

Penerapan Metode Klasifikasi Decision Tree dalam Prediksi Kanker Paru-Paru Menggunakan Algoritma C4.5

Riska Rofiani^{1,*}, Luluk Oktaviani², Dwi Vernanda³ Tri Hendriawan⁴

^{1,2,3,4}Jurusan Teknologi Informasi dan Komputer, Sistem Informasi, Politeknik Negeri Subang, Subang, Indonesia

Email: ^{1,*}Riska.10108050@student.polsub.ac.id, ²Luluk.10108032@student.polsub.ac.id, ³nanda@polsub.ac.id, ⁴tri@polsub.ac.id

^{*)} Riska.10108050@student.polsub.ac.id

Abstrak— Penelitian ini bertujuan untuk memprediksi kanker paru-paru menggunakan teknik Data Mining dengan algoritma Decision Tree, terutama algoritma C4.5. Dataset yang digunakan merupakan kumpulan data dari Kaggle yang mencakup atribut seperti usia, jenis kelamin, kebiasaan merokok, konsumsi alkohol, aktivitas, dan faktor lainnya yang diduga berperan dalam risiko terkena kanker paru-paru. Hasil klasifikasi dari model yang dibangun menunjukkan tingkat akurasi yang tinggi dalam memprediksi kemungkinan seseorang terkena kanker paru-paru. Analisis mendalam terhadap model mengindikasikan bahwa kebiasaan merokok adalah faktor dominan yang sangat memengaruhi prediksi penyakit ini. Faktor usia juga memainkan peran penting dalam penentuan risiko terkena kanker paru-paru, di mana individu dengan usia tertentu cenderung memiliki risiko lebih tinggi. Model klasifikasi yang dikembangkan dalam penelitian ini telah memberikan hasil yang konsisten dan dapat diandalkan dalam mengidentifikasi risiko kanker paru-paru. Keberhasilan model ini dalam memprediksi dengan tingkat akurasi yang tinggi menunjukkan potensi besar untuk mendeteksi dini penyakit ini. Penekanan yang besar pada kebiasaan merokok sebagai faktor risiko utama kanker paru-paru disorot dalam hasil penelitian ini. Model berhasil menggambarkan bahwa individu yang memiliki kebiasaan merokok memiliki kemungkinan lebih tinggi terkena kanker paru-paru daripada individu yang tidak merokok. Diharapkan model prediksi ini dapat digunakan sebagai alat penting dalam mengidentifikasi individu dengan risiko tinggi terkena kanker paru-paru secara dini. Dengan demikian, perawatan yang tepat dapat diberikan pada tahap awal penyakit, meningkatkan peluang kesembuhan dan memperbaiki prognosis. Namun, penelitian ini juga menunjukkan bahwa ada beberapa atribut lain yang mungkin memiliki pengaruh yang signifikan dalam risiko terkena kanker paru-paru. Oleh karena itu, penelitian lebih lanjut diperlukan untuk menggali lebih dalam atribut-atribut lain yang berkontribusi terhadap prediksi penyakit ini secara lebih komprehensif. Hasil klasifikasi yang kuat dari model ini menunjukkan bahwa analisis data menggunakan teknik Data Mining dan algoritma C4.5 dapat menjadi solusi efektif dalam deteksi dini dan pencegahan penyakit kanker paru-paru. Model ini dapat memberikan kontribusi yang signifikan dalam upaya penyelamatan nyawa dan perbaikan kualitas hidup bagi individu yang berisiko terkena penyakit ini.

Kata Kunci: Kanker Paru-Paru, Prediksi, Data Mining, Algoritma C4.5, Pohon Keputusan.

Abstract— This study aims to predict lung cancer using Data Mining techniques with the Decision Tree algorithm, particularly the C4.5 algorithm. The dataset used is a collection of data from Kaggle that includes attributes such as age, gender, smoking habits, alcohol consumption, activity level, and other factors suspected to influence the risk of lung cancer. The classification results of the constructed model indicate a high level of accuracy in predicting the likelihood of someone developing lung cancer. In-depth analysis of the model suggests that smoking habits significantly influence the prediction of this disease. Age also plays a crucial role in determining the risk of developing lung cancer, where individuals of certain ages tend to have a higher risk. The classification model developed in this research has consistently delivered reliable results in identifying the risk of lung cancer. The success of this model in predicting with high accuracy demonstrates significant potential for early detection of this disease. The substantial emphasis on smoking as a primary risk factor for lung cancer is highlighted in this research. The model successfully illustrates that individuals with smoking habits have a higher likelihood of developing lung cancer compared to non-smokers. It is expected that this predictive model can serve as a crucial tool in identifying individuals at high risk of developing lung cancer at an early stage. Consequently, appropriate treatment can be provided in the initial stages of the disease, increasing the chances of recovery and improving the prognosis. However, this study also indicates that there might be other attributes that could significantly influence the risk of developing lung cancer. Therefore, further research is needed to delve deeper into other attributes contributing to predicting this disease more comprehensively. The robust classification results from this model indicate that data analysis using Data Mining techniques and the C4.5 algorithm can be an effective solution for early detection and prevention of lung cancer. This model can make a significant contribution to life-saving efforts and improving the quality of life for individuals at risk of this disease.

Keywords: Lung Cancer, Prediction, Data Mining, C4.5 Algorithm, Decision Tree.

1. PENDAHULUAN

Paru-paru merupakan bagian dari sistem pernapasan manusia dan berperan penting dalam memenuhi kebutuhan oksigen tubuh. Selain itu, paru-paru juga berfungsi sebagai tempat pertukaran oksigen dari udara dengan karbon dioksida dari darah. Dalam situasi tertentu, paru-paru mungkin mengalami masalah yang dapat berdampak buruk pada fungsi sistem pernapasan. Jika paru-paru tidak berfungsi normal maka dapat menimbulkan penyakit. Paru-paru merupakan organ yang berperan penting dalam proses pernafasan. Respirasi adalah proses

kimia yang melibatkan pelepasan energi yang tersimpan dalam sumber energi dengan menggunakan oksigen (O₂) dan menghasilkan karbon dioksida (CO₂), air (H₂O) dan energi[1]. Saat ini yang sering terjadi pada manusia yaitu penyakit paru – paru. Penyebab umum penyakit paru-paru biasanya adalah menghirup udara yang terkontaminasi debu, asap, virus atau bakteri yang dapat menyebabkan infeksi saluran pernafasan. Penyakit paru-paru dapat menyerang individu di berbagai kelompok umur, mulai dari bayi hingga orang dewasa. Penting untuk dicatat bahwa penyakit paru-paru seringkali sulit disembuhkan[2].

Menurut Organisasi Kesehatan Dunia (WHO) Kanker paru-paru adalah salah satu kelompok penyakit kanker yang paling mematikan. Penyakit ini dapat disebabkan oleh faktor-faktor yang berkaitan dengan paru-paru itu sendiri atau faktor-faktor yang berasal dari luar paru-paru. Di Indonesia, kanker paru-paru merupakan penyakit yang cukup umum, dan Indonesia menempati peringkat ke-4 dalam jumlah penderita kanker paru-paru terbanyak di seluruh dunia. Penderita kanker paru-paru di Indonesia sebagian besar adalah pria yang berusia di atas 40 tahun. Hal ini menunjukkan pentingnya upaya pencegahan dan pengelolaan penyakit ini, terutama di kalangan populasi yang rentan. Kanker paru-paru adalah penyakit yang dicirikan oleh pertumbuhan sel yang tidak terkontrol dalam jaringan paru-paru[3]. Jika tidak diobati, pertumbuhan sel ini dapat menyebar keluar dari paru-paru melalui proses yang disebut metastasis, yang dapat memengaruhi jaringan yang berdekatan atau bahkan bagian tubuh lainnya. Kanker paru-paru dapat didiagnosis melalui pemeriksaan foto rontgen dada dan tomografi komputer (CT scan). Diagnosis yang pasti biasanya memerlukan biopsi, yaitu pengambilan sampel jaringan tubuh untuk pemeriksaan laboratorium. Prosedur ini biasanya dilakukan melalui bronkoskopi atau dipandu dengan bantuan CT.

Kanker paru-paru merupakan salah satu dari tiga penyakit paling berbahaya di dunia dan dapat berkembang dengan sangat cepat. Beberapa faktor mempengaruhi kecepatan berkembangnya kanker paru-paru. Faktor-faktor yang mempengaruhi pesatnya perkembangan kanker paru-paru di seluruh dunia antara lain paparan asap tembakau dalam jangka panjang, faktor genetik, radon dan polusi udara global, termasuk paparan asap rokok dalam jangka panjang, usia tua, faktor genetik, paparan radon dan polusi udara. Upaya pencegahan dan pengobatan dini sangat penting untuk mengatasi penyakit ini[4].

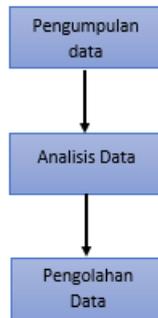
Oleh sebab itu maka perlu kiranya suatu model prediksi penyakit paru-paru secara dini merupakan sebuah langkah yang sangat penting dalam upaya pencegahan dan pengobatan lebih efektif terhadap kanker paru-paru. Model prediksi ini dapat membantu mengidentifikasi individu yang berisiko tinggi untuk penyakit ini sehingga mereka dapat segera mendapatkan perawatan dan intervensi yang diperlukan. Penelitian sebelumnya yang dilakukan oleh Eva dan Andreas menggunakan algoritma Naïve Bayes dan mencapai tingkat akurasi sebesar 73% merupakan langkah awal yang baik. Namun, dengan menggunakan algoritma C4.5, yang dikenal memiliki akurasi yang lebih baik dalam prediksi, diharapkan peningkatan dalam akurasi prediksi penyakit paru-paru. Dengan meningkatnya akurasi dalam prediksi, model ini memiliki potensi untuk digunakan dalam praktik medis dan dunia kesehatan secara luas. Dengan mendeteksi penyakit paru-paru lebih awal, penanganan dapat dimulai lebih cepat, dan ini dapat meningkatkan peluang kesembuhan atau pengendalian penyakit. Pengembangan dan penerapan model prediksi semacam ini adalah salah satu contoh bagaimana teknologi dan analisis data dapat memberikan kontribusi yang berharga dalam bidang kesehatan, khususnya dalam upaya mendeteksi penyakit secara dini dan menyelamatkan nyawa[5].

Data mining merupakan metode yang dapat digunakan untuk menghitung jumlah orang yang berisiko terkena kanker paru-paru. Penambangan data adalah cabang ilmu komputer yang berkembang pesat dan telah diterapkan di berbagai bidang seperti sains, teknik, dan bisnis. Tujuan utama dari data mining adalah untuk mendapatkan wawasan berharga dari kumpulan data yang besar. Banyak teknik data mining yang berbeda telah dikembangkan, termasuk teknik asosiasi, clustering, sequence modeling, dan klasifikasi. Waktu juga merupakan faktor penting dalam analisis data. Banyak data kehidupan nyata yang mencerminkan situasi atau status berbagai objek pada waktu tertentu, dan jenis data ini disebut data temporal. Dalam konteks deteksi kanker paru-paru, data mining dapat digunakan untuk menganalisis data medis yang mencakup informasi temporal, seperti riwayat kesehatan pasien dari waktu ke waktu[6]. Dengan menggunakan teknik-teknik data mining yang sesuai, informasi penting dapat diidentifikasi, pola-pola yang berkaitan dengan perkembangan penyakit dapat ditemukan, dan ini dapat membantu dalam mendukung diagnosis dini dan pengambilan keputusan dalam pengelolaan penyakit. Oleh karena itu, penggunaan data mining dalam analisis data medis memiliki potensi besar untuk meningkatkan deteksi dini dan penanganan kanker paru-paru.

Dalam penelitian ini, kami menerapkan teknik data mining yang menggunakan metode Decision Tree dengan algoritma C4.5. Fokus utama kami adalah untuk mencapai tingkat akurasi yang tepat dalam memprediksi penyakit paru-paru[7]. Algoritma C4.5 merupakan salah satu pendekatan yang digunakan dalam konstruksi pohon

keputusan untuk mengklasifikasikan data berdasarkan karakteristik atau atribut spesifik. Tujuan kami adalah untuk menghasilkan model prediksi penyakit paru-paru yang efektif dan akurat dengan mengadopsi algoritma ini. Dengan penerapan algoritma C4.5, kami berharap dapat menyoroti cara kami membangun model prediksi yang berpotensi memberikan hasil yang lebih andal dan relevan terkait deteksi dini penyakit paru-paru[8].

2. METODE PENELITIAN



Gambar 1. Metode Penelitian

2.1 Pengumpulan Data dan Pengolahan data awal

2.1.1. Pengumpulan Data

- a. Sumber Data: Dataset ini diambil dari situs web Kaggle dan terdiri dari 150 dataset yang relevan dengan kanker paru-paru.
- b. Atribut yang Tersedia: Dataset ini memiliki 10 atribut yang berbeda, yaitu:
 1. Usia: Informasi usia subjek dalam dataset.
 2. Jenis Kelamin: Jenis kelamin subjek, misalnya, pria atau wanita.
 3. Merokok: Status merokok subjek (Ya/Tidak).
 4. Alkohol: Status konsumsi alkohol subjek (Ya/Tidak).
 5. Jari Kuning: Detail terkait kondisi jari kuning yang terkait dengan penyakit.
 6. Bekerja: Status pekerjaan subjek (Ya/Tidak).
 7. Aktivitas Bergadang: Informasi tentang kebiasaan bergadang subjek.
 8. Aktivitas Olahraga: Informasi tentang kebiasaan berolahraga subjek.
 9. Penyakit Bawaan: Kondisi kesehatan bawaan subjek.
 10. Hasil: Hasil dari penelitian atau tes terkait penyakit kanker paru-paru.

2.1.2. Proses Pengumpulan Data

- a. Kegiatan Pengumpulan: Informasi tentang subjek dalam dataset dikumpulkan dari sumber yang terpercaya. Pengumpulan data ini mungkin melibatkan survei, catatan medis, atau sumber data kesehatan lainnya.
- b. Ketelitian Data: Proses pengumpulan data dilakukan dengan memastikan konsistensi, validitas, dan integritas data. Setiap atribut yang dikumpulkan harus memiliki dokumentasi dan spesifikasi yang jelas.
- c. Privasi dan Etika: Penting untuk memastikan bahwa data yang dikumpulkan dilakukan sesuai dengan peraturan privasi yang berlaku dan etika penelitian, seperti memastikan anonimitas subjek atau mendapatkan izin sebelum mengumpulkan data sensitif.
- d. Pengolahan Awal: Setelah pengumpulan, data dapat mengalami proses pengolahan awal untuk membersihkan data dari duplikasi, kesalahan, atau nilai yang hilang agar dapat digunakan dalam analisis lebih lanjut.

Proses pengumpulan data ini merupakan langkah penting dalam memastikan bahwa dataset yang diperoleh valid, lengkap, dan sesuai dengan kebutuhan analisis dan pengujian model.

Tabel 1. Dataset Prediksi Penyakit Kanker Paru - Paru

No	Usia	JK	Merokok	Alkohol	Jari Kuning	Bekerja	Aktivitas _Begadang	Aktivitas _Olahraga	Penyakit _Bawaan	Hasil
1	Tua	Pria	Pasif	Jarang	Ada	Tidak	Ya	Sering	Tidak	Ya
2	Tua	Pria	Aktif	Sering	Tidak	Tidak	Ya	Jarang	Ada	Tidak
3	Muda	Pria	Aktif	Sering	Tidak	Tidak	Ya	Jarang	Tidak	Tidak
4	Tua	Pria	Aktif	Sering	Tidak	Ya	Tidak	Jarang	Ada	Tidak
5	Muda	Wanita	Pasif	Sering	Ada	Ya	Tidak	Sering	Ada	Ya
6	Muda	Wanita	Pasif	Jarang	Tidak	Ya	Tidak	Sering	Ada	Tidak
7	Tua	Wanita	Pasif	Jarang	ada	Tidak	Tidak	Sering	Tidak	Ya
8	Muda	Pria	Aktif	Sering	Tidak	Tidak	Ya	Sering	Tidak	Tidak
9	Tua	Wanita	Aktif	Jarang	Ada	Ya	Ya	Jarang	Ada	Ya
10	Muda	Wanita	Pasif	Jarang	Ada	Ya	Ya	Jarang	Ada	Ya
11	Tua	Wanita	Pasif	Sering	Ada	Ya	Tidak	Sering	Ada	Ya
12	Tua	Wanita	Aktif	Sering	Ada	Tidak	Tidak	Jarang	Tidak	Tidak
13	Muda	Pria	Aktif	Sering	Ada	Tidak	Ya	Jarang	Tidak	Tidak
14	Tua	Wanita	Aktif	Jarang	Ada	Ya	Ya	Jarang	Ada	Tidak
15	Muda	Wanita	Pasif	Sering	Tidak	Ya	Ya	Sering	Ada	Ya
16	Muda	Wanita	Pasif	Jarang	Ada	Ya	Ya	Jarang	Ada	Ya
17	Tua	Wanita	Pasif	Jarang	Ada	Ya	Tidak	Sering	Ada	Ya
18	Tua	Wanita	Aktif	Jarang	Ada	Tidak	Tidak	Jarang	Tidak	Tidak
19	Muda	Pria	Aktif	Jarang	Ada	Tidak	Ya	Jarang	Tidak	Tidak
20	Tua	Wanita	Aktif	Sering	Tidak	Ya	Ya	Jarang	Ada	Tidak
21	Muda	Wanita	Pasif	Sering	Tidak	Ya	Ya	Sering	Ada	Ya
22	Tua	Pria	Pasif	Sering	Ada	Tidak	Ya	Sering	Tidak	Ya
23	Tua	Pria	Aktif	Sering	Ada	Tidak	Ya	Jarang	Ada	Tidak
24	Muda	Pria	Aktif	Jarang	Ada	Tidak	Ya	Jarang	Tidak	Tidak

2.2 Analisis Data

a. Analisis Data

1. Tujuan Utama: Tahap analisis data ini merupakan langkah awal yang penting sebelum proses pengolahan data. Fokus utamanya adalah untuk memahami dataset secara menyeluruh agar data siap untuk diolah atau diuji dengan menggunakan algoritma C4.5.
2. Deteksi Data Duplikat: Analisis data dilakukan untuk mengidentifikasi dan menghapus data duplikat, jika ada. Duplikasi dalam dataset dapat mengganggu analisis dan interpretasi hasil.
3. Validasi Konsistensi Data: Data diperiksa untuk memastikan konsistensi, baik dalam format maupun nilai. Hal ini termasuk dalam memeriksa data untuk memastikan bahwa setiap entri sesuai dengan aturan dan tipe data yang diharapkan.
4. Perbaiki Kesalahan Data: Jika ada kesalahan atau nilai yang tidak konsisten dalam dataset, langkah-langkah perbaikan akan dilakukan. Misalnya, memperbaiki format tanggal, menangani nilai yang hilang, atau melakukan transformasi untuk memastikan kesesuaian dengan algoritma C4.5.
5. Kriteria Algoritma C4.5: Dataset dianalisis untuk memastikan bahwa data memenuhi kriteria yang dibutuhkan oleh algoritma C4.5, seperti data numerik, kategori, dan kecocokan dengan tipe algoritma.

b. Proses Analisis Data

1. Tools atau Teknik yang Digunakan: Proses analisis data ini mungkin melibatkan penggunaan berbagai alat atau teknik seperti penggunaan software statistik, pemrograman untuk pembersihan data, atau visualisasi data untuk pemahaman yang lebih baik.
2. Validasi dan Verifikasi: Setelah analisis, data divalidasi ulang untuk memastikan bahwa semua masalah yang terdeteksi sebelumnya telah diperbaiki dan dataset siap untuk digunakan pada tahap berikutnya, yaitu pengolahan data dan penerapan algoritma C4.5.

Proses analisis data ini memastikan bahwa dataset yang akan digunakan untuk pengembangan model menggunakan algoritma C4.5 telah melewati tahap pengujian dan pembersihan yang diperlukan untuk meminimalkan potensi kesalahan dan memastikan keandalan hasil analisis[8].

2.3 Pengolahan Data

Tahap pengolahan data dalam metode penelitian adalah proses sistematis yang mencakup serangkaian langkah penting untuk mempersiapkan data sebelum dilakukan analisis[9]. Pengolahan dataset yang merupakan tidak lanjut dari analisis data, yang diterapkan dalam algoritma C4.5 terhadap prediksi kanker paru-paru yang akan menggunakan algoritma C 45 dan tols *Rapid Miner*. Proses ini terdiri dari beberapa tahap yang dijelaskan sebagai berikut:

- a. Pemilihan Data: Tahap pertama melibatkan pemilihan data yang relevan dengan pertanyaan penelitian. Data ini dapat berasal dari platform kaggle
- b. Pengumpulan Data: Setelah pemilihan data, langkah selanjutnya adalah mengumpulkan informasi dari sumber-sumber yang telah dipilih. Proses ini bisa melibatkan pengambilan data dari penelitian mandiri
- c. Pembersihan Data: Proses pembersihan data dilakukan untuk mengatasi data yang tidak lengkap, tidak akurat, atau tidak relevan. Ini melibatkan identifikasi dan penanganan nilai yang hilang, kesalahan entri, atau ketidakkonsistenan dalam data untuk memastikan kebersihan dan keakuratan sebelum analisis.
- d. Transformasi Data: Data sering kali memiliki format yang berbeda-beda. Tahap transformasi ini melibatkan normalisasi, pengkodean ulang variabel, atau perubahan format data ke bentuk yang konsisten untuk memungkinkan analisis yang efektif.
- e. Integrasi Data: Jika data berasal dari beberapa sumber, proses integrasi dilakukan. Ini memungkinkan penggabungan data dari berbagai sumber menjadi satu set data yang lengkap dan terintegrasi.
- f. Reduksi Dimensi Data: Untuk mengurangi kompleksitas, terutama pada data yang besar, reduksi dimensi dilakukan dengan menghilangkan atribut yang tidak relevan atau redundan, sehingga mempermudah analisis.

- g. Validasi Data: Langkah terakhir adalah memverifikasi keakuratan, kekonsistenan, dan keabsahan data. Proses validasi ini memastikan bahwa data yang telah diolah siap untuk digunakan dalam analisis tanpa adanya kesalahan yang mendasar.

Tahap pengolahan data merupakan fondasi penting sebelum dilakukannya analisis lebih lanjut. Proses yang cermat dan tepat dalam tahap ini membantu memastikan bahwa data yang digunakan untuk menjawab pertanyaan penelitian telah disiapkan dengan baik, yang pada gilirannya akan mempengaruhi kualitas dan keandalan kesimpulan yang dihasilkan dari penelitian.

2.3.1. Algoritma C4.5

Rumus untuk menghitung nilai gain dalam algoritma C4.5 adalah salah satu langkah kunci dalam pembentukan pohon keputusan. Gain digunakan untuk mengevaluasi atribut mana yang paling informatif atau yang paling baik dalam membagi data. Secara umum, nilai gain mengukur seberapa baik sebuah atribut dalam memisahkan data menjadi kelas yang lebih homogen[10].

Dalam konteks penelitian yang menggunakan algoritma C4.5 untuk memprediksi kanker paru-paru, perhitungan nilai gain akan digunakan untuk menentukan atribut mana yang paling signifikan dalam memprediksi atau mengelompokkan data yang diberikan. Ini bisa berarti mengidentifikasi atribut apa yang paling kuat dalam menentukan apakah seseorang memiliki risiko tinggi atau rendah terhadap kanker paru-paru. Secara matematis, rumus untuk menghitung nilai gain pada C4.5 adalah sebagai berikut:

Rumus pertama adalah:

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

Keterangan :

S : Himpunan Kasus

A : Atribut

n : Jumlah Partisi Atribut A

|S_i| : Jumlah kasus pada partisi ke-i

|S| : Jumlah Kasus dalam S

Dan rumus yang kedua adalah :

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \log_2 p_i$$

Keterangan :

S : Himpunan kasus

n : Banyaknya partisi S

p_i : probabilitas yang didapat dari kasus i dibagi total kasus.

Perhitungan nilai gain digunakan untuk menentukan atribut mana yang paling baik dalam membagi data menjadi subset yang lebih homogen terkait kelas target, dalam konteks ini, apakah seseorang berisiko tinggi atau rendah terhadap kanker paru-paru berdasarkan atribut-atribut yang diberikan.

Dalam implementasi algoritma C4.5, atribut dengan nilai gain tertinggi akan dipilih sebagai atribut untuk membuat keputusan di setiap node dalam pembentukan pohon keputusan.

2.3.2 RapidMiner

RapidMiner adalah salah satu perangkat lunak untuk pengolahan data mining yang bersifat open source atau terbuka. Perangkat ini dikembangkan menggunakan bahasa pemrograman Java, sehingga dapat digunakan pada berbagai sistem operasi[11]. RapidMiner dalam konteks text mining fokus pada analisis teks, yaitu kegiatan yang melibatkan ekstraksi pola-pola dari dataset besar dan menggabungkannya dengan metode statistika, kecerdasan buatan, dan database. Tujuan utama dari analisis teks ini adalah untuk menghasilkan informasi yang bermutu tinggi dari teks yang telah diolah.

RapidMiner adalah sebuah platform perangkat lunak yang mengkhususkan diri dalam analisis data. Dikembangkan dengan menggunakan bahasa pemrograman Java, RapidMiner dapat dioperasikan di berbagai sistem operasi dan menawarkan alat analisis data yang komprehensif tanpa perlu pengetahuan teknis yang mendalam. Fitur utama RapidMiner mencakup beragam kemampuan, termasuk pemodelan prediktif, analisis

statistik, dan terutama, penambangan data (data mining) termasuk analisis teks. Platform ini memungkinkan pengguna untuk menjalankan berbagai tugas analisis data dengan antarmuka yang intuitif, memungkinkan pengolahan data, pembuatan model prediktif, klustering, analisis teks, dan visualisasi data. RapidMiner digunakan secara luas di berbagai sektor industri seperti keuangan, kesehatan, e-commerce, dan pemasaran. Kemampuannya dalam analisis data memberikan wawasan yang berharga untuk pengambilan keputusan yang didukung oleh data. Platform ini tersedia dalam versi open source yang memungkinkan akses gratis untuk sebagian besar fungsionalitas dasar, dan memiliki komunitas pengguna yang aktif. Hal ini memungkinkan pengguna untuk berbagi pengetahuan, pengalaman, dan solusi terkait analisis data, serta memfasilitasi pertukaran informasi yang berguna[12]. RapidMiner membantu pengguna untuk memproses data, mengidentifikasi pola-pola, membangun model prediktif, dan membuat keputusan yang didukung oleh informasi yang dihasilkan dari analisis data. Dengan antarmuka yang mudah digunakan dan kemampuan yang luas, platform ini sangat berguna bagi para profesional dari berbagai bidang untuk menjalankan analisis data yang mendalam.

3. HASIL DAN PEMBAHASAN

Dalam penelitian ini metode yang digunakan adalah C4.5, dengan proses pengolahan data meliputi tahap seleksi dan transformasi data. Bagan ini akan digunakan sebagai panduan untuk memprediksi gejala kanker paru-paru.

3.1 Proses Pengumpulan Data

Dalam penelitian ini data dikumpulkan dengan menggunakan teknik pengambilan data sekunder yang tersedia di salah satu platform penyedia data yaitu Kaggle. Penggunaan data sekunder dipilih karena sumber data sudah tersedia sehingga tidak diperlukan observasi, wawancara atau survei untuk mengumpulkan data. Dataset yang digunakan untuk penelitian terdiri dari 150 record yang kemudian dibagi menjadi 80% untuk data training dan 20% untuk data testing. Berikut ini adalah tahapan proses decision tree menggunakan rapidminer dengan menggunakan data training terkait Ya dan Tidak hasil dari prediksi terkenanya kanker paru yang jumlahnya 120 record dengan hasil 57 “Ya” dan 63”Tidak”. Berikut Tabel 2 dibawah merupakan daftar atribut keseluruhan dari data set yang di peroleh dari Kaggle.

Tabel 2. daftar atribut data keseluruhan

Artibut	Tipe Data	Ketrangan
Usia	Kategorik	Tua/Muda
Jenis Kelamin	Kategorik	Pria/Wanita
Merokok	Kategorik	Pasif/Aktif
Bekerja	Kategorik	Tidak/Ya
Rumah Tangga	Kategorik	Tidak/YA
Alkohol	Kategorik	Jarang/Sering
Jari Kuning	Kategorik	Ada/Tidak
Aktivitas Bergadang	Kategorik	Seriing/Jarang
Aktivitas Olahraga	Kategorik	Ya/Tidak
Asuransi	Kategorik	Tidak/Ada
Penyakit Bawaan	Kategorik	Tidak/Ada
Hasil	Label	Ya/Tidak

Berikut adalah penjelsan mengenai Tabel 2:

- a. Usia: Data usia pasien dalam dataset ini diidentifikasi dengan atribut "Usia" dan diklasifikasikan menjadi "Tua" dan "Muda".
- b. Jenis_Kelamin: Atribut "Jenis_Kelamin" menggambarkan jenis kelamin pasien dalam dataset ini dengan opsi nilai "Pria" dan "Wanita".
- c. Merokok: Atribut "Merokok" menunjukkan status perokok pasien, dengan nilai yang bisa berupa "Pasif" atau "Aktif".
- d. Alkohol: Atribut "Alkohol" menunjukna sataus pasien sering tau jarang meminum alkohol
- e. Jari Kuning: Atribut "Ajri Kuning" menunjukan tingkat kesehatan pasien dengan berupa "ada" atau "tidak".
- f. Bekerja: Atribut "Bekerja" menunjukkan status pekerjaan pasien, dengan nilai "Ya" untuk yang bekerja dan "Tidak" untuk yang tidak bekerja.
- g. Aktivitas_Begadang: Atribut "Aktivitas_Begadang" mengindikasikan apakah pasien sering melakukan begadang atau tidak, dengan nilai "Ya" dan "Tidak".
- h. Aktivitas_Olahraga: Atribut "Aktivitas_Olahraga" mencerminkan kebiasaan berolahraga pasien, dengan nilai "Sering" atau "Jarang".
- i. Penyakit_Bawaan: Atribut "Penyakit_Bawaan" menggambarkan apakah pasien mempunyai riwayat penyakit bawaan atau tidak, dengan nilai "Ada" atau "Tidak".
- j. Hasil: Atribut "Hasil" adalah label yang mengidentifikasi apakah pasien menderita penyakit paru-paru, dengan nilai "Ya" (menderita kanker paru-paru) atau "Tidak" (tidak menderita kanker paru-paru).

3.2 Data Seletion

Langkah ini diambil karena beberapa atribut kurang relevan dalam memprediksi gejala penyakit paru. Proses ini bertujuan untuk memilih atribut yang paling tepat digunakan untuk memprediksi penyakit paru-paru. Tabel 3 dibawah ini berisi daftar atribut yang akan digunakan dan diproses oleh software RapidMiner.

Table 3. AtributTerpilih

No	Artibut Terpilih	Keterangan
1	Usia	Artibut Fitur
2	Jenis Kelamin	Artibut Fitur
3	Merokok	Artibut Fitur
4	Alkohol	Artibut Fitur
5	Jari Kuning	Artibut Fitur
6	Aktivitas Bergadang	Artibut Fitur
7	Aktivitas Olahraga	Artibut Fitur
8	Penyakit Bawaan	Artibut Fitur
9	Hasil	Label

Atribut yang telah terpilih adalah Usia, Jenis_Kelamin, Merokok, Alkohol, Jari Kuning Aktivitas_Begadang, Aktivitas_Olahraga, Penyakit_Bawaan, dan Hasil. Pemilihan atribut tersebut didasarkan pada korelasi mereka dengan gejala penyakit paru-paru.[13] Atribut – atribut ini akan diproses menggunakan algoritma C4.5 melalui software RapidMiner. Proses ini akan membuat pohon keputusan dan mengukur keakuratan prediksi kanker paru-paru. Langkah ini diambil karena beberapa atribut kurang relevan dalam memprediksi gejala penyakit paru. Proses ini bertujuan untuk memilih atribut yang paling tepat digunakan untuk memprediksi penyakit paru-paru. Tabel 3.2 dibawah ini berisi daftar atribut yang akan digunakan dan diproses oleh software RapidMiner

3.3 Data Transformation

Langkah transformasi ini melibatkan perubahan format data agar sesuai untuk diproses lebih lanjut dalam proses data mining. Transformasi dataset ini dilakukan menggunakan RapidMiner dengan mengubah tipe data dari polinomial menjadi binomial pada atribut yang digunakan, sebagaimana yang terlihat pada Gambar 3.1 data keseluruhan.

	kok ordinal	Bekerja polynomial	Rumah_Ta... polynomial	Aktivitas_B... polynomial	Aktivitas_O... polynomial	Asuransi polynomial	Penyakit_B... polynomial	Hasil polynomial label
1		Tidak	Ya	Ya	Sering	Ada	Tidak	Ya
2		Tidak	Ya	Ya	Jarang	Ada	Ada	Tidak
3		Tidak	Ya	Ya	Jarang	Ada	Tidak	Tidak
4		Ya	Tidak	Tidak	Jarang	Ada	Ada	Tidak
5		Ya	Tidak	Tidak	Sering	Tidak	Ada	Ya
6		Ya	Tidak	Tidak	Sering	Tidak	Ada	Tidak
7		Tidak	Ya	Tidak	Sering	Tidak	Tidak	Ya
8		Tidak	Ya	Ya	Sering	Tidak	Tidak	Tidak
9		Ya	Ya	Ya	Jarang	Ada	Ada	Ya
10		Ya	Tidak	Ya	Jarang	Ada	Ada	Ya
11		Ya	Ya	Tidak	Sering	Ada	Ada	Ya
12		Tidak	Ya	Tidak	Jarang	Ada	Tidak	Tidak
13		Tidak	Ya	Ya	Jarang	Ada	Tidak	Tidak
14		Ya	Tidak	Ya	Jarang	Ada	Ada	Tidak
15		Ya	Tidak	Ya	Sering	Tidak	Ada	Ya
16		Ya	Tidak	Ya	Jarang	Ada	Ada	Ya
17		Ya	Ya	Tidak	Sering	Ada	Ada	Ya

Gambar 2. Data yang terdapat dari kaggle

3.4 Data Mining

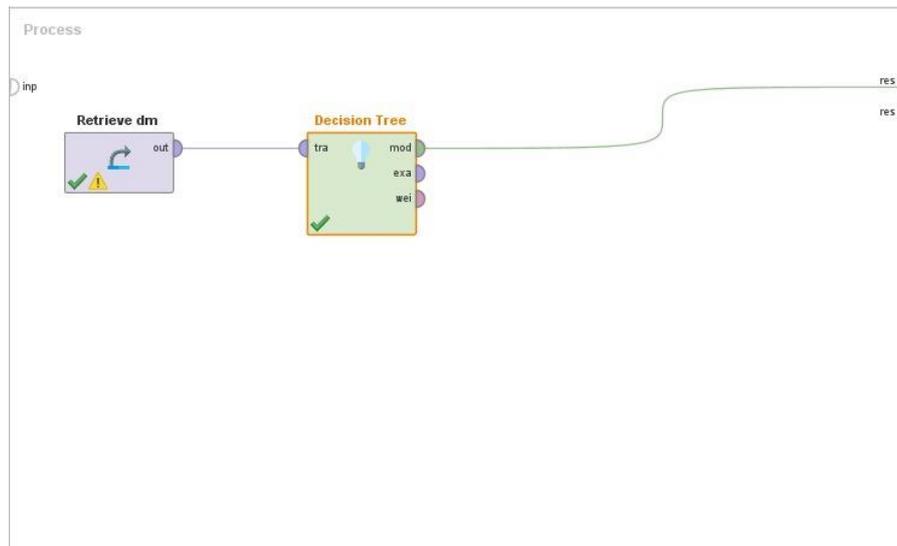
Algoritma C4.5 adalah salah satu algoritma populer dalam bidang data mining yang digunakan untuk membangun model prediksi berdasarkan kumpulan data. Tujuan utamanya adalah untuk melakukan analisis terhadap data yang telah melalui proses seleksi dan transformasi guna membuat prediksi mengenai kemungkinan pasien terkena kanker paru-paru. Proses data mining dimulai dengan pengumpulan data dari berbagai sumber seperti catatan medis, tes laboratorium, riwayat penyakit, gaya hidup, dan faktor risiko lainnya yang relevan dengan kanker paru-paru. Data-data ini kemudian menjalani serangkaian tahapan, termasuk pembersihan data (data cleansing) untuk menghilangkan data yang tidak lengkap, tidak akurat, atau tidak relevan. Selanjutnya, data tersebut disiapkan melalui tahap transformasi untuk mengonversi data mentah menjadi format yang sesuai untuk analisis.

Algoritma C4.5 bekerja dengan membangun pohon keputusan (decision tree) dari data yang telah dipersiapkan[14]. Pohon keputusan adalah representasi grafis dari semua kemungkinan keputusan yang mungkin diambil berdasarkan atribut-atribut yang ada. C4.5 bekerja dengan mengidentifikasi atribut terpenting dalam data untuk membuat keputusan yang optimal. Algoritma ini menggunakan metode pembelajaran mesin untuk memahami pola-pola dalam data yang memungkinkan pengklasifikasian pasien sebagai penderita kanker paru-paru atau tidak. Saat model pohon keputusan sudah dibangun, algoritma ini memungkinkan untuk melakukan prediksi terhadap pasien baru berdasarkan atribut-atribut yang dimiliki. Misalnya, atribut seperti usia, riwayat merokok, paparan asap rokok, riwayat keluarga, dan hasil tes medis dapat menjadi faktor-faktor yang digunakan oleh model untuk memprediksi kemungkinan seseorang terkena kanker paru-paru. Pada akhirnya, model yang dihasilkan oleh algoritma C4.5 akan dapat membantu dalam memberikan informasi prediktif yang berguna kepada para profesional medis atau peneliti untuk mengidentifikasi risiko kanker paru-paru pada pasien berdasarkan informasi yang ada. Dengan demikian, ini dapat membantu dalam upaya pencegahan, deteksi dini, dan penanganan penyakit secara lebih efisien.

3.5 Implementasi menggunakan RapidMiner

RapidMiner merupakan salah satu perangkat lunak populer yang digunakan dalam proses analisis data, termasuk dalam konteks data mining. Dalam konteks spesifik Anda, penggunaan RapidMiner bertujuan untuk memproses data menggunakan algoritma C4.5. Setelah proses ini selesai, diharapkan akan dihasilkan pohon keputusan dan aturan-aturan terkait dari analisis tersebut. Proses yang melibatkan algoritma C4.5 dalam RapidMiner dimulai dengan tahap pengolahan data. Data yang telah disiapkan sebelumnya, yang mungkin telah melalui tahap pembersihan, transformasi, dan pemilihan atribut, akan dimuat ke dalam perangkat lunak. RapidMiner akan membantu Anda dalam menyiapkan data tersebut, memungkinkan visualisasi yang lebih baik, serta memfasilitasi penggunaan algoritma C4.5 untuk membangun model pohon keputusan.

Algoritma C4.5 akan bekerja pada RapidMiner dengan melakukan analisis terhadap data yang dimasukkan, mengidentifikasi pola-pola yang ada, dan membangun model pohon keputusan berdasarkan atribut-atribut yang relevan. Proses ini melibatkan penentuan atribut terpenting yang mempengaruhi hasil prediksi terkait kemungkinan pasien terkena kanker paru-paru. Saat algoritma C4.5 selesai dieksekusi, RapidMiner akan menghasilkan pohon keputusan yang dapat divisualisasikan. Pohon keputusan ini akan memperlihatkan cabang-cabang keputusan berdasarkan atribut-atribut yang paling signifikan. Selain pohon keputusan, RapidMiner juga dapat menyajikan aturan-aturan yang terkait dengan model yang telah dibangun. Aturan-aturan ini merupakan representasi dari kondisi-kondisi yang mempengaruhi keputusan prediksi terkait kanker paru-paru berdasarkan data yang diolah[15]. Saat menggunakan RapidMiner dengan algoritma C4.5, interpretasi model pohon keputusan dan aturan-aturan ini penting untuk memahami faktor-faktor yang mempengaruhi prediksi kanker paru-paru. Informasi ini bisa memberikan wawasan yang berharga kepada para profesional kesehatan atau peneliti dalam mengidentifikasi faktor-faktor risiko yang perlu diperhatikan serta membantu dalam pengambilan keputusan yang lebih tepat terkait dengan pencegahan, diagnosis, dan perawatan kanker paru-paru. akan dihasilkan pohon keputusan seperti yang terlihat pada gambar di bawah ini[13].

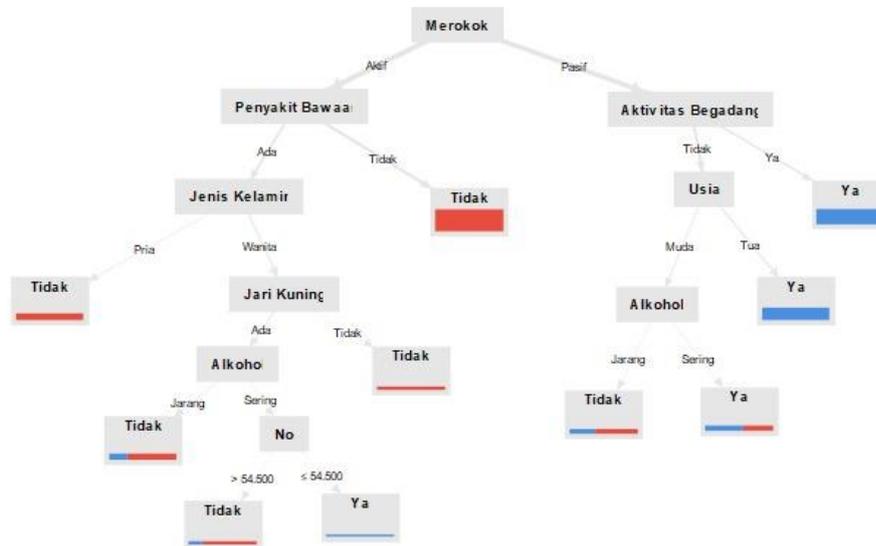


Gambar 3.. penggunaan rapid miner

Pada gambar di atas merupakan rumus untuk perhitungan menggunakan rapidminer

a. Pohon Keputusan.

Dari hasil pengujian pada RapidMiner, diperoleh pohon keputusan seperti yang terlihat pada Gambar 4 di bawah ini.



Gambar 4. Pohon Keputusan

Hasil dari pohon keputusan ini yaitu :

1. Perokok aktif
 - a. Jika merokk aktif dan memepunyai penyakit bawaan maka dia ya, dia dapat menghidap penyakit paru.
 - b. Jika merokok tetapi tidak mempunyai penyakit bawaan maka dia tidak terkena penyakit paru
 - c. Jika dia merokok aktif dan mempunyai penyakit bawaan dan berjenis kelamin laki laki maka dia tidak terkena penyakit paru
 - d. Jika dia mereokok aktif dan mempunyaipenyakit bawaan dan berjenis kelamin perempuan maka dia terkena penyakit paru
 - e. Jika dia meroko aktif dan mempunyai penyakit bawaan dan berjenis kelamin perempuan tetapi tidak berjari kuning maka dia tidak terkena penyakit paru
 - f. Jika dia meroko aktif dan mempunyai penyakit bawaan dan berjenis kelamin perempuan mempunyai jari kuning maka dia terkena penyakit paru
 - g. Jika dia merokok aktif dan mempunyai penyakit bawaan dan berjenis kelamin perempuan dan mempunyai jari kuning tetapi dia mengkonsumsi alkohol maka dia terkena penyakit paru
 - h. Jia dia merokok aktif mempunyai penyakit bawaan dan berjenis kelamin perempuan dan mempunyai jari kuning dan tidak mengkonsumsi alkohol maka dia tidak terkena penyakit paru
2. Perokok pasif
 - a. Jika dia merokok pasif tidak bergadang maka dia tidak terkena penyakit paru
 - b. Jika dia merokok pasif dan bergadang maka dia dapat terkena penyakit paru
 - c. Jika dai merokok pasif dan bergadang dan berusia muda maka dia terkena penyakit paru
 - d. Jika dia meroko pasif dan bergadang dan berusia tua maka dia tidak terkena penyait paru
 - e. Jika dia merokok pasif dan bergadang, berusia tua dan mengkonsumsi alkohol maka dia terkena penyakit paru
 - f. Jika dia merokok pasif dan bergadang, berusia tua dan jarang mengkonsumsi alkohol maka dia tidak terkena penyakit paru

b. Deskripsi Pohon Keputusan

Berdasarkan hasil pohon keputusan di atas dapat disimpulkan bahwa telah terjadi beberapa kondisi atau aturan. Deskripsi pohon keputusan di Rapidminer berikut ini dapat Anda lihat pada gambar 5 dibawah ini.

Tree

```

Merokok = Aktif
| Penyakit Bawaan = Ada
| | Jenis Kelamin = Pria: Tidak {Ya=0, Tidak=12}
| | Jenis Kelamin = Wanita
| | | Jari Kuning = Ada
| | | | Alkohol = Jarang : Tidak {Ya=3, Tidak=8}
| | | | Alkohol = Sering
| | | | No > 54.500: Tidak {Ya=1, Tidak=4}
| | | | No ≤ 54.500: Ya {Ya=2, Tidak=0}
| | | | Jari Kuning = Tidak: Tidak {Ya=0, Tidak=5}
| Penyakit Bawaan = Tidak: Tidak {Ya=0, Tidak=44}
Merokok = Pasif
| Aktivitas Begadang = Tidak
| | Usia = Muda
| | | Alkohol = Jarang : Tidak {Ya=3, Tidak=5}
| | | Alkohol = Sering: Ya {Ya=5, Tidak=4}
| | Usia = Tua: Ya {Ya=24, Tidak=0}
| Aktivitas Begadang = Ya: Ya {Ya=30, Tidak=0}
    
```

Gambar 5. Deskripsi pohon keputusan

Berdasarkan gambar diatas, dapat dijelaskan bahwa dari 9 atribut penilaian terhadap prediksi kanker paru-paru, terdapat 1atribut yang memiliki pengaruh paling signifikan terhadap prediksi kanker paru-paru yaitu atribut merokok. Dalam konteks ini, dilakukan pengujian terhadap kebenaran data yang digunakan, yaitu apakah metode klasifikasi C4.5 dapat diterapkan untuk menilai prediksi terkena kanker paru-paru dengan bantuan perangkat lunak RapidMiner.

4. KESIMPULAN

Penelitian ini bertujuan untuk mengembangkan model prediksi penyakit paru-paru menggunakan algoritma C4.5 dalam konteks data mining. Dengan menggabungkan teknik data mining dan algoritma C4.5, penelitian ini memberikan kontribusi dalam upaya pencegahan dan pengelolaan penyakit paru-paru, khususnya kanker paru-paru. Hasil penelitian menunjukkan bahwa algoritma C4.5 mampu membangun model prediksi penyakit paru-paru dengan tingkat akurasi yang memadai. Proses analisis data, seleksi atribut, dan transformasi data memastikan kualitas dataset yang digunakan dalam pengembangan model. Atribut yang terpilih, seperti usia, jenis kelamin, kebiasaan merokok, konsumsi alkohol, dan faktor-faktor lainnya, secara signifikan berkontribusi terhadap kemampuan prediktif model. Model ini diharapkan dapat membantu identifikasi dini individu yang berisiko tinggi terkena penyakit paru-paru, terutama kanker paru-paru. Dengan deteksi dini, intervensi dan pengobatan dapat dimulai lebih awal, meningkatkan peluang kesembuhan dan pengendalian penyakit. Penerapan teknologi data mining, khususnya algoritma C4.5, memberikan bukti bahwa pendekatan ini dapat menjadi alat yang efektif dalam praktik medis dan bidang kesehatan. Namun demikian, penelitian ini juga mengakui beberapa keterbatasan, termasuk ketergantungan pada dataset yang tersedia, kompleksitas faktor-faktor penyebab penyakit paru-paru, dan kebutuhan untuk validasi lebih lanjut dengan dataset yang lebih luas. Pengembangan model ini sebaiknya terus menerus diperbarui dan dioptimalkan seiring dengan peningkatan pemahaman tentang faktor risiko penyakit paru-paru. Dalam konteks pencegahan penyakit paru-paru, termasuk kanker paru-paru, upaya kolektif untuk meningkatkan kesadaran masyarakat, mengedukasi tentang gaya hidup sehat, dan promosi kebijakan kesehatan yang mendukung sangat penting. Integrasi model prediksi ini dengan praktik klinis dan kebijakan kesehatan dapat memberikan dampak positif dalam mengurangi beban penyakit paru-paru di masyarakat. Secara keseluruhan, penelitian ini memberikan kontribusi pada pengembangan teknologi di bidang kesehatan, khususnya dalam deteksi dini penyakit paru-paru. Dengan pendekatan yang terintegrasi antara data mining dan algoritma C4.5, diharapkan penelitian ini dapat menjadi landasan untuk pengembangan lebih lanjut dalam upaya pencegahan dan penanganan penyakit paru-paru secara dini.

UCAPAN TERIMAKASIH

1. Terima kasih kepada semua pihak yang memberikan dukungan tak terhingga dalam menyelesaikan penelitian ini.
2. Ucapan terima kasih kepada Allah SWT atas rahmat dan petunjuk-Nya yang telah melimpahkan kesuksesan dalam penelitian ini.
3. Penghargaan yang tulus kepada kedua orang tua yang senantiasa memberikan doa dan dukungan yang tak henti-hentinya.
4. Terima kasih kepada dosen pembimbing yang telah memberikan arahan, bimbingan, dan wawasan berharga yang menjadi kunci kesuksesan penelitian ini.
5. Terimakasih kepada rekan-rekan seperjuangan di kelompok 3B atas dukungan dan kerjasama keras dalam menyelesaikan penelitian ini.

REFERENCES

- [1] L. Sari, A. Romadloni, and R. Listyaningrum, "Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest," *Infotekmesin*, vol. 14, no. 1, pp. 155–162, Jan. 2023, doi: 10.35970/infotekmesin.v14i1.1751.
- [2] J. Homepage *et al.*, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Implementation of Decision Tree Algorithm and Support Vector Machine for Lung Cancer Classification Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru," vol. 3, pp. 15–19, 2023.
- [3] C. A. Rahman and A. Kudus, "Penggunaan Metode K Nearest Neighborhood untuk Imputasi Data Tersensor Kanan pada Pasien Kanker Paru-Paru Sel Kecil," *Bandung Conference Series: Statistics*, vol. 2, no. 2, pp. 441–448, Aug. 2022, doi: 10.29313/bcss.v2i2.4615.
- [4] Angelina M. T. I. Sambu *et al.*, "Penggunaan Bahasa Pemrograman Python Dalam Analisis Faktor Penyebab Kanker Paru-Paru," *Jurnal Publikasi Teknik Informatika*, vol. 2, no. 2, pp. 88–99, Jul. 2023, doi: 10.55606/jupti.v2i2.1742.
- [5] A. Rifa'i and Y. Prabowo, "Krea-TIF: Jurnal Teknik Informatika Diagnosis Kanker Paru-Paru dengan Sistem Fuzzy," vol. 10, no. 1, pp. 19–28, 2022, doi: 10.32832/kreatif.v10i1.6317.
- [6] A. Desiani *et al.*, "Perbandingan Klasifikasi Penyakit Kanker Paru-Paru menggunakan Support Vector Machine dan K-Nearest Neighbor," *Jurnal PROCESSOR*, vol. 18, no. 1, Apr. 2023, doi: 10.33998/processor.2023.18.1.700.
- [7] P. Rawinda Meliala, "Perbandingan Algoritma Machine Learning untuk Survivabilitas dan Biaya Pengobatan Pasien Kanker Paru-paru di Taiwan," 2021.
- [8] M. Abdul, R. Wahid, A. Nugroho, and A. H. Anshor, "Prediksi Penyakit Kanker Paru-Paru Dengan Algoritma Regresi Linier," *Bulletin of Information Technology (BIT)*, vol. 4, no. 1, pp. 63–74, 2023, doi: 10.47065/bit.v3i1.
- [9] Bambang Pulu Hartato, "Penerapan Convolutional Neural Network pada Citra Rontgen Paru-Paru untuk Deteksi SARS-CoV-2," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 747–759, Aug. 2021, doi: 10.29207/resti.v5i4.3153.
- [10] M. Y. Haffandi, E. Haerani, F. Syafria, and L. Oktavia, "KLASIFIKASI PENYAKIT PARU-PARU DENGAN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER," *Jurnal Teknik Informasi dan Komputer (Tekinkom)*, vol. 5, no. 2, p. 176, Dec. 2022, doi: 10.37600/tekinkom.v5i2.649.
- [11] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, "ANALISIS DATA MINING DATA NETFLIX MENGGUNAKAN APLIKASI RAPID MINER," *JBASE - Journal of Business and Audit Information Systems*, vol. 4, no. 1, Apr. 2021, doi: 10.30813/jbase.v4i1.2729.
- [12] S. Widaningsih, "PERBANDINGAN METODE DATA MINING UNTUK PREDIKSI NILAI DAN WAKTU KELULUSAN MAHASISWA PRODI TEKNIK INFORMATIKA DENGAN ALGORITMA C4,5, NAÏVE BAYES, KNN DAN SVM," *Jurnal Tekno Insentif*, vol. 13, no. 1, pp. 16–25, Apr. 2019, doi: 10.36787/jti.v13i1.78.
- [13] F. Faisal Nugraha, I. Sunandar, and C. Juliane, "Penerapan Data Mining Dengan Metode Klasifikasi Menggunakan Algoritma C4.5," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 4, 2022, [Online]. Available: <http://jurnal.mdp.ac.id>

- [14] M. Zuhra Almasah and A. Wahyu Wijayanto, "Eigen Mathematics Journal Comparison of Data Mining Methods in Classifying Village Status of Purwakarta and West Bandung Regencies (Podes 2021)," vol. 6, no. 1, 2023, doi: 10.29303/emj.v6i1.156.
- [15] M. R. Nahjan, N. Heryana, and A. Voutama, "IMPLEMENTASI RAPIDMINER DENGAN METODE CLUSTERING K-MEANS UNTUK ANALISA PENJUALAN PADA TOKO OJ CELL," 2023.