



# Implementasi Metode Decision Tree Dalam Prediksi Kanker Paru Paru Dengan Rapidminer

Anzori<sup>1\*</sup>, Zaehol Fatah<sup>2</sup>

<sup>1</sup>Teknologi Informasi, Universitas Ibrahimy

<sup>2</sup>Sistem Informasi, Universitas Ibrahimy

<sup>1\*</sup>anzoribws@gmail.com, <sup>2</sup>zaeholfatah@gmail.com

## Article History:

Received Okt 12<sup>th</sup>, 2024

Revised Okt 20<sup>th</sup>, 2024

Accepted Nov 15<sup>th</sup>, 2024

## Abstrak

kanker merupakan salah satu faktor utama penyebab kematian. Kanker juga memiliki posisi kedua dengan angka kematian yang sangat tinggi. Dalam penelitian ini memprediksi kanker paru-paru dengan menggunakan data mining dengan metode decision tree. Metode decision tree memprediksi akurasi dengan tingkat akurasi 94.66% dari data yang ada menjadi potensi besar dalam prediksi penyakit kanker paru-paru. Oleh karena itu, data mining dengan metode decision tree sebagai solusi ke efektifan dalam memprediksi terkenanya kanker paru paru dan pencegahannya. Meode ini juga menjadi upaya dalam penyelamatan nyawa dan perbaikan kualitas hidup bagi individu yang berisiko terkena penyakit kanker paru-paru.

**Kata Kunci :** Data Mining, Rapiitmainer, Decision Tree, Prediksi Kanker

## Abstract

*Cancer is one of the main factors causing death. Cancer also has the second position with a very high death rate. In this study, predicting lung cancer uses data mining with the decision tree method. The decision tree method predicts accuracy with a high level of accuracy because active smokers with an accuracy level of 94.66% from existing data have great potential in predicting lung cancer. Therefore, data mining using the decision tree method is a solution to be effective in predicting disease. Lung cancer and its prevention. This method is also an effort to save lives and improve the quality of life for individuals at risk of lung cancer.*

**Keyword :** Data Mining, Rapiitmainer, Decision Tree, Cancer Prediction

## PENDAHULUAN

Penyakit Jantung masuk ke dalam kelompok penyakit kardiovaskuler dan salah satu penyebab utama dari kehilangan nyawa secara global.[1] Kanker 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.[2] Menurut WHO (2022) pada tahun 2020, terdapat 2.2 juta penderita kanker paru-paru di seluruh dunia dan 1.8 juta mengalami kematian. Di Amerika diperkirakan terdapat 236.740 kasus kanker paru-paru dan 130.180 kasus mengalami kematian.[3] Kanker paru-paru adalah gangguan pada paru-paru karena perubahan sel epitel saluran pernapasan yang menyebabkan pembelahan dan pertumbuhan sel tak terkendali. Kanker paru-paru disebabkan oleh beberapa faktor seperti paparan radiasi, perokok, keturunan, jenis kelamin, polusi udara dan pola hidup tidak sehat. Kanker paru-paru dapat dideteksi ketika kanker telah memasuki tahap stadium lanjut. Banyaknya data diagnosis kanker paru-paru saat ini, dapat digunakan untuk memprediksi kanker paru-paru berdasarkan pola pada data.[4] Kanker paru-paru juga merupakan masalah kesehatan masyarakat yang serius, yang menyebabkan sejumlah besar kematian di seluruh dunia. Estimasi GLOBOCAN 2020 tentang kejadian dan kematian akibat kanker yang dibuat oleh Badan Internasional untuk Penelitian Kanker (IARC) menunjukkan bahwa kanker paru-paru tetap menjadi penyebab utama kematian akibat kanker, dengan perkiraan 1,8 juta kematian (18%) pada tahun 2020.

Merokok tembakau (termasuk rokok, cerutu, dan pipa) merupakan faktor risiko utama kanker paru-paru, tetapi juga dapat memengaruhi orang yang bukan perokok. Faktor risiko lainnya termasuk paparan asap rokok orang lain, bahaya pekerjaan (seperti asbes, radon, dan bahan kimia tertentu), polusi udara, sindrom kanker turunan, dan penyakit paru kronis sebelumnya. Metode diagnostik untuk kanker paru-paru meliputi pemeriksaan fisik, pencitraan (seperti rontgen dada, pemindaian tomografi terkomputasi, dan pencitraan resonansi magnetik), pemeriksaan bagian dalam paru-paru



menggunakan bronkoskopi, pengambilan sampel jaringan (biopsi) untuk pemeriksaan histopatologi dan definisi subtype spesifik (NSCLC versus SCLC), dan pengujian molekuler untuk mengidentifikasi mutasi genetik atau biomarker spesifik untuk memandu pilihan pengobatan terbaik.

Oleh sebab itu maka perlu kiranya suatu model prediksi penyakit paru-paru secara dini merupakan sebaik 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. Dengan adanya metode decision tree 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.[7]

Dalam penelitian ini, kami menerapkan teknik data mining yang menggunakan metode Decision Tree. Fokus utama kami adalah untuk mencapai tingkat akurasi yang tepat dalam memprediksi penyakit paru-paru, kami berharap dapat menyortir cara kami membangun model prediksi yang berpotensi memberikan hasil yang lebih andal dan relevan terkait deteksi dini penyakit paru-paru.[8]

## TINJAUAN PUSTAKA

### 2.1 Data Mining

Data Mining adalah proses penggalian informasi dan pola yang bermanfaat dari suatu data yang sangat besar. Proses data mining terdiri dari pengumpulan data, ekstraksi data, analisa data, dan statistik data. Ia juga umum dikenal sebagai knowledge discovery, knowledge extraction, data/pattern analysis, information harvesting, dan lainnya.

Menurut Muflikhah (2018), data mining dapat didefinisikan sebagai penguraian kompleks dari sekumpulan data menjadi informasi yang memiliki potensi secara implisit (tidak nyata/jelas) yang sebelumnya belum diketahui. Ia juga dapat didefinisikan sebagai penggalian dan analisis dengan menggunakan peralatan otomatis atau semi otomatis, dari sebagian besar data yang memiliki tujuan yaitu menemukan pola yang memiliki arti atau maksud.[9]

### 2.2 Decision Tree

Decision tree merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (tree) di mana setiap node merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas. Node yang paling atas dari decision tree disebut sebagai root. Decision tree merupakan metode klasifikasi yang paling populer digunakan. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun mudah untuk dipahami.

Simpul dalam menyatakan pengujian terhadap suatu atribut (digambarkan dengan kotak), cabang menyatakan hasil dari suatu pengujian (digambarkan dengan panah yang memiliki label dan arah), sementara daun menyatakan kelas yang diprediksi (digambarkan dengan lingkaran).[10]

### 2.3 RapiMainer

RapidMiner merupakan sebuah software atau perangkat lunak yang digunakan untuk proses pengolahan data. RapidMiner merupakan sebuah perangkat lunak yang bersifat terbuka (open source), yang digunakan untuk melakukan proses analisis data mining, text mining, dan analisis prediksi. Dalam proses ini, RapidMiner menggunakan berbagai teknik deskriptif dan prediksi untuk dapat membuat sebuah keputusan yang paling baik.[11]

RapidMiner menyediakan GUI (Graphic User Interface) untuk merancang sebuah pipeline analitis, yang akan menghasilkan file XML (Extensible Markup Language) yang mendefinisikan proses analitis keinginan pengguna untuk



diterapkan ke dalam bentuk data. File ini kemudian dibaca oleh RapidMiner untuk menjalankan analisis secara otomatis.[11]

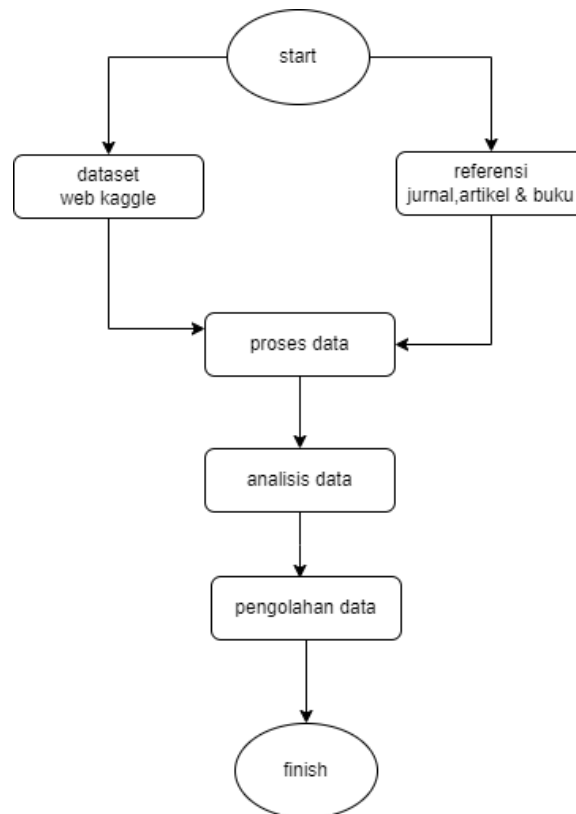
Berkat adanya kecanggihan teknologi algoritma komputasi dan analisis data berbasis komputer, data mining dapat diolah menggunakan software Rapidminer.[12] dengan adanya Rapidminer menjadi lebih mudah dalam penggunaan atau implementasi data mining yang memang dirancang khusus tanpa coding namun sudah dilengkapi dengan fitur-fitur yang sangat dibutuhkan dan digunakan dalam proses implementasi data mining.

#### 2.4 Kanker Paru Paru

Kanker paru-paru adalah kondisi kesehatan yang bermula dalam jaringan paru-paru, organ pernapasan utama di dalam tubuh. Lung cancer atau kanker paru-paru merupakan jenis kanker yang tumbuh pada organ paru-paru, yaitu di mana terdapat sel-sel kanker yang berkembang secara tidak terkendali di dalamnya. Kanker ini biasanya berkembang secara perlahan dan mungkin tidak menunjukkan gejala pada tahap awal.[13]

Kanker paru-paru terjadi saat sel di paru-paru tumbuh tidak teratur dan membentuk tumor. Tumor bisa non-kanker atau kanker. Tumor ganas adalah kumpulan dari sejumlah besar sel kanker yang mempunyai kemampuan untuk menyebar ke bagian tubuh yang lain. Tingkat kematian dari kanker paru-paru tergantung dari jenis kankernya.[14]

### METODE



Gambar 1. Gambar Tahapan Metode Waterfall

#### 3.1 Pengumpulan Data

Dataset didapatkan dengan pengelolaan data yang didapat dari web Kaggle yang dalam data tersebut memiliki beberapa macam atribut yang berbeda, yaitu;

1. Usia: Informasi usia ada dalam dataset.
2. Jenis kelamin: Jenis kelamin subjek, seperti pria atau wanita.
3. Merokok: Status merokok (Pasif/Aktif).
4. Rumah tangga: Status rumah tangga (Ya/Tidak).
5. Aktivitas bedagang: informasi kebiasaan dalam berdagang.
6. Aktivitas olahraga: informasi kebiasaan dalam berolahraga
7. Asuransi: Status asuransi (Ya/Tidak)
8. Penyakit bawaan: Status penyakit (Ada/Tidak)
9. Hasil: hasil dari tes terkait terkena penyakit kanker paru paru.



Pada pengumpulan data ini, dataset yang digunakan berasal dari website Kaggle yang terhitung dari beberapa kumpulan sekian data. Sedangkan dengan atribut menyesuaikan dengan dataset yang sudah diambil dari sumber terpilih dan terpercaya akan kejelasan data tersebut.

### 3.2 Proses Pengumpulan Data

Kegiatan pengumpulan data merupakan langkah awal yang krusial dalam penelitian kesehatan, di mana informasi tentang subjek dalam dataset dikumpulkan dari sumber-sumber terpercaya seperti survei, catatan medis, dan sumber data kesehatan lainnya. Ketelitian data menjadi fokus utama dalam proses ini, dengan memastikan konsistensi, validitas, dan integritas data yang dikumpulkan; setiap atribut harus didokumentasikan dengan jelas dan memiliki spesifikasi terperinci untuk menjaga kualitas data. Selain itu, aspek privasi dan etika sangat penting, di mana peneliti harus memastikan bahwa pengumpulan data dilakukan sesuai dengan peraturan privasi yang berlaku serta etika penelitian, termasuk menjaga anonimitas subjek dan mendapatkan izin sebelum mengumpulkan data sensitif. Setelah pengumpulan, tahap pengolahan awal dilakukan untuk membersihkan data dari duplikasi, kesalahan, atau nilai yang hilang agar siap untuk analisis lebih lanjut. Dengan demikian, setiap langkah dalam proses ini berkontribusi pada keberhasilan penelitian kesehatan secara keseluruhan.

Pada proses pengumpulan data ini menjadi Langkah penting dalam penelitian ini, dimana dataset yang diperoleh merupakan data yang valid, lengkap dan sesuai dengan kebutuhan yang akan dijadikan bahan analisis dan pengujian ini.

Tabel 1. Dataset Prediksi Penyakit Kanker Paru-Paru

| No | Usia | Jenis Kelamin | Merokok | Bekerja | Rumah Tangga | Aktivitas Bedagang | Aktivitas Olahraga | Asuransi | Penyakit Bawaan | Hasil |
|----|------|---------------|---------|---------|--------------|--------------------|--------------------|----------|-----------------|-------|
| 1  | Tua  | Pria          | Pasif   | Tidak   | Ya           | Ya                 | Sering             | Ada      | Tidak           | Ya    |
| 2  | Tua  | Pria          | Aktif   | Tidak   | Ya           | Ya                 | Jarang             | Ada      | Ada             | Tidak |
| 3  | Muda | Pria          | Aktif   | Tidak   | Ya           | Ya                 | Jarang             | Ada      | Tidak           | Tidak |
| 4  | Tua  | Pria          | Aktif   | Ya      | Tidak        | Tidak              | Jarang             | Ada      | Ada             | Tidak |
| 5  | Muda | Wanita        | Pasif   | Ya      | Tidak        | Tidak              | Sering             | Tidak    | Ada             | Ya    |
| 6  | Muda | Wanita        | Pasif   | Ya      | Tidak        | Tidak              | Sering             | Tidak    | Ada             | Tidak |
| 7  | Tua  | Wanita        | Pasif   | Tidak   | Ya           | Tidak              | Sering             | Tidak    | Tidak           | Ya    |
| 8  | Muda | Pria          | Aktif   | Tidak   | Ya           | Ya                 | Sering             | Tidak    | Tidak           | Tidak |
| 9  | Tua  | Wanita        | Aktif   | Ya      | Ya           | Ya                 | Jarang             | Ada      | Ada             | Ya    |
| 10 | Muda | Wanita        | Pasif   | Ya      | Tidak        | Ya                 | Jarang             | Ada      | Ada             | Ya    |
| 11 | Tua  | Wanita        | Pasif   | Ya      | Ya           | Tidak              | Sering             | Ada      | Ada             | Ya    |
| 12 | Tua  | Wanita        | Aktif   | Tidak   | Ya           | Tidak              | Jarang             | Ada      | Tidak           | Tidak |
| 13 | Muda | Pria          | Aktif   | Tidak   | Ya           | Ya                 | Jarang             | Ada      | Tidak           | Tidak |
| 14 | Tua  | Wanita        | Aktif   | Ya      | Tidak        | Ya                 | Jarang             | Ada      | Ada             | Tidak |
| 15 | Muda | Wanita        | Pasif   | Ya      | Tidak        | Ya                 | Sering             | Tidak    | Ada             | Ya    |
| 16 | Muda | Wanita        | Pasif   | Ya      | Tidak        | Ya                 | Jarang             | Ada      | Ada             | Ya    |
| 17 | Tua  | Wanita        | Pasif   | Ya      | Ya           | Tidak              | Sering             | Ada      | Ada             | Ya    |
| 18 | Tua  | Wanita        | Aktif   | Tidak   | Ya           | Tidak              | Jarang             | Ada      | Tidak           | Tidak |
| 19 | Muda | Pria          | Aktif   | Tidak   | Ya           | Ya                 | Jarang             | Ada      | Tidak           | Tidak |
| 20 | Tua  | Wanita        | Aktif   | Ya      | Tidak        | Ya                 | Jarang             | Ada      | Ada             | Tidak |
| 21 | Muda | Wanita        | Pasif   | Ya      | Tidak        | Ya                 | Sering             | Tidak    | Ada             | Ya    |
| 22 | Tua  | Pria          | Pasif   | Tidak   | Ya           | Ya                 | Sering             | Ada      | Tidak           | Ya    |
| 23 | Tua  | Pria          | Aktif   | Tidak   | Ya           | Ya                 | Jarang             | Ada      | Ada             | Tidak |
| 24 | Muda | Pria          | Aktif   | Tidak   | Ya           | Ya                 | Jarang             | Ada      | Tidak           | Tidak |
| 25 | Tua  | Pria          | Aktif   | Ya      | Tidak        | Tidak              | Jarang             | Ada      | Ada             | Tidak |

### 3.3 pengolahan data

Tahap pengolahan data dalam metode penelitian adalah proses sistematis yang mencakup serangkaian langkah penting untuk mempersiapkan data sebelum dilakukan analisis. Dataset diolah dengan tidak lanjut dari analisis data yang penerapannya dengan metode decision tree terhadap prediksi kanker paru-paru yang tolsnya menggunkan Rapit Mainer. Dalam proses ini ada beberapa tahapan yang akan dijelaskan sebagai berikut:

- a. Pemilihan data: pada tahap pertama ini pemilihan data yang relevan dengan data berasal dari website Kaggle.



- b. Pengumpulan data: setelah melakukan pemilihan data maka Langkah selanjutnya mengumpulkan informasi dari sumber yang sudah terpilih. Dimana pada proses ini akan mengambil data secara mandiri.
- c. Pembersihan data: pada proses ini data akan mengatasi data yang tidak lengkap, atau tidak akurat. Proses ini melibatkan indentifikasi nilai yang hilang atau tidak konsistennya dalam yang diteliti yaitu untuk memastikan keakuratan data sebelum dilakukannya analisis.
- d. Traformasi data: dalam tranformasi data terkadang sering memiliki format berbeda-beda dimana pada tahapan ini melibatkan tentang normalisasi, pengkodean, atau mengubah format kedalam bentuk konsisten yaitu dalam rangka agar analisis lebih efektif.
- e. Integrasi data: data yang didapatkan dari beberapa sumber, maka akan dilakukan proses integrasi. Hal ini memungkinkan penggabungan data dari bebagai sumber untuk dijadikan satu set data yang lebih lengkap dan terintegrasi.
- f. Reduksi dimensi: melakukan penghilangan atribut yang tidak relavan yaitu untuk mengurangi kompleksitas sehingga mempermudah dalam analisis.
- g. Validasi data: pada Langkah terakhir ini memverifikasi akurat data, konsisten data dan kejelasan data yang mana dalam prosesnya untuk memastikan bahwa data tersebut sudah siap untuk digunakan untuk analisis tanpa memperdulikan lagi adanya kesalahan.

Pada tahapan pengolahan data ini menjadi fondasi penting sebelum peneliti melakukan analisis lebih lanjut. Hal ini sangat berguna dalam memastikan bahwa data siap untuk digunakan, yang berguna dalam mempengaruhi kualitas kesimpulan dari hasil penelitian ini.

### 3.4 Analisis Data

#### a. Analisis Data

Tahap analisis data merupakan langkah awal yang penting sebelum proses pengolahan data, dengan tujuan utama untuk memahami dataset secara menyeluruh agar data siap untuk diolah atau diuji menggunakan metode decision tree. Dalam tahap ini, analisis dilakukan untuk mendeteksi dan menghapus data duplikat, jika ada, karena duplikasi dalam dataset dapat mengganggu analisis dan interpretasi hasil. Selain itu, validasi konsistensi data juga menjadi fokus, di mana data diperiksa untuk memastikan konsistensi baik dalam format maupun nilai; hal ini mencakup pemeriksaan setiap entri agar sesuai dengan aturan dan tipe data yang diharapkan. Dengan demikian, tahap analisis ini sangat penting untuk memastikan kualitas dan keandalan data sebelum melanjutkan ke proses pengolahan lebih lanjut.

#### b. Proses Analisis Data

Proses analisis data melibatkan penggunaan berbagai alat dan teknik, seperti software statistik, pemrograman untuk pembersihan data, dan visualisasi data untuk meningkatkan pemahaman terhadap dataset. Beberapa software yang umum digunakan dalam analisis ini termasuk SPSS, Microsoft Excel, dan bahasa pemrograman R, masing-masing menawarkan fitur yang berbeda untuk memudahkan pengguna dalam mengolah dan menganalisis data. Setelah tahap analisis, penting untuk melakukan validasi dan verifikasi data guna memastikan bahwa semua masalah yang terdeteksi sebelumnya telah diperbaiki dan dataset siap digunakan pada tahap berikutnya. Proses ini memastikan bahwa hasil analisis tidak hanya akurat tetapi juga dapat diandalkan untuk pengambilan keputusan yang lebih baik.

Pada proses analisis ini memastikan dataset yang akan digunakan untuk melewati tahap pengujian yang sangat diperlukan meminimal adanya kesalahan data dan memastikan hasil analisis yang dapat diandalkan.

## HASIL DAN PEMBAHASAN

Dalam penelitian ini metode yang digunakan decision tree dengan proses pengolahan data melalui beberapa langkah seleksi dan traformasi data. Pada bagian ini menggunakan panduan untuk memprediksi terkena kanker paru-paru.

Dalam proses pengumpulan data bahwa Teknik yang digunakan yaitu secara sekunder yang mana salah satu platform menyediakan data yaitu website Kaggle. Dataset yang digunakan untuk penelitian terdiri beberapa record yang kemudian dibagi menjadi 80% untuk data training dan 20% untuk data testing. Berikut ini adalah tahapan proses decion tree menggunakan rapidminer dengan menggunakan data training terkait Ya dan Tidak hasil dari prediksi terkenanya kanker paru.



Tabel 2. Daftar Atribut Keseluruhan

| Atribut            | Tipe Data | Keterangan    |
|--------------------|-----------|---------------|
| Usia               | kategori  | Tua/Muda      |
| Jenis kelamin      | kategori  | Pria/Wanita   |
| Merokok            | kategori  | Pasif/Aktif   |
| Bekerja            | kategori  | Ya/Tidak      |
| Rumah tangga       | kategori  | Ya/Tidak      |
| Aktivitas bedagang | kategori  | Ya/Tidak      |
| Aktivitas olahraga | kategori  | Sering/Jarang |
| Asuransi           | kategori  | Ada/Tidak     |
| Penyakit Bawaan    | kategori  | Ada/Tidak     |
| Hasil              | label     | Ya/Tidak      |

#### 4.1 Data Selection

Pada tahap ini ada beberapa atribut yang kurang relevan untuk diprediksi terkena kanker paru-paru. Dalam proses ini memiliki tujuan untuk memilih atribut yang paling tepat untuk digunakan memprediksi penyakit kanker paru-paru. Dibawah ini berisi daftar atribut yang digunakan dan akan diproses oleh software RapiMainer yang didalamnya sudah lengkap dengan beberapa kebutuhan dalam Teknik data mining.

Tabel 3. Daftar Atribut Terpilih

| No | Atribut Terpilih    | Keterangan    |
|----|---------------------|---------------|
| 1  | Usia                | Atribut Fitur |
| 2  | Merokok             | Atribut Fitur |
| 3  | Bekerja             | Atribut Fitur |
| 4  | Rumah Tangga        | Atribut Fitur |
| 5  | Aktivitas Berdagang | Atribut Fitur |
| 6  | Asuransi            | Atribut Fitur |
| 7  | Hasil               | Label         |

Atribut yang terpilih adalah usia, merokok, bekerja, rumah tangga, aktivitas berdagang, asuransi, dan hasil. Pemilihan atribut dipilih berdasarkan dengan hasil korelasi mereka yang terkena penyakit kanker paru-paru. Atribut ini akan di proses dengan menggunakan decision tree dengan software RapiMainer. Pada proses ini akan dibuat pohon keputusan yaitu untuk mengukur akurat prediksi terkena penyakit kanker paru-paru. Tujuan dari proses ini memilih atribut yang paling tepat untuk memprediksi penyakit kanker paru-paru.maka pembahasan dibawah ini menampilkan daftar data atribut yang akan diproses oleh software RapiMainer.

#### 4.2 Data Transformation

Langkah traformasi ini mengikuti perubahan format data,agar data yang didapat sesuai dengan yang diproses lebih lanjtu kedalam proses Teknik data mining. Biasanya tranformasi dataset dilakukan dengan rapitmainer yaitu mengubah tipe data dari polynomial menjadi binomial kepada atribut yang sudah digunakan. Berikut bisa terlihat pada gambar dibawah ini.



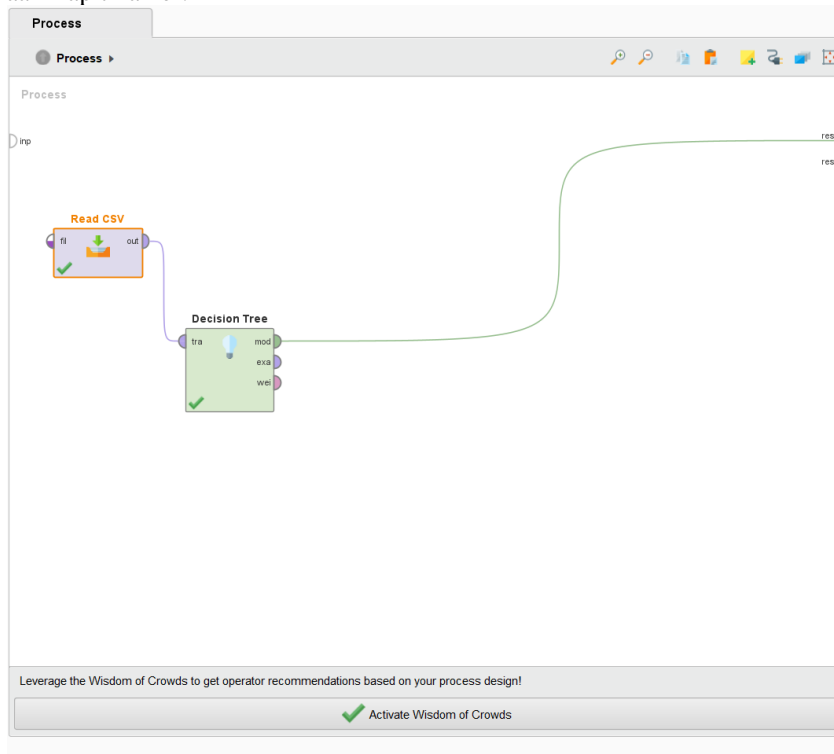


| kok_ominal | Bekerja_polynomial | Rumah_Tan..._polynomial | Aktivitas_Be..._polynomial | Aktivitas_OL..._polynomial | Asuransi_polynomial | Penyakit_Ba..._polynomial | Hasil_binominal_label |
|------------|--------------------|-------------------------|----------------------------|----------------------------|---------------------|---------------------------|-----------------------|
| 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                    |
| 18         | Tidak              | Ya                      | Tidak                      | Jarang                     | Ada                 | Tidak                     | Tidak                 |

Gambar 2. Data Set Yang Didapat Dari Kaggle

#### 4.3 Implementasi RapiMainer

Pada proses implementasi RapiMainer ini akan mengelola dataset yang nanti akan memunculkan hasil akurasi data dengan membangun pohon keputusan(decision tree) dari data yang sudah ada. Pada bagian pohon keputusan ini data yang diambil dari atribut yang sudah ada. Selanjutnya, software RapiMainer akan menjalankan sistem yang ada untuk memprediksi dari data yang ada tentang prediksi terkena penyakit kanker paru-paru.berikut merupakan gambaran dari implementasi penggunaan RapiMainer.

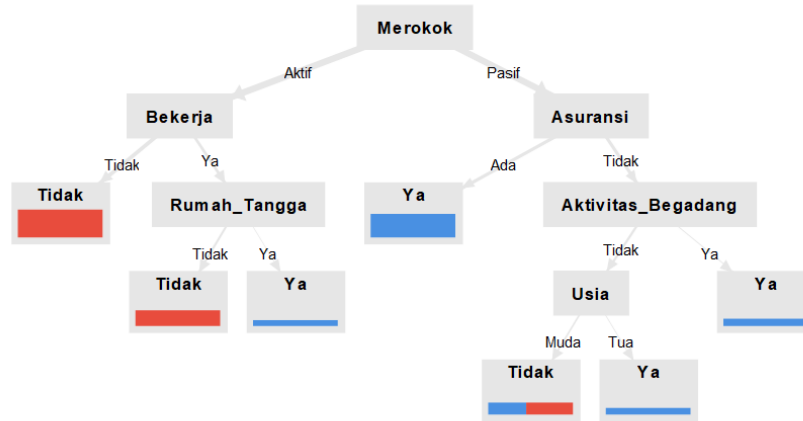


Gambar 3. Implementasi RapiMainer

Gambar diatas menunjukkan rumus untuk menghitung penggunaan software rapiMainer yang didalamnya sudah dilengkapi dengan fitur-fitur yang dibutuhkan oleh data mining.

a. Pohon Keputusan

Dari hasil pengujian pada RapidMiner, diperoleh pohon keputusan dapat dilihat pada Gambar 5 dibawah ini.



Gambar 4. Pohon Keputusan

b. Deskripsi Pohon Keputusan

Berdasarkan pohon keputusan di atas, dapat disimpulkan bahwa terdapat beberapa kondisi atau aturan. Dapat dilihat deskripsi pohon keputusan Rapidminer pada gambar 5 di bawah ini.



Gambar 5. Deskripsi Pohon Keputusan

Berdasarkan hasil gambar diatas dapat dijelaskan bahwasanya 7 atribut memiliki penilaian tentang prediksi terkena kanker paru-paru, dalam hal ini ada 1 atribut yang sangat berpengaruh terhadap prediksi terkena kanker paru-paru yaitu atribut merokok. Hal ini Sudah dilakukan pengujian terhadap kevalidan data yang ada, yaitu menggunakan metode decision tree dengan diterapkan dengan software RapiMainer untuk prediksi terkenanya kanker paru-paru.

### KESIMPULAN

Dalam uji coba yang di lakukan dengan metode decision tree menggunakan 10 artibut maka menghasilkan akurasi 94.66% dari data yang ada. Terbukti bahwa metode ini memberikan keakuratan dataset menjadi data yang lebih detail setelah dilakukannya seleksi data atribut dengan dijadikan atribut terpilih. Keterbatasan data yang masih tergantung kepada data yang ada dan butuh kepada validasi data yang lebih luas. Data mining dengan metode decision tree sangat diharapkan bisa menjadi landasan pengembangan lebih baik yaitu sebagai upaya dalam pencegahan penyakit kanker paru-paru.





### UCAPAN TERIMA KASIH

Pembuatan” Implementasi Metode Decision Tree Dalam Prediksi Kanker Paru Paru Dengan RapiMainer” ini tidak lepas dari bimbingan dan bantuan dari berbagai pihak. Ribuan terima kasih saya ucapkan kepada:

1. dosen pembimbing yang telah meluangkan waktunya untuk memberikan arahan, bimbingan dan serta wawasan yang menjadikan salah satu kunci keberhasilan penelitian ini.
2. Pihak kampus yang telah memberikan fasilitas yang sangat memadai sehingga penelitian ini bisa selesai tepat waktu.

### DAFTAR PUSTAKA

- A. Yogiarto, A. Homaidi, and Z. Fatah, “Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung,” *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 1720–1728, 2024, doi: 10.33379/gtech.v8i3.4495.
- H. Widya, N. Surya Putra, V. Atina, and J. Maulindar, “Penerapan Algoritme Decision Tree Pada Klasifikasi Penyakit Kanker Paru-Paru,” *J. Ilm. Tek. Inform. dan Sist. Inf.*, 2023, [Online]. Available: <https://www.kaggle.com/datasets/mysarahmadbhat/lung-cancer>,
- F. Faurika, A. Naseh Khudori, and M. S. Haris, “Application of Cross-Validation Techniques to Handle Overfitting in a Case Study of Decision Tree Implementation for Lung Cancer Prediction,” *ROUTERS J. Sist. dan Teknol. Inf.*, vol. 2, no. 2, pp. 111–120, 2024, doi: 10.25181/rt.v2i2.3631.
- A. Naseh Khudori and M. Syauiq Haris, “Implementasi Decision tree Untuk Prediksi Kanker Paru-Paru,” *J. Ris. Sist. Inf. Dan Tek. Inform. (JURASIK)*, vol. 9, no. 1, pp. 94–106, 2024, [Online]. Available: <https://tunasbangsa.ac.id/ejurnal/index.php/jurasik>
- A. Rifa’i and Y. Prabowo, “Diagnosis Kanker Paru-Paru dengan Sistem Fuzzy,” *Krea-TIF J. Tek. Inform.*, vol. 10, no. 1, pp. 19–28, 2022, doi: 10.32832/kreatif.v10i1.6317.
- A. Desiani *et al.*, “Perbandingan Klasifikasi Penyakit Kanker Paru-Paru menggunakan Support Vector Machine dan K-Nearest Neighbor,” *J. Process.*, vol. 18, no. 1, pp. 54–62, 2023, doi: 10.33998/processor.2023.18.1.700.
- P. R. MELIALA, “Perbandingan Algoritma Machine Learning Untuk Survivabilitas Dan Biaya Pengobatan Pasien Kanker Paru-Paru Di Taiwan,” 2021, [Online]. Available: <https://dspace.uui.ac.id/handle/123456789/37617%0Ahttps://dspace.uui.ac.id/bitstream/handle/123456789/37617/19916038.pdf?sequence=1&isAllowed=y>
- M. A. Rahman Wahid, A. Nugroho, and A. Halim Anshor, “Prediksi Penyakit Kanker Paru-Paru Dengan Algoritma Regresi Linier,” *Bull. Inf. Technol.*, vol. 4, no. 1, pp. 63–74, 2023, doi: 10.47065/bit.v4i1.501.
- R. S. Wahono, *Data Mining Data mining*, vol. 2, no. January 2013. 2023. [Online]. Available: [https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book\\_part](https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part)
- M. R. Hidayatsyah, “Penerapan Metode Decision Tree Dalam Pemberian Pinjaman Kepada Debitur,” *PENERAPAN Metod. Decis. TREE DALAM PEMBERIAN PINJAMAN Kpd. DEBITUR DENGAN Algoritm. (Studi*, vol. 5, p. 22, 2013, [Online]. Available: <https://repository.uin-suska.ac.id/3213/>
- A. N. Yuliarina and H. Hendry, “Comparison of Prediction Analysis of Gofood Service Users Using the Knn & Naive Bayes Algorithm With Rapidminer Software,” *J. Tek. Inform.*, vol. 3, no. 4, pp. 847–856, 2022, doi: 10.20884/1.jutif.2022.3.4.294.
- Y. R. Sari, A. Sudewa, D. A. Lestari, and T. I. Jaya, “Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer,” *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, p. 192, 2020, doi: 10.24114/cess.v5i2.18519.
- A. siloa. Hopital, “kanker paru-paru: penyebab,gejala dan cara mengobatinya  
Url:<https://www.siloamhospitals.com/informasi-siloam/artikel/kanker-paru-paru>,” *artikel*.  
bummmrungrad internasional Hospital, “bummmrungrad internasional  
Hospitalurl:<https://www.bumrungrad.com/id/conditions/lung-cancer>”.