Klasifikasi Penyakit Kanker Paru-Paru Menggunakan Metode C4.5

Khairudin[1], Syaeful Machfud[2], Yono Cahyono[3]*

[1,2,3]Fakultas Ilmu Komputer Program Studi Teknik Informatika, Universitas Pamulang

Korespondensi : imandayantinureza@gmail.com

ABSTRACT

Lung cancer is one of the leading causes of death worldwide. Early detection of this disease is crucial to improving patients' chances of recovery. In this study, lung cancer prediction was conducted using the C4.5 algorithm applied to a secondary dataset from Kaggle. The dataset consists of 150 records, divided into 80% for training data and 20% for testing data. The classification process was performed using RapidMiner software. The results indicate that smoking is the most significant factor influencing lung cancer prediction. The generated decision tree model demonstrated a relatively high accuracy level, making it a useful tool for assisting in the early detection of lung cancer in patients.

Article History

Received : 10-10-2024
Revised : 10-11-2024
Accepted : 25-12-2024

Keywords

Lung Cancer
Risk Classification
Decision Tree

ABSTRAK

Kanker paru-paru merupakan salah satu penyakit yang memiliki tingkat kematian tinggi di dunia. Deteksi dini terhadap penyakit ini sangat penting untuk meningkatkan peluang kesembuhan pasien. Dalam penelitian ini, dilakukan prediksi kanker paru-paru menggunakan algoritma C4.5 yang diterapkan pada dataset sekunder dari Kaggle. Data yang digunakan terdiri dari 150 record, yang dibagi menjadi 80% untuk data training dan 20% untuk data testing. Proses klasifikasi dilakukan menggunakan perangkat lunak RapidMiner. Hasil penelitian menunjukkan bahwa atribut merokok memiliki pengaruh paling signifikan terhadap prediksi kanker paru-paru. Model pohon keputusan yang dihasilkan memiliki tingkat akurasi yang cukup tinggi, sehingga dapat digunakan sebagai alat bantu dalam mendeteksi kemungkinan pasien menderita kanker paru-paru.

PENDAHULUAN

Kanker paru-paru merupakan salah satu penyakit dengan tingkat kematian yang tinggi di dunia. Penyakit ini sering kali terdeteksi pada tahap lanjut, sehingga peluang penyembuhannya menjadi lebih kecil. Faktor risiko utama yang berkontribusi terhadap kanker paru-paru meliputi kebiasaan merokok, paparan asap rokok, konsumsi alkohol, kurangnya aktivitas fisik, serta faktor genetik. Oleh karena itu, diperlukan suatu metode yang dapat membantu dalam mendeteksi risiko kanker paru-paru secara lebih dini agar dapat dilakukan tindakan pencegahan dan pengobatan yang tepat.[1]

Kemajuan teknologi dalam bidang kecerdasan buatan (Artificial Intelligence) dan data mining memberikan peluang besar dalam analisis dan prediksi berbagai penyakit, termasuk kanker paru-paru. Salah satu metode yang banyak digunakan dalam klasifikasi data medis adalah algoritma C4.5, yang mampu membangun model prediksi berdasarkan pola dalam data. Dengan metode ini, faktor-faktor risiko dari pasien dapat dianalisis untuk menghasilkan keputusan yang dapat membantu tenaga medis dalam menilai kemungkinan seseorang terkena kanker paru-paru.[2]

Beberapa penelitian terdahulu telah membuktikan efektivitas metode data mining dalam prediksi penyakit, khususnya kanker paru-paru:

1. Penelitian oleh Prasetyo et al. (2020) mengimplementasikan algoritma C4.5 dalam prediksi kanker paru-paru menggunakan data rekam medis. Hasil penelitian menunjukkan bahwa metode ini memiliki tingkat akurasi yang cukup tinggi dalam mengklasifikasikan pasien berdasarkan faktor risiko seperti usia, kebiasaan merokok, dan riwayat kesehatan keluarga.
2. Penelitian oleh Susanti dan Wijaya (2021) membandingkan beberapa algoritma klasifikasi seperti Naïve Bayes, Decision Tree, dan Random Forest dalam memprediksi kanker paru-

paru. Hasilnya menunjukkan bahwa algoritma Decision Tree (C4.5) memiliki performa yang lebih baik dalam hal interpretasi hasil dan akurasi dibandingkan metode lainnya.

3. Penelitian oleh Rahman et al. (2022) menggunakan pendekatan Machine Learning berbasis Support Vector Machine (SVM) untuk mendeteksi kanker paru-paru dari dataset pasien. Penelitian ini menunjukkan bahwa SVM mampu memberikan akurasi tinggi, tetapi memerlukan waktu komputasi yang lebih lama dibandingkan Decision Tree.

Berdasarkan penelitian-penelitian terdahulu tersebut, dapat disimpulkan bahwa algoritma C4.5 merupakan salah satu metode yang efektif dalam mengklasifikasikan risiko kanker paru-paru. Oleh karena itu, penelitian ini menggunakan data sekunder dari Kaggle yang telah dikategorikan ke dalam berbagai atribut terkait risiko kanker paru-paru. Dengan memanfaatkan RapidMiner sebagai alat analisis, algoritma C4.5 diterapkan untuk membangun pohon keputusan guna mengidentifikasi hubungan antara faktor-faktor risiko dan kemungkinan seseorang terkena kanker paru-paru.

Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam dunia medis, khususnya dalam membantu mendeteksi risiko kanker paru-paru lebih awal serta memberikan wawasan bagi tenaga kesehatan dalam mengambil tindakan preventif yang lebih efektif.

TINJAUAN PUSTAKA

Penyakit Paru-Paru

Penyakit paru-paru merupakan istilah luas yang mencakup berbagai gangguan yang menyerang sistem pernapasan, terutama paru-paru dan saluran pernapasan. Paru-paru memiliki peran penting dalam proses pertukaran gas, yaitu menyerap oksigen dari udara dan mengeluarkan karbon dioksida dari tubuh. Gangguan pada organ ini dapat menyebabkan kesulitan bernapas, menurunnya kadar oksigen dalam darah, serta berdampak serius pada kesehatan secara keseluruhan. Penyakit paru-paru dapat bersifat akut, seperti pneumonia, atau bersifat kronis, seperti penyakit paru obstruktif kronis (PPOK). Berbagai faktor dapat menjadi penyebabnya, termasuk infeksi, paparan zat berbahaya, serta faktor genetik.[3]

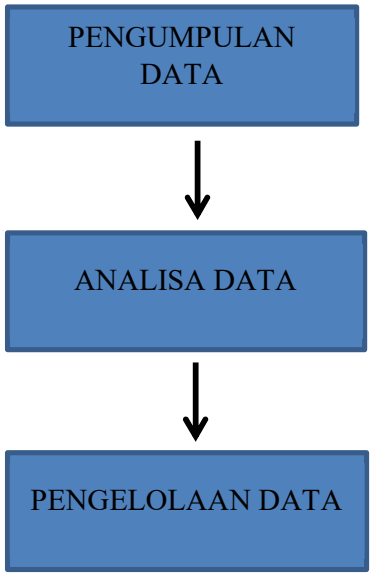
Decision Tree

Decision Tree adalah struktur pohon yang digunakan untuk membuat keputusan berdasarkan serangkaian aturan. Setiap simpul dalam pohon mewakili keputusan berdasarkan fitur tertentu. cabang (branches) yang menunjukkan hasil dari tes tersebut, dan simpul daun (leaf nodes) yang merepresentasikan prediksi akhir atau hasil klasifikasi. Decision Tree bekerja dengan membagi data ke dalam subset yang lebih kecil berdasarkan nilai fitur, menggunakan proses pembagian berulang (recursive splitting). Model ini membagi dataset menjadi subset yang lebih kecil berdasarkan nilai fitur, sehingga memudahkan klasifikasi atau regresi.[4] Proses pembagian ini berlangsung secara rekursif, dimulai dari simpul akar, di mana seluruh dataset dipertimbangkan. Setiap simpul pohon kemudian mengevaluasi fitur tertentu untuk menentukan bagaimana data sebaiknya dibagi. Sering digunakan dalam analisis kredit, diagnosis medis, dan analisis risiko.[5]

METODE

Dalam penelitian ini, kami menerapkan teknik data mining yang menggunakan metode Decision Tree[6] dengan algoritma C4.5. Fokus utama kami adalah untuk mencapai tingkat akurasi yang tepat dalam memprediksi penyakit paru-paru. 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.

ISSN: 2774-4345, DOI: 10.31284/j.kernel.2024.v5i2.7315



Gambar 1. Metode Penelitian

Pengumpulan Data :

1. Dataset ini diambil dari situs web Kaggle dan terdiri dari 150 dataset yang relevan dengan kanker paru-paru.
2. 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.

A. Analisa Data : data yang diperoleh kan di analisa dan di olah menggunakan metode Decision Tree
B. Pengelolaan Data
    1. 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.
    2. Ketelitian Data: Proses pengumpulan data dilakukan dengan memastikan konsistensi, validitas, dan integritas data. Setiap atribut yang dikumpulkan harus memiliki dokumentasi dan spesifikasi yang jelas.

3. 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.
4. 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.

HASIL DAN PEMBAHASAN

1. Proses Pengumpulan Data

Penelitian ini menggunakan teknik pengumpulan data sekunder yang diperoleh dari platform penyedia data, Kaggle. Pemilihan data sekunder dilakukan karena sumber data telah tersedia, sehingga tidak diperlukan proses observasi, wawancara, atau survei untuk mengumpulkan data. Dataset yang digunakan dalam penelitian ini berjumlah 150 record, yang kemudian dibagi menjadi 80% untuk data pelatihan dan 20% untuk data pengujian. Proses pembentukan decision tree dilakukan menggunakan RapidMiner dengan data pelatihan yang berisi 120 record, terdiri dari 57 kasus dengan hasil “Ya” dan 63 kasus dengan hasil “Tidak” dalam prediksi kanker paru-paru. Tabel 2 berikut menunjukkan daftar atribut lengkap dari dataset yang diperoleh melalui Kaggle.

Tabel 1. Daftar Atribut Data Keseluruhan

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
Alkohol	Kategorik	Jarang/Sering
Jari Kuning	Kategorik	Ada/Tidak
Aktivitas Bergadang	Kategorik	Sering/Jarang
Aktivitas Olahraga	Kategorik	Ya/Tidak
Asuransi	Kategorik	Tidak/Ada
Penyakit Bawaan	Kategorik	Tidak/Ada
Hasil	Kategorik	Ya/Tidak

Berikut adalah penjelasan mengenai Tabel 2:

- Usia: Data usia pasien dalam dataset ini diidentifikasi dengan atribut "Usia" dan diklasifikasikan menjadi "Tua" dan "Muda".
- Jenis_Kelamin: Atribut "Jenis_Kelamin" menggambarkan jenis kelamin pasien dalam dataset ini dengan opsi nilai "Pria" dan "Wanita".
- Merokok: Atribut "Merokok" menunjukkan status perokok pasien, dengan nilai yang bisa berupa "Pasif" atau "Aktif".
- Alkohol: Artibut “Alkohol” menunjukna sataus pasien sering tau jarang meminum alkohol

ISSN: 2774-4345, DOI: 10.31284/j.kernel.2024.v5i2.7315

- Jari Kuning: Artibut"Ajri Kuning" menunjukan tingkat kesehatan pasien dengan berupa "ada" atau"tidak".
- Bekerja: Atribut "Bekerja" menunjukkan status pekerjaan pasien, dengan nilai "Ya" untuk yang bekerja dan "Tidak" untuk yang tidak bekerja.
- Aktivitas_Begadang: Atribut "Aktivitas_Begadang" mengindikasikan apakah pasien sering melakukan begadang atau tidak, dengan nilai "Ya" dan "Tidak".
- Aktivitas_Olahraga: Atribut "Aktivitas_Olahraga" mencerminkan kebiasaan berolahraga pasien, dengan nilai "Sering" atau "Jarang".
- Penyakit_Bawaan: Atribut "Penyakit_Bawaan" menggambarkan apakah pasien mempunyai riwayat penyakit bawaan atau tidak, dengan nilai "Ada" atau "Tidak".
- 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).

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.

Tabel 2. Atribut Terpilih

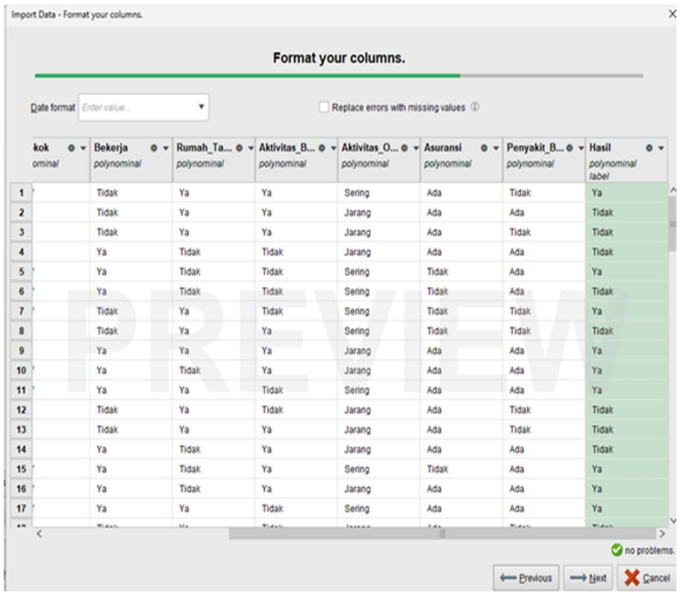
No	Artibut Terpilih	Keterangan
1	Usia	Artibut Fitur
2	Jenis Kelamin	Artibut Fitur
3	Merokok	Artibut Fitur
4	Alkohol	Artibut Fitur
5	Jari Kuning	Artibut Fitur
6	Aktivitas Bergadang	Artibut Fitur
7	Aktivitas Olahraga	Artibut Fitur
8	Penyakit Bawaan	Artibut Fitur

Atribut yang dipilih dalam penelitian ini meliputi Usia, Jenis Kelamin, Merokok, Alkohol, Jari Kuning, Aktivitas Begadang, Aktivitas Olahraga, Penyakit Bawaan, dan Hasil. Pemilihan atribut tersebut didasarkan pada tingkat korelasinya dengan gejala penyakit paru-paru. Atribut-atribut ini kemudian diproses menggunakan algoritma C4.5 dalam perangkat lunak RapidMiner untuk membangun pohon keputusan serta mengevaluasi akurasi prediksi kanker paru-paru. Pemilihan ini dilakukan karena beberapa atribut lainnya dianggap kurang relevan dalam proses prediksi. Tujuan utama dari langkah ini adalah menentukan atribut yang paling efektif dalam memprediksi penyakit paru-paru. Tabel 3 berikut menyajikan daftar atribut yang akan digunakan dalam analisis menggunakan RapidMiner.

ISSN: 2774-4345, DOI: 10.31284/j.kernel.2024.v5i2.7315

2. 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



Gambar 2. Data yang terdapat dari kaggle

3. Data Mining

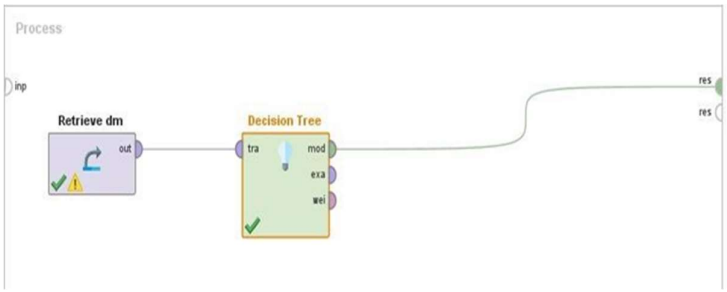
Algoritma C4.5 merupakan salah satu metode yang banyak digunakan dalam data mining untuk membangun model prediksi berdasarkan sekumpulan data. Tujuan utamanya adalah menganalisis data yang telah diseleksi dan ditransformasikan guna memprediksi kemungkinan pasien menderita kanker paru-paru. Proses data mining dimulai dengan mengumpulkan data dari berbagai sumber, seperti rekam medis, hasil tes laboratorium, riwayat kesehatan, gaya hidup, serta faktor risiko lainnya yang berhubungan dengan kanker paru-paru. Data tersebut kemudian melalui tahap pembersihan (data cleansing) untuk menghilangkan informasi yang tidak lengkap, tidak akurat, atau tidak relevan. Selanjutnya, data diubah dalam tahap transformasi agar sesuai untuk analisis lebih lanjut.

Algoritma C4.5[7] bekerja dengan membangun pohon keputusan (decision tree) berdasarkan data yang telah diproses. Pohon keputusan ini merepresentasikan berbagai kemungkinan keputusan yang dapat diambil berdasarkan atribut-atribut yang tersedia. Algoritma ini menentukan atribut paling signifikan dalam data untuk menghasilkan keputusan yang optimal. Dengan menggunakan metode pembelajaran mesin, C4.5 mengidentifikasi pola dalam data untuk mengklasifikasikan pasien sebagai penderita kanker paru-paru atau bukan. Setelah pohon keputusan terbentuk, model ini dapat digunakan untuk memprediksi kondisi pasien baru berdasarkan atribut yang dimilikinya. Misalnya, faktor seperti usia, riwayat merokok, paparan asap rokok, riwayat keluarga, dan hasil tes medis dapat menjadi variabel yang digunakan untuk memperkirakan risiko seseorang terkena kanker paru-paru.

Pada akhirnya, model yang dihasilkan oleh algoritma C4.5 dapat memberikan informasi prediktif yang bermanfaat bagi tenaga medis atau peneliti dalam mengidentifikasi risiko kanker paru-paru berdasarkan data yang tersedia. Dengan demikian, algoritma ini berkontribusi dalam upaya pencegahan, deteksi dini, serta perencanaan penanganan penyakit yang lebih efektif.

4. 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 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.



Gambar 3. Penggunaan rapid miner

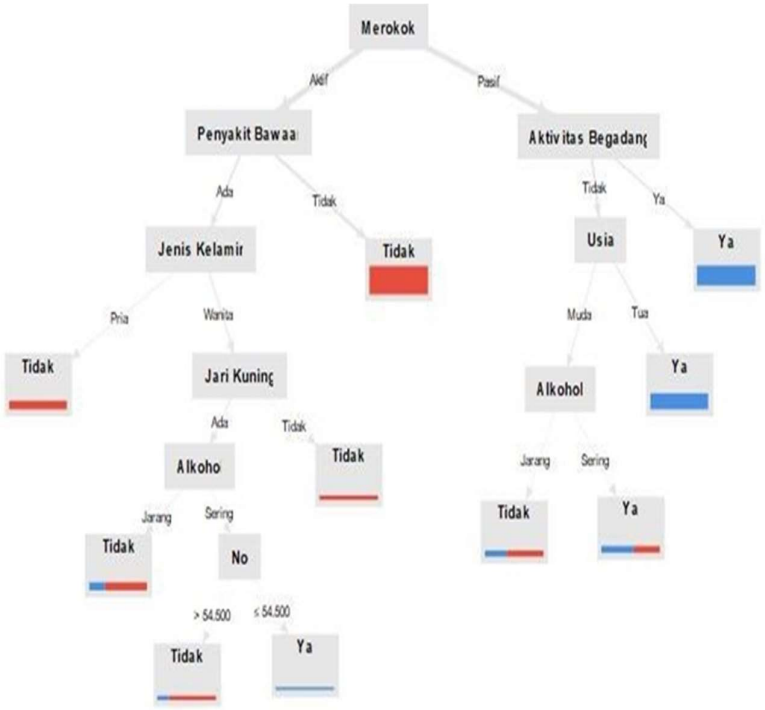
Saat menggunakan RapidMiner dengan algoritma C4.5, interpretasi model pohon keputusan dan aturan-aturan ini penting untuk memahami faktor-faktor yang mempengaruhi prediksi kanker paru-paru. Informasi ini bisa memberikan wawasan yang berharga kepada para profesional kesehatan atau peneliti dalam mengidentifikasi 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.

Pada gambar di atas merupakan rumus untuk perhitungan menggunakan rapidminer

Pohon Keputusan.

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

ISSN: 2774-4345, DOI: 10.31284/j.kernel.2024.v5i2.7315



Gambar 4. Pohon keputusan

Hasil dari pohon keputusan ini yaitu :

1. Perokok aktif

- Jika merokok aktif dan memepunyai penyakit bawaan maka dia ya, dia dapat menghidap penyakit paru.
- Jika merokok tetapi tidak mempunyai penyakit bawaan maka dia tidak terkena penyakit paru
- Jika dia merokok aktif dan mempunyai penyakit bawaan dan berjenis kelamin laki laki maka dia tidak terkena penyakit paru
- Jika dia mereokok aktif dan mempunyaipenyakit bawaan dan berjenis kelamin perempuan maka dia terkena penyakit paru
- Jika dia meroko aktif dan mempunyai penyakit bawaan dan berjenis kelamin perempuan tetapi tidak berjari kuning maka dia tidak terkena penyakit paru
- Jika dia meroko aktif dan mempunyai penyakit bawaan dan berjenis kelamin perempuan mempunyai jari kuning maka dia terkena penyakit paru
- 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
- 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

- Jika dia merokok pasif tidak bergadang maka dia tidak terkena penyakit paru
- Jika dia merokok pasif dan bergadang maka dia dapat terkena penyakit paru
- Jika dai merokok pasif dan bergadang dan berusia muda maka dia terkena penyakit paru
- Jika dia meroko pasif dan bergadang dan berusia tua maka dia tidak terkena penyait paru

ISSN: 2774-4345, DOI: 10.31284/j.kernel.2024.v5i2.7315

- Jika dia merokok pasif dan bergadang, berusia tua dan mengkonsumsi alkohol maka dia terkena penyakit paru
- Jika dia merokok pasif dan bergadang, berusia tua dan jarang mengkonsumsi alkohol maka dia tidak terkena penyakit paru

Berdasarkan gambar 4 , dapat dijelaskan bahwa dari 9 artibut penilaian terhadap prediksi kanker paru-paru, terdapat 1artibut yang memiliki pengaruh paling signifikan terhadapa prediksi kanker paru- paru yaitu artibut 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.

KESIMPULAN

Dalam penelitian ini, data diperoleh dari platform Kaggle menggunakan teknik pengambilan data sekunder, yang terdiri dari 150 record dan dibagi menjadi 80% untuk training serta 20% untuk testing. Proses analisis menggunakan algoritma C4.5 dalam RapidMiner untuk membangun model prediksi kanker paru-paru. Dari hasil pengolahan data, sembilan atribut terpilih memiliki pengaruh terhadap prediksi kanker paru-paru, dengan atribut "Merokok" menjadi faktor paling signifikan. Pohon keputusan yang dihasilkan menunjukkan bahwa individu dengan kebiasaan merokok aktif dan memiliki penyakit bawaan memiliki risiko lebih tinggi terkena kanker paru-paru, terutama jika berjenis kelamin perempuan dan memiliki jari kuning. Sementara itu, perokok pasif juga memiliki risiko, terutama jika memiliki kebiasaan begadang dan mengonsumsi alkohol. Penelitian ini membuktikan bahwa algoritma C4.5 yang diterapkan dalam RapidMiner dapat digunakan untuk mengklasifikasikan risiko kanker paru-paru berdasarkan faktor-faktor tertentu. Model yang dihasilkan dapat membantu dalam mendukung keputusan medis, terutama dalam upaya deteksi dini dan pencegahan penyakit.

DAFTAR PUSTAKA

[1] R. Rofiani, L. Oktaviani, D. Vernanda, dan T. Hendriawan, “Penerapan Metode Klasifikasi Decision Tree dalam Prediksi Kanker Paru-Paru Menggunakan Algoritma C4.5,” J. Tekno Kompak, vol. 18, no. 1, hal. 126, 2024, doi: 10.33365/jtk.v18i1.3525.

[2] R. Pambudi, A. R. Harahap, F. D. Saputra, dan M. Jusub, “Klasifikasi Penyakit Paru-paru Menggunakan Metode Decision Tree,” vol. 3, no. 9, hal. 2397–2402, 2024.

[3] M. Y. Haffandi, E. Haerani, F. Syafria, dan L. Oktavia, “Klasifikasi Penyakit Paru-Paru Dengan Menggunakan Metode Naïve Bayes Classifier,” J. Tek. Inf. dan Komput., vol. 5, no. 2, hal. 176, 2022, doi: 10.37600/tekinkom.v5i2.649.

[4] E. Muningsih, “Kombinasi Metode K-Means Dan Decision Tree Dengan Perbandingan Kriteria Dan Split Data,” J. Teknoinfo, vol. 16, no. 1, hal. 113, 2022, doi: 10.33365/jti.v16i1.1561.

[5] R. Robianto, S. H. Sitorus, dan U. Ristian, “Penerapan Metode Decision Tree Untuk Mengklasifikasikan Mutu Buah Jeruk Berdasarkan Fitur Warna Dan Ukuran,” Coding J. Komput. dan Apl., vol. 9, no. 01, hal. 76, 2021, doi: 10.26418/coding.v9i01.45907.

[6] R. T. Handayanto dan H. Herlawati, “Prediksi Kelas Jamak dengan Deep Learning Berbasis Graphics Processing Units,” J. Kaji. Ilm., vol. 20, no. 1, hal. 67–76, 2020, doi: 10.31599/jki.v20i1.71.

[7] D. Anandra, L. Latipah, dan A. Ambarwati, “Implementasi Metode Preference Selection Index Pada Pemilihan Penerima Bantuan SEMBAKO,” JOINTECS (Journal Inf. Technol. Comput. Sci., vol. 7, no. 1, hal. 9, 2022, doi: 10.31328/jointecs.v7i1.2660.