\%$$

Berdasarkan table 2 diatas, hasil nilai *Accuracy* yang didapat 89%, *Precision* = 70%, dan *Recall* = 74.5% dengan jumlah prediksi benar 83 dari total data tesing sebanyak 93 data. Berikut ini merupakan gambar dari hasil prediksi:

```
Label Sebenarnya :  [1 0 1 1 1 1 1 1 1 1 0 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1
 1 1 0 1 1 1 1 0 1 0 1 1 1 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1]
Hasil Prediksi :  [1 0 1 1 1 1 1 1 1 1 0 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1
 1 1 0 1 1 1 1 0 1 0 1 1 1 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1]
```

Gambar 3. Hasil prediksi

Dan berikut ini merupakan hasil prediksi dalam bentuk persen:

```
Prediksi Benar :  83 data
Prediksi Salah :  10 data
Akurasi        :  89.24731182795699 %
```

Gambar 4. Hasil prediksi dalam persen

Berdasarkan hasil pengujian yang sudah dilaksanakan, Algoritme C4.5 dapat digunakan untuk memprediksi kanker paru-paru dengan tingkat akurasi sebesar 89%. Seperti penelitian yang telah dilakukan oleh Eva Wulandari dan Andreas Perdana [15], Anwar Rifai dan Yani Prabowo [16], dan Yunianto, Dkk [17] penelitian ini membuahkan hasil yang dapat membuktikan bahwa algoritme C4.5 juga dapat digunakan untuk memprediksi kanker paru-paru.

5. Simpulan

Pada penelitian ini, telah dilakukan klasifikasi data gejala Kanker Paru-paru menggunakan metode *Decision tree* dengan algortima C4.5 dengan menggunakan Bahasa pemrograman *python*. Pada dataset terdapat 16 atribut yaitu Gender, *Chronic disease, Allergy, Yellow fingers, Age, Anxiety, Chest pain, Peer pressure, Swallowing difficulty*, *Shortness of breath, Wheezing, Smoking, Coughing, Alcohol consuming, Fatigue*, serta *Lung cancer*. Dari 309 dataset, 93 diantaranya dijadikan sebagai data test dan hasil dari pengujian menggunakan algoritme C4.5 memberikan hasil kinerja yang cukup baik dengan *Precision* sebesar 70%, *Recall* sebesar 74.5%, dan akurasi yang didapat sebesar 89% dengan evaluasi dan validasi menggunakan parameter *confussion matrix*, dari hasil akurasi yang didapat, diketahui bahwa jumlah prediksi benar adalah 83 data dari total jumlah data test yaitu sebanyak 93 data. Dengan hasil tersebut, dapat dibuktikan bahwa algoritme *Decision tree* dapat digunakan menjadi model pada klasifikasi untuk penelitian ini. Menurut penulis, penelitian yang dilakukan masih dapat dapat ditingkatkan dengan mengimplementasikan dan memadankan performa antara algoritme-algoritme data mining yang lain. Dengan adanya perbandingan performa maka dapat dilihat kinerja berbagai algoritme data mining yang lain.

Daftar Referensi

[1] A. S. Rachmawati, “Prevalensi Kanker Di Rumah Sakit Jasa Kartini Kota Tasikmalaya Tahun 2018,” *Jurnal Kesehatan komunitas Indonesia*, vol. 16, no. 1, pp. 119–126, 2020.

[2] S. Hartini, B. Dwi Winarsih, E. Galih Zulva Nugroho, P. Studi Profesi Ners, P. S. Studi, dan I. Keperawatan, “Peningkatan Pengetahuan Perawat Untuk Perawatan Anak Penderita Kanker,” *Jurnal Pengabdian Kesehatan*, vol. 3, no. 2, pp. 141–149, 2020, [Daring]. Tersedia pada: http://jpk.jurnal.stikescendekiautamakudus.ac.id

[3] A. Lestari dan Y. Budiyarti, “Study Fenomenologi: Psikologis Pasien Kanker Yang Menjalani Kemoterapi,” *Jurnal Keperawatan Suaka Insan* |, vol. 5, no. 1, pp. 52–66, 2020.

[4] L. Rahayuwati, I. A. Rizal, T. Pahria, M. Lukman, dan N. Juniarti, “Pendidikan Kesehatan tentang Pencegahan Penyakit Kanker dan Menjaga Kualitas Kesehatan,” *Media Karya Kesehatan*, vol. 3, no. 1, pp. 59–69, 2020.

[5] S. Rahmawati, P. Onkogen, D. T. Suppressor, dan G. Pada Karsinogenesis, “Peran Onkogen dan Tumor Suppressor Gene pada Karsinogenesis,” *JK Unila*, vol. 5, no. 1, pp. 61–68, 2021.

[6] A. Reynaldi, Y. Trisyani, dan D. Adiningsih, “Kualitas Hidup Pasien Kanker Paru Stadium Lanjut,” *JNC*, vol. 3, no. 2, pp. 71–79, 2020.

[7] Juwita, N. Amalita, dan M. D. Parma, “Faktor-Faktor Risiko yang Mempengaruhi Kanker Paru-Paru dengan Menggunakan Analisis Regresi Logistik,” *UNPjoMath*, vol. 4, no. 1, pp. 38–42, 2021.

[8] R. L. Siegel, K. D. Miller, H. E. Fuchs, dan A. Jemal, “Cancer statistics, 2022,” *CA Cancer J Clin*, vol. 72, no. 1, pp. 7–33, Jan 2022, doi: 10.3322/caac.21708.

[9] F. A. Hermawati, “Sistem Deteksi Keganasan Kanker Paru-Paru pada CT Scan dengan Menggunakan Metode Mask Region-based Convolutional Neural Network (Mask R-CNN),” *Konferensi Nasional Ilmu Komputer (KONIK)*, pp. 193–197, 2021.

[10] A. Dewi *dkk.*, “Kanker Paru, Kanker Paling Mematikan Di Indonesia: Apa Saja Yang Telah Kita Atasi Dan Apa Yang Kita Bisa Lakukan,” dalam *Dialog Pemangku Kepentingan dengan tema “Kanker Paling Mematikan di Indonesia: Seberapa Jauh Kita Atasi dan Apa yang Dapat Kita Lakukan?,”* Jakarta Pusat: Pusat Kajian Jaminan Sosial Universitas Indonesia (PKJS-UI), Feb 2021, pp. 1–29, 2020.

[11] Z. Nabila, A. Rahman Isnain, dan Z. Abidin, “Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritme K-Means,” *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 2, pp. 100–108, 2021, [Daring]. Tersedia pada: http://jim.teknokrat.ac.id/index.php/JTSI

[12] K. K. Ningrum, J. Maulindar, dan A. Farida, “Penerapan Algoritme K-Means Untuk Pengelompokkan Penilaian Akhir Semester Di Sdn Kadokan 01 Sukoharjo,” *INFOTECH journal*, vol. 9, no. 1, pp. 190–197, 2023, doi: 10.31949/infotech.v9i1.xxx.

[13] Y. Rosela, “Implementasi Klasifikasi Decision Tree Menganalisa Status Penjualan Barang Menggunakan C4.5 (Studi Kasus: PT. Matahari Department Store Medan Mall),” *Jurnal Pelita Informatika*, vol. 7, no. 3, hlm. 404–411, 2019.

[14] F. S. Pamungkas, B. D. Prasetya, dan I. Kharisudin, “Perbandingan Metode Klasifikasi Supervised Learning pada Data Bank Customers Menggunakan Python,” *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 3, pp. 689–694, 2019, [Daring]. Tersedia pada: https://journal.unnes.ac.id/sju/index.php/prisma/

[15] E. Wulandari dan A. Perdana, “Klasifikasi Kanker Paru-Paru Menggunakan Metode Naive Bayes,” *I-Robot Journal*, vol. 6, no. 2, pp. 20–24, 2022.

[16] A. Rifa’i dan Y. Prabowo, “Diagnosis Kanker Paru-Paru dengan Sistem Fuzzy,” *Krea-TIF: Jurnal Teknik Informatika*, vol. 10, no. 1, pp. 19–28, 2022, doi: 10.32832/kreatif.v10i1.6317.

[17] M. Yunianto, F. Anwar, D. Nur Septianingsih, T. Dwi Ardyanto, dan R. Farits Pradana, “Klasifikasi Kanker Paru Paru Menggunakan Naïve Bayes Dengan Variasi Filter Dan Ekstraksi Ciri Gray Level Co-Occurance Matrix (GLCM),” *Indonesian Journal of Applied Physics*, vol. 11, no. 2, pp. 256–268, 2021.

[18] W. Musu dan A. Ibrahim, “Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritme C4.5,” *Prosidingseminar Ilmiah Sistem Informasi Dan Teknologi Informasi*, vol. 10, no. 1, pp. 186–195, 2021.

[19] D. Darwis, N. Siskawati, dan Z. Abidin, “Penerapan Algoritme Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional,” *Jurnal TEKNO KOMPAK*, vol. 15, no. 1, pp. 131–145, 2021.

[20] M. Fahmi Panwar, E. Purwanto, dan V. Atina, “Sistem Pakar Tes Psikologi Dengan Menggunakan Metode Dominance Influence Steadiness And Compliance,” dalam *Prosiding Seminar Nasional*, 2022, pp. 240–244.