



## KLASIFIKASI PENYAKIT KANKER PARU MENGGUNAKAN PERBANDINGAN ALGORITMA *MACHINE LEARNING*

Oleh:

**Jatnika Fahmi Idris<sup>1</sup>**

**Rafid Ramadhani<sup>2</sup>**

**Muhammad Malik Mutoffar<sup>3</sup>**

Sekolah Tinggi Teknologi Bandung

Alamat: JL. Soekarno Hatta No.378, Kb. Lega, Kec. Bojongloa Kidul, Kota Bandung,  
Jawa Barat (40235)

Korespondensi penulis: [fahmijatnika014@gmail.com](mailto:fahmijatnika014@gmail.com)

**Abstract.** *The lungs are an important organ that plays a role in maintaining the respiratory and blood circulation systems in the human body. Smoking is one of the causes of disease and not only affects active smokers, but also people who are close to smokers or passive smokers, because smoking is one of the causes of lung cancer. Therefore, technology from the field of artificial intelligence can be used to predict, recognize patterns and classify whether or not a person is affected by lung cancer. Machine learning helps in obtaining accuracy values based on the algorithm used. This research compares 7 algorithms with existing accuracy values, the Random Forest, Tree, Neural Network, and Logistic Regression algorithms have accuracy values above 90%, while the SVM, Naïve Bayes, and KNN algorithms have accuracy values below 90%. It can be concluded that an algorithm that has a higher level of accuracy can be used in the classification of whether or not lung cancer is affected.*

**Keywords:** *Lungs, Smoking, Artificial Intelligence, Machine learning, Algorithms.*

**Abstrak.** Paru-paru adalah organ penting yang berperan dalam menjaga sistem pernafasan dan peredaran darah dalam tubuh manusia. Merokok merupakan salah satu penyebab timbulnya penyakit dan tidak hanya menimpa perokok aktif, namun juga orang-

# KLASIFIKASI PENYAKIT KANKER PARU MENGGUNAKAN PERBANDINGAN ALGORITMA MACHINE LEARNING

orang yang dekat dengan perokok atau perokok pasif, karena merokok adalah salah satu penyebab kanker paru-paru. Maka dari itu sebuah teknologi dari bidang kecerdasan buatan bisa dimanfaatkan untuk memprediksi, pengenalan pola serta mengklasifikasi terkena dan tidaknya seseorang oleh penyakit kanker paru-paru. *Machine learning* membantu dalam mendapatkan nilai akurasi yang di dapat berdasarkan algoritma yang digunakan. Penelitian ini membandingkan 7 algoritma dengan hasil nilai akurasi yang ada, untuk algoritma *Random Forest*, *Tree*, *Neural Network*, dan *Logistic Regression* memiliki nilai akurasi di atas 90% sementara untuk algoritma *SVM*, *Naïve Bayes*, dan *kNN* memiliki nilai akurasi di bawah 90%. Dapat di ambil kesimpulan bahwa algoritma yang memiliki tingkat akurasi yang lebih tinggi dapat digunakan dalam klasifikasi penentuan terkena dan tidak nya kanker paru-paru.

**Kata kunci:** Paru-Paru, Merokok, Kecerdasan Buatan, *Machine learning*, Algoritma.

## LATAR BELAKANG

Paru-paru merupakan organ penting yang berperan dalam sistem pernapasan dan peredaran darah tubuh manusia. Merokok adalah penyebab utama kanker paru-paru, namun perokok pasif kecil kemungkinannya berisiko terkena kanker paru-paru. Angka kejadian kanker paru-paru di Indonesia setiap harinya meningkat pesat hingga menduduki peringkat ke-8 di Asia Tenggara dan meningkat sebesar 10,85 persen dalam lima tahun terakhir (Vierisyah et al. n.d.). Kanker merupakan penyakit yang dapat menyebabkan kematian di berbagai negara. Kanker telah menjadi penyebab paling umum kematian manusia dalam beberapa tahun terakhir. Menurut WHO, kanker menyebabkan 9,6 juta kematian di seluruh dunia pada tahun 2018 (Naufal, Adiwijaya, and Astuti 2020). Menurut Globocan (2018), angka kejadian kanker paru secara global sebesar 11,6 dan angka kematian sebesar 18%, sedangkan angka kejadian kanker paru di Indonesia sebesar 8,6% atau sebanyak 30.023 kasus dengan angka kematian sebesar 12,6% atau 26095 kematian akibat kanker paru-paru (Putra, Utami, and P.Kurniawan 2022).

Merokok merupakan salah satu penyebab timbulnya penyakit dan tidak hanya menimpa perokok aktif, namun juga orang-orang yang dekat dengan perokok atau perokok pasif. Perokok pasif lebih mungkin mengalami dampak penyakit dibandingkan perokok aktif. Namun, jika 1% populasi manusia adalah perokok pasif, jumlah dokter spesialis paru yang ada saat ini tidak akan mampu mengatasinya. Ini adalah masalah yang

perlu diatasi. Pengguna dapat melakukan diagnosis awal terhadap gejala dan pengobatan yang dideritanya melalui sistem pakar. Dalam penelitian ini sistem pakar menggunakan metode faktor kepercayaan yang dapat memberikan kepastian terhadap fakta yang ada. Perhitungan dilakukan berdasarkan nilai pengalaman dari gejala penyakit. Sistem pakar yang dihasilkan disebut Diagperosif, dimana sistem mendiagnosis penyakit berdasarkan gejala yang dimasukkan oleh pengguna. Penyakit yang didiagnosis Diagperosif adalah asma, bronkitis, PPOK dan kanker paru-paru. Pasalnya, kebiasaan merokok lebih banyak dialami oleh laki-laki sehingga menyebabkan kanker pada jaringan paru-paru (Khultsum, Sarasati, and Taufik 2022).

Klasifikasi adalah jenis analisis data yang membantu orang memprediksi label kelas mana yang harus diklasifikasikan ke dalam sampel. Banyak teknik klasifikasi berbeda telah diusulkan di berbagai bidang seperti pembelajaran mesin, sistem pakar, dan *statistic* (Annisa 2019). Bila kanker paru-paru sudah diketahui maka segera dilakukan pencegahan dan pengobatan terhadap kanker paru-paru untuk mencegah efek samping bahkan kematian pada seseorang. Deteksi dini dan pengobatan dini sangat penting. Pembelajaran mesin adalah bagian dari bidang kecerdasan buatan yang berfokus pada penerapan algoritma dan metode tertentu untuk prediksi, pengenalan pola, dan klasifikasi (Angkasa and Pangaribuan 2022). Algoritma-algoritma di *Machine learning* yaitu *naural network*, *decision Tree*, *k-nearest neighbor*, *naïve bayes*, *Random Forest* dan lain sebagainya (Sowah et al. 2020).

## **KAJIAN TEORITIS**

Pada penelitian sebelumnya yang berjudul Klasifikasi penderita kanker Paru Paru Menggunakan Algoritma *Artificial Neural Network (ANN)* mendapat nilai akurasi sebesar 95.12% (Putra et al. 2022).

Pada penelitian lain yang berjudul Perbandingan Klasifikasi Penyakit Kanker Paru-paru menggunakan *Support Vector Machine* dan *K-Nearest Neighbor* mendapatkan akurasi 90% (Desiani et al. 2023).

Pada penelitian sebelumnya yang berjudul Klasifikasi Kanker Paru-Paru Menggunakan Metode *Naive Bayes* mendapat nilai akurasi sebesar 94,62%. (Wulandari 2022).

# KLASIFIKASI PENYAKIT KANKER PARU MENGGUNAKAN PERBANDINGAN ALGORITMA MACHINE LEARNING

Pada penelitian yang berjudul Penerapan Data *Mining* dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma *Random Forest* mendapat nilai akurasi sebesar 98,4%. (Sari, Romadloni, and Listyaningrum 2023).

Pada penelitian yang berjudul Implementasi Algoritma *Decision Tree* dan *Support Vector Machine* untuk Klasifikasi Penyakit Kanker Paru mendapatkan akurasi tingkat akurasi terbaik terdapat pada algoritma *Support Vector Machine (SVM)* menggunakan *Forward Selection splitting* data 80:20 dengan tingkat akurasi sebesar 62,3%. (Septhya et al. 2023).

Pada penelitian Penerapan Metode Mobile-Net Untuk Klasifikasi Citra Penyakit Kanker Paru-Paru. Mendapatkan hasil menerapkan metode *CNN* menggunakan arsitektur Mobile-Net yang telah dimodifikasi dilengkapi dengan tahapan preprocessing yaitu proses segmentasi menggunakan *K-Means* dengan hasil akurasi 96,70%. (Khultsum et al. 2022).

Pada penelitian yang berjudul Analisis Terjadinya Kanker Paru-Paru Pada Pasien Menggunakan *Decision Tree* : Penerapan Algoritma C4.5 Dan *RapidMiner* Untuk Menentukan Risiko Kanker Pada Gejala Pasien dengan hasil akurasi sebesar 93.80% (Mesin, Dan, and Jtmei 2023).

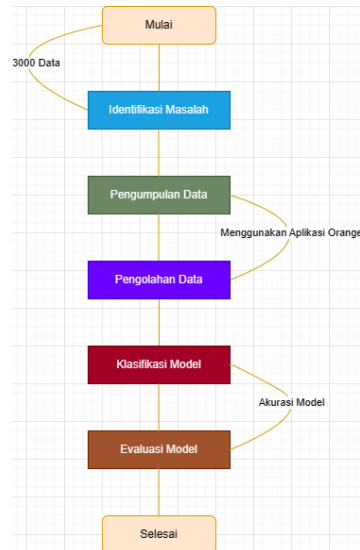
## METODE PENELITIAN

- A. Penambangan** data adalah proses menganalisis kumpulan data besar untuk menemukan pola yang berguna, informasi tersembunyi, atau informasi yang dapat digunakan untuk membuat keputusan yang lebih baik. Penambangan data menggunakan teknik statistik, matematika, dan kecerdasan buatan untuk memeriksa dan menganalisis kumpulan data besar dengan tujuan menemukan pola yang tidak langsung terlihat (Mesin et al. 2023).
- B. Orange** adalah teknologi pembelajaran mesin sumber terbuka atau perangkat lunak penambangan data. Orange dapat digunakan untuk analisis penelitian dan visualisasi data. Ini menyediakan platform untuk pemilihan eksperimen, pemodelan prediktif, dan sistem rekomendasi dan dapat digunakan dalam penelitian genomik, biomedis, bioinformatika, dan pendidikan. Orange selalu lebih baik dalam faktor inovasi, kualitas, atau keandalan. Orange memudahkan pengguna bermain dengan data sumber terbuka dan melakukan proses analisis data secara intuitif (Wiguna and Rifai 2021).

**C. Confussion Matrix** Terdapat suatu konsep dalam datamining dimana penghitungan akurasi dapat menggunakan salah satu metode yaitu *confusion matrix*, dimana pada *confusion matrix* ini perhitungan dapat menghasilkan 4 keluaran yaitu akurasi, *recall*, *precision*, dan *error rate*. Hasil evaluasi model klasifikasi didasarkan pada pengujian untuk memperkirakan objek benar dan salah (Pratama, Hellyana, and Fadlilah 2022).

**D. Cross Validation** Metode validasi silang digunakan untuk mempercepat waktu komputasi pada saat membangun model sekaligus memastikan hasil evaluasi model yang dihasilkan akurat. Validasi silang, atau rotasi perkiraan, adalah teknik validasi yang digunakan untuk mengevaluasi hasil analisis. Teknik validasi silang sering digunakan untuk memprediksi tingkat akurasi suatu model prediktif, dan merupakan salah satu teknik validasi silang dimana data dibagi menjadi K bagian dari dataset yang memiliki jumlah yang sama yaitu data (Firmansyach, Hayati, and Arie Wijaya 2023).

## E. Tahapan Penelitian



### 1) Identifikasi Masalah

Berdasarkan latar belakang yang ada, penelitian ini mengidentifikasi masalah untuk membandingkan algoritma dengan menggunakan 7 algoritma yaitu *Neural Network*, *Decision Tree*, *Naive Bayes*, *K-Nearest Neighbor*, *Logistic Regression*, *Random Forest* dan *Support Vector Machine*. Algoritma

# KLASIFIKASI PENYAKIT KANKER PARU MENGGUNAKAN PERBANDINGAN ALGORITMA MACHINE LEARNING

untuk memahami model klasifikasi dengan akurasi tertinggi untuk menentukan seseorang terkena kanker atau tidak.

## 2) Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah Dataset Predic Terkena Penyakit Paru-Paru. Basis data tersebut berupa kumpulan data untuk mendapatkan nilai akurasi terbaik dari penelitian yang di lakukan dan menggunakan 9 atribut yaitu usia, jenis kelamin, merokok, bekerja, rumah tangga, aktifitas begadang, aktifitas olahraga, asuransi kesehatan, penyakit bawaan.

**Tabel 1. Data Atribut**

No	Atribut	Keterangan
1	Usia	Digunakan
2	Jenis Kelamin	Digunakan
3	Merokok	Digunakan
4	Bekerja	Digunakan
5	Rumah Tangga	Digunakan
6	Aktifitas Begadang	Digunakan
7	Aktifitas Olahraga	Digunakan
8	Asuransi Kesehatan	Digunakan
9	Penyakit Bawaan	Digunakan

## 3) Pengolahan Data

Data diolah dengan 3 tahap yaitu data cleaning, normaliasi data, dan pembagian data.

- *Data Cleaning*: Proses data *cleaning* berdasarkan *record* dari setiap atribut dan pada dataset ini tidak memiliki *missing value*.
- *Normalisasi Data*: Data terkadang berisi nilai dengan rentang yang berbeda-beda. Oleh karena itu mempengaruhi hasil pengukuran analisis data, sehingga harus ada cara untuk menormalkan data tersebut.

Normalisasi data adalah proses menempatkan skala nilai suatu atribut ke dalam rentang yang lebih kecil dengan bobot yang sama. Pengukuran nilai atribut data baru dapat menghapus fitur yang memiliki noise tinggi dan bernilai rendah. Metode ini menormalkan data kolom atribut entitas ke rentang [0,1] menggunakan skala min dan max. Rumus konversi standar ditunjukkan pada Persamaan.

$$n_i^1 = \frac{n_i - \text{min}_A}{\text{maks}_A - \text{min}_A}$$

$n_i^1$  adalah hasil min-max scalling

$n_i$  adalah data yang akan di normalisasi

$\text{min}_A$  adalah nilai minimum pada atribut

$\text{maks}_A$  adalah nilai maksimal pada atribut

- Pembagian Data: Data yang di olah dalam pembelajaran mesin di gunakan 100% untuk pengujian. Padaa tahapan ini uji coba terhadap pengolaha data dengan menggunakan *software* Orange.

#### 4) Klasifikasi Model

##### A. *Decission Tree*

*Decission Tree* merupakan metode klasifikasi dan prediksi yang sangat efektif dan terkenal. Pada pohon keputusan ini, data berupa fakta diubah menjadi pohon keputusan yang berisi aturan dan tentunya lebih mudah dipahami dalam bahasa alami. Model pohon keputusan banyak digunakan ketika data mempunyai nilai diskrit. Meskipun, hal ini tidak menghalangi kemungkinan juga dapat digunakan untuk data yang mengandung atribut numerik (Verawati and Hasibuan 2021).

##### B. *Naïve Bayes*

Pengklasifikasi *Naive Bayes* adalah metode klasifikasi berdasarkan teorema Bayes. Metode klasifikasi menggunakan metode probabilitas dan statistic yang dikemukakan oleh ilmuwanasal Inggris Thomas Bayes, yaitu memprediksi kemungkinan masa depan berdasarkan pengalaman periode sebelumnya, sehingga dikenal dengan teorema Bayes. (Alvina Felicia Watratan, Arwini Puspita. B, and Dikwan Moeis 2020).

# KLASIFIKASI PENYAKIT KANKER PARU MENGGUNAKAN PERBANDINGAN ALGORITMA MACHINE LEARNING

## C. *K-Nearest Neighbor*

*K-NN* atau *K-Nearest Neighbor* merupakan algoritma yang menemukan kasus dengan mengevaluasi seberapa dekat kasus baru dengan kasus lama, Ide utama *K-NN* adalah mencari jarak terpendek antara data yang akan diestimasi dengan data terdekatnya, Nilai jarak data uji dan data latih diurutkan berdasarkan nilai terkecil (Safitri, Hilabi, and Nurapriani 2023).

## D. *Random Forest*

*Random Forest* adalah salah satu metode yang digunakan untuk klasifikasi ketika membangun banyak pohon klasifikasi. Metode ini dapat meningkatkan hasil akurat dengan menghasilkan node anak untuk setiap *node* (*node* di atasnya) dan memilihnya secara acak, Kemudian, hasil klasifikasi setiap pohon dikumpulkan dan hasil klasifikasi yang paling sering dipilih (Sandag 2020).

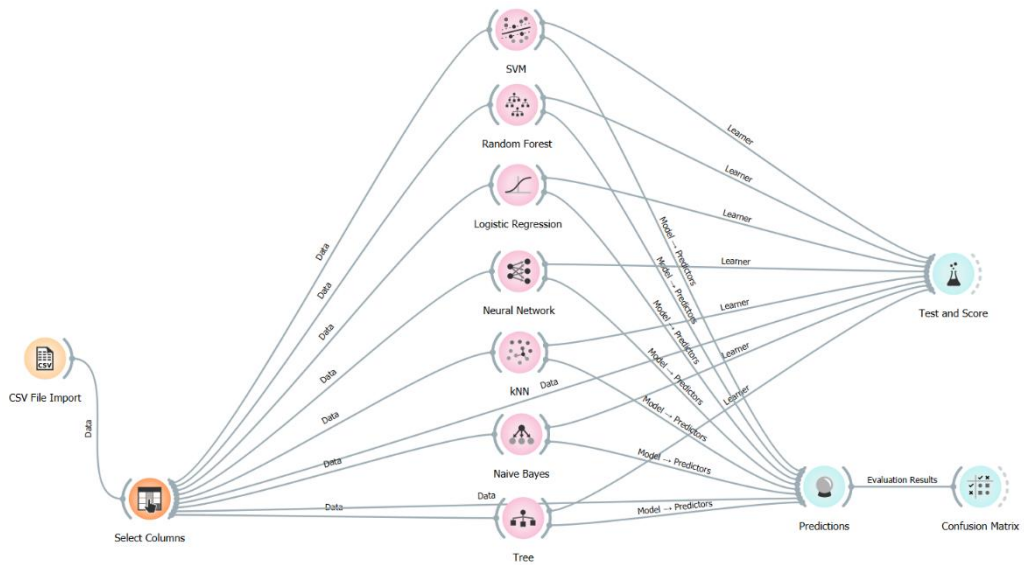
## E. *Support Vector Machines*

*Support vector machine* merupakan sistem pembelajaran yang menggunakan hipotesis berupa fungsi linier tentang fungsi besar berdimensi dan dilatih menggunakan algoritma pembelajaran berdasarkan teori optimasi. Dari perspektif geometris, *SVM* sangat cepat dan efisien untuk masalah klasifikasi teks. *Biner* klasifikasi dapat dianggap sebagai *hyperplane* dalam ruang fitur yang memisahkan titik-titik yang mewakili kasus positif dari kelas yang mewakili keadaan negatif *Neural Network* (Sandag 2020).

## F. *Logistic Regresion*

*Regresi logistik* (logistik regresi) adalah bagian dari analisis regresi yang digunakan apabila variabel terikat (*respon*) merupakan variabel dikotomis. Variabel dikotomi biasanya hanya terdiri dari dua nilai yang mewakili ada tidaknya suatu peristiwa, biasanya diberi angka 0 atau 1 (Alnur et al. 2023).





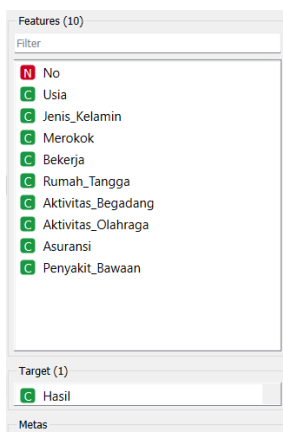
**Gambar 1. Model Klasifikasi**

### 5) Evaluasi Model

Berdasarkan hasil pengujian di dapatkan tingkat akurasi dari masing-masing algoritma yang akan di bandingkan tingkat akurasi mana yang paling baik.

## HASIL DAN PEMBAHASAN

Data set yang telah dipilih berdasarkan informasi yang tersedia menjadikan “hasil” sebagai target untuk mendapatkan akurasi pengujian yang akan di lakukan.

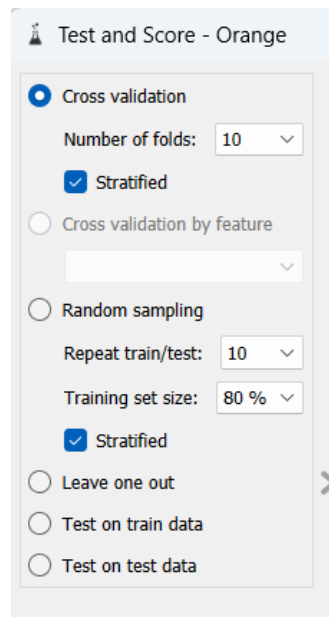


**Gambar 2. Informasi Data Set**

# KLASIFIKASI PENYAKIT KANKER PARU MENGGUNAKAN PERBANDINGAN ALGORITMA MACHINE LEARNING

Klasifikasi dilakukan secara otomatis oleh aplikasi *orange* dengan hasil yang terlihat pada widget Test and Score dengan menggunakan konfigurasi *10-fold cross validation* sebagai metode evaluasi seperti yang terlihat pada gambar

Hasil pada *widget Test and Score* ini merupakan nilai *accuracy*, *precision* dan *recall* ketujuh algoritma seperti terlihat pada Gambar 3.



**Gambar 3. Konfigurasi widget Test and Score**

Model	AUC	CA	F1	Prec	Recall	MCC
Tree	0.994	0.952	0.951	0.955	0.952	0.906
Naive Bayes	0.852	0.871	0.871	0.871	0.871	0.741
kNN	0.804	0.774	0.774	0.775	0.774	0.549
Logistic Regression	0.993	0.947	0.946	0.952	0.947	0.898
Random Forest	0.998	0.976	0.976	0.976	0.976	0.952
	0.993	0.947	0.946	0.952	0.947	0.898
SVM	0.820	0.553	0.485	0.576	0.553	0.107

**Gambar 4. Hasil Cross Validation**

Terlihat bahwa algoritma *Random Forest* memiliki nilai *accuracy*, *precision* dan *recall* tertinggi di dibandingkan algoritma yang lainnya. Performa ketujuh algoritma dalam mengklasifikasikan data set yang digunakan terlihat pada *widget Confussion Matrix*, dimana di peroleh nilai prediksi untuk kelas *true negative*, *false positive*, *false negative*, dan *true positive* ketujuh algoritma seperti pada gambar berikut.

		Predicted		$\Sigma$
		Tidak	Ya	
Actual	Tidak	15375	273	15648
	Ya	482	13870	14352
$\Sigma$		15857	14143	30000

**Gambar 5. Hasil Confusion Matrix Random Forest**

Dari hasil *confusion matrix* algoritma *Random Forest* pada gambar, terlihat bahwa dari 15648 data status pasien bukan kanker paru-paru, berhasil diprediksi dengan benar 15375 data sedangkan 273 data sisanya gagal. Sedangkan dari 14352 data status pasien penderita kanker paru-paru, berhasil di prediksi 13870 data dengan benar sedangkan 482 data sisanya gagal.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Population}} \times 100\%$$

$$\text{Accuracy} = \frac{15375 + 13870}{15648 + 14352} \times 100\%$$

$$\text{Accuracy} = \frac{29245}{30000} \times 100\%$$

$$\text{Accuracy} \approx 97.483\%$$

So, the accuracy is approximately 97.483%.

**Gambar 6. Penghitungan Akurasi Random Forest**

Hasil dari gambar tersebut pengolahan data menggunakan aplikasi orange dengan algoritma *Random Forest* didapat akurasi sebesar 97.48%.

		Predicted		$\Sigma$
		Tidak	Ya	
Actual	Tidak	14018	1630	15648
	Ya	11789	2563	14352
$\Sigma$		25807	4193	30000

# KLASIFIKASI PENYAKIT KANKER PARU MENGGUNAKAN PERBANDINGAN ALGORITMA MACHINE LEARNING

**Gambar 7. Hasil Confussion Matrix Support Vector Machines**

Dari hasil confusion matrix algoritma SVM pada gambar, terlihat bahwa dari 15648 data status pasien bukan kanker paru-paru, berhasil diprediksi dengan benar 14018 data sedangkan 1630 data sisanya gagal. Sedangkan dari 14352 data status pasien penderita kanker paru-paru, berhasil di prediksi 2563 data dengan benar sedangkan 11789 data sisanya gagal.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Population}} \times 100\%$$

$$\text{Accuracy} = \frac{14018 + 2563}{15648 + 14352} \times 100\%$$

$$\text{Accuracy} = \frac{16581}{30000} \times 100\%$$

$$\text{Accuracy} \approx 55.27\%$$

So, the accuracy is approximately 55.27%.

**Gambar 8. Penghitungan akurasi Support Vector Machines**

Hasil dari gambar tersebut pengolahan data menggunakan aplikasi orange dengan algoritma *Support Vector Machines* didapat akurasi sebesar 55.27%.

		Predicted		Σ
		Tidak	Ya	
Actual	Tidak	15648	0	<b>15648</b>
	Ya	1602	12750	<b>14352</b>
Σ		<b>17250</b>	<b>12750</b>	<b>30000</b>

**Gambar 9. Hasil Confussion Matrix Neural Network**

Dari hasil *confussion matrix* algoritma *Neural Network* pada gambar, terlihat bahwa dari 15648 data status pasien bukan kanker paru-paru, berhasil diprediksi dengan benar 15648 data sedangkan 0 data data sisanya gagal. Sedangkan dari 14352 data status pasien penderita kanker paru-paru, berhasil di prediksi 12750 data dengan benar sedangkan 1602 data sisanya gagal.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Population}} \times 100\%$$

$$\text{Accuracy} = \frac{15648 + 12750}{15648 + 14352} \times 100\%$$

$$\text{Accuracy} = \frac{28498}{30000} \times 100\%$$

$$\text{Accuracy} \approx 94.993\%$$

So, the accuracy is approximately 94.993%.

### Gambar 10. Hasil Perhitungan Akurasi *Neural Network*

Hasil dari gambar tersebut pengolahan data menggunakan aplikasi orange dengan algoritma *Neural Network* didapat akurasi sebesar 94.993%

		Predicted		$\Sigma$
		Tidak	Ya	
Actual	Tidak	13680	1968	15648
	Ya	1916	12436	14352
$\Sigma$		15596	14404	30000

### Gambar 11. Hasil *Confussion Matrix Naïve Bayes*

Dari hasil *confussion matrix algoritma Naïve Bayes* pada gambar, terlihat bahwa dari 15648 data status pasien bukan kanker paru-paru, berhasil diprediksi dengan benar 13680 data sedangkan 1968 data data sisanya gagal. Sedangkan dari 14352 data status pasien penderita kanker paru-paru, berhasil di prediksi 12436 data dengan benar sedangkan 1916 data sisanya gagal.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Population}} \times 100\%$$

$$\text{Accuracy} = \frac{13680 + 12436}{15648 + 14352} \times 100\%$$

$$\text{Accuracy} = \frac{26116}{30000} \times 100\%$$

$$\text{Accuracy} \approx 87.053\%$$

So, the accuracy is approximately 87.053%.

# KLASIFIKASI PENYAKIT KANKER PARU MENGGUNAKAN PERBANDINGAN ALGORITMA MACHINE LEARNING

**Gambar 12. Hasil Akurasi Naïve Bayes**

Hasil dari gambar tersebut pengolahan data menggunakan aplikasi orange dengan algoritma Naïve Bayes didapat akurasi sebesar 87.053%

		Predicted		Σ
		Tidak	Ya	
Actual	Tidak	15586	62	15648
	Ya	1389	12963	14352
Σ		16975	13025	30000

**Gambar 13. Hasil Confussion Matrix Decission Tree**

Dari hasil *confussion matrix algoritma Decission Tree* pada gambar, terlihat bahwa dari 15648 data status pasien bukan kanker paru-paru, berhasil diprediksi dengan benar 15586 data sedangkan 62 data sisanya gagal. Sedangkan dari 14352 data status pasien penderita kanker paru-paru, berhasil di prediksi 12963 data dengan benar sedangkan 1389 data sisanya gagal.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Population}} \times 100\%$$

$$\text{Accuracy} = \frac{15586 + 12963}{15648 + 14352} \times 100\%$$

$$\text{Accuracy} = \frac{28549}{30000} \times 100\%$$

$$\text{Accuracy} \approx 95.16\%$$

So, the accuracy is approximately 95.16%.

**Gambar 13. Hasil Akurasi Decission Tree**

Hasil dari gambar tersebut pengolahan data menggunakan aplikasi orange dengan algoritma *Decission Tree* didapat akurasi sebesar 95.16%.

		Predicted		$\Sigma$
		Tidak	Ya	
Actual	Tidak	12044	3604	15648
	Ya	3165	11187	14352
$\Sigma$		15209	14791	30000

**Gambar 14. Hasil *confussion Matrix K-neareast neighbor***

Dari hasil *confussion matrix* algoritma *K-neirest neighbor* pada gambar, terlihat bahwa dari 15648 data status pasien bukan kanker paru-paru, berhasil diprediksi dengan benar 12044 data sedangkan 3604 data data sisanya gagal. Sedangkan dari 14352 data status pasien penderita kanker paru-paru, berhasil di prediksi 11187 data dengan benar sedangkan 3165 data sisanya gagal.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Population}} \times 100\%$$

$$\text{Accuracy} = \frac{12044 + 11187}{15648 + 14352} \times 100\%$$

$$\text{Accuracy} = \frac{23231}{30000} \times 100\%$$

$$\text{Accuracy} \approx 77.437\%$$

So, the accuracy is approximately 77.437%.

**Gambar 15. Hasil Akurasi *K-neirest neighbor***

Hasil dari gambar tersebut pengolahan data menggunakan aplikasi orange dengan algoritma *kNN* didapat akurasi sebesar 77.437 %.

		Predicted		$\Sigma$
		Tidak	Ya	
Actual	Tidak	15648	0	15648
	Ya	1602	12750	14352
$\Sigma$		17250	12750	30000

**Gambar 16. Hasil *Confussion Matrix Logistic Regression***

# KLASIFIKASI PENYAKIT KANKER PARU MENGGUNAKAN PERBANDINGAN ALGORITMA MACHINE LEARNING

Dari hasil *confusion matrix* algoritma *Logistic Regression* pada gambar, terlihat bahwa dari 15648 data status pasien bukan kanker paru-paru, berhasil diprediksi dengan benar 15648 data sedangkan 0 data sisanya gagal. Sedangkan dari 14352 data status pasien penderita kanker paru-paru, berhasil di prediksi 12750 data dengan benar sedangkan 1602 data sisanya gagal.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Population}} \times 100\%$$

$$\text{Accuracy} = \frac{15648 + 12750}{15648 + 14352} \times 100\%$$

$$\text{Accuracy} = \frac{28498}{30000} \times 100\%$$

$$\text{Accuracy} \approx 94.993\%$$

So, the accuracy is approximately 94.993%.

## Gambar 17. Akurasi *Logistic Regression*

Hasil dari gambar tersebut pengolahan data menggunakan aplikasi orange dengan algoritma *Logistic Regression* didapat akurasi sebesar 94.993%

Pengujian menggunakan aplikasi Orange dengan masing-masing algoritma dan nilai *performance* dari setiap algoritma dapat dibandingkan dengan melihat *tabel*

Jadi berdasarkan nilai *performance* untuk penentuan dengan nilai akurasi yang ada, untuk algoritma *Random Forest*, *Tree*, *Neural Network*, dan *Logistic Regression* memiliki nilai akurasi di atas 90% sementara untuk algoritma *SVM*, *Naïve Bayes*, dan *kNN* memiliki nilai akurasi di bawah 90%. Dapat di ambil kesimpulan bahwa algortima yang memiliki tingkat akurasi yang lebih tinggi dapat digunakan dalam klasifikasi penentuan terkena dan tidak nya kanker paru-paru.

**Tabel 2. Hasil Komparasi Nilai Akurasi**

No	Nama Algoritma	Nilai Akurasi
1	<i>Random Forest</i>	97.48 %
2	<i>SVM</i>	55.27 %
3	<i>Neural Network</i>	94.993 %



4	Naïve Bayes	87.053 %
5	<i>Tree</i>	95.16 %
6	kNN	77.437 %
7	<i>Logistic Regression</i>	94.993 %

## KESIMPULAN DAN SARAN

Dari hasil penelitian yang dilakukan klasifikasi 30000 data selama pengujian untuk algoritma *Random Forest*, *Tree*, *Neural Network*, dan *Logistic Regression* memiliki nilai akurasi di atas 90%. Penerapan data *mining* menggunakan metode Algoritma *Random Forest*, *Tree*, *Neural Network*, dan *Logistic Regression* ini dapat mempercepat pengambilan keputusan dalam meprediksi untuk kkasifikasi penentuan terkena dan tidaknya kanker paru-paru.

Penelitian selanjutnya sebaiknya menambah atribut dan memperbanyak data uji agar nilai akurasi lebih akurat dan tinggi. Dapat membandingkan dengan metode lainnya sehingga dapat membandingkan metode mana yang lebih baik.

## UCAPAN TERIMA KASIH

Dalam kesempatan ini, penulis ingin menyampaikan ucapan terima kasih yang tulus kepada semua pihak yang telah memberikan dukungan dan bantuan dalam penyelesaian artikel ini. Tanpa bantuan mereka, artikel ini tidak mungkin menjadi kenyataan.

Pertama-tama, penulis ingin mengucapkan terima kasih kepada pihak penyandang dana penelitian yang telah memberikan dukungan finansial untuk penelitian ini. Dukungan ini memungkinkan penulis untuk melakukan penelitian dengan lebih fokus dan mendalam.

Selanjutnya, penulis ingin mengungkapkan rasa terima kasih kepada semua pihak yang telah memberikan fasilitas dan dukungan teknis selama proses penelitian. Fasilitas yang disediakan sangat membantu dalam pengumpulan data dan analisis yang akurat.

Tak lupa, terima kasih kepada semua pihak yang telah memberikan masukan, saran, dan kritik konstruktif selama penyusunan artikel ini. Masukan tersebut sangat berharga dan telah meningkatkan kualitas artikel.

# KLASIFIKASI PENYAKIT KANKER PARU MENGGUNAKAN PERBANDINGAN ALGORITMA MACHINE LEARNING

Apabila artikel ini merupakan bagian dari skripsi, tesis, disertasi, makalah konferensi, atau hasil penelitian lainnya, penulis ingin menyampaikan apresiasi kepada dosen pembimbing dan tim penguji yang telah memberikan panduan dan masukan yang sangat berarti.

Akhir kata, penulis menyadari bahwa kerjasama dan dukungan dari berbagai pihak telah menjadi kunci keberhasilan dalam penyelesaian artikel ini. Semoga hasil penelitian ini dapat memberikan kontribusi positif dan bermanfaat bagi perkembangan ilmu pengetahuan. Sekali lagi, terima kasih kepada semua yang telah berperan dalam kesuksesan artikel ini.

## DAFTAR REFERENSI

- Alnur, Bima, Mulyono, Fitri Amillia, and Sutoyo Sutoyo. 2023. "JITE (Journal of Informatics and Telecommunication Engineering)." *Journal of Informatics and Telecommunication Engineering* 7(1):102–11.
- Alvina Felicia Watratan, Arwini Puspita. B, and Dikwan Moeis. 2020. "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia." *Journal of Applied Computer Science and Technology* 1(1):7–14. doi: 10.52158/jacost.v1i1.9.
- Angkasa, Vincent, and Jefri Junifer Pangaribuan. 2022. "Komparasi Tingkat Akurasi *Random Forest* Dan KNN Untuk Mendiagnosis Penyakit Kanker Payudara." *Journal Information System Development (ISD)* 7(1):49–61.
- Annisa, Riski. 2019. "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung." *Jurnal Teknik Informatika Kaputama (JTIK)* 3(1):22–28.
- Desiani, Anita, Sri Indra Maiyanti, Yuli Andriani, Bambang Suprihatin, Ali Amran, Nyanyu Chika Marselina, and Aulia Salsabila. 2023. "Perbandingan Klasifikasi Penyakit Kanker Paru-Paru Menggunakan Support Vector Machine Dan K-Nearest Neighbor." *Jurnal PROCESSOR* 18(1):54–62. doi: 10.33998/processor.2023.18.1.700.
- Firmansyach, Wildan Attariq, Umi Hayati, and Yudhistira Arie Wijaya. 2023. "Analisa

- Terjadinya Overfitting Dan Underfitting Pada Algoritma Naive Bayes Dan Decision Tree Dengan Teknik Cross Validation.” *JATI (Jurnal Mahasiswa Teknik Informatika)* 7(1):262–69. doi: 10.36040/jati.v7i1.6329.
- Khultsum, Umi, Fajar Sarasati, and Ghofar Taufik. 2022. “Penerapan Metode Mobile-Net Untuk Klasifikasi Citra Penyakit Kanker Paru-Paru.” *JURIKOM (Jurnal Riset Komputer)* 9(5):1366. doi: 10.30865/jurikom.v9i5.4918.
- Mesin, Jurnal Teknik, Elektro Dan, and Informatika Jtmei. 2023. “Analisis Terjadinya Kanker Paru-Paru Pada Pasien Menggunakan Decision Tree : Penerapan Algoritma C4 . 5 Dan RapidMiner Untuk Menentukan Risiko Kanker Pada Gejala Pasien Deigo Anugrah Pratama Program Studi Sistem Informasi FTI Universitas Bina Sarana Infor.” 2(4).
- Naufal, Shidqi Aqil, Adiwijaya Adiwijaya, and Widi Astuti. 2020. “Analisis Perbandingan Klasifikasi Support Vector Machine (SVM) Dan K-Nearest Neighbors (KNN) Untuk Deteksi Kanker Dengan Data Microarray.” *JURIKOM (Jurnal Riset Komputer)* 7(1):162. doi: 10.30865/jurikom.v7i1.2014.
- Pratama, Eva Argarini, Corie Mei Hellyana, and Nuzul Imam Fadlilah. 2022. “Perbandingan 3 Algoritma Klasifikasi Data Mining Dalam Pro-Kontra Bahaya Rokok Elektrik.” *Jurnal Teknoinfo* 16(1):93. doi: 10.33365/jti.v16i1.1534.
- Putra, Tommy Dwi, Ema Utami, and Mei P.Kurniawan. 2022. “Klasifikasi Penderita Kanker Paru Paru Menggunakan Algoritma Artificial Neural Network (ANN).” *Explore* 12(2):13. doi: 10.35200/explore.v12i2.568.
- Safitri, Dinda, Shofa Shofiah Hilabi, and Fitria Nurapriani. 2023. “Analisis Penggunaan Algoritma Klasifikasi Dalam Prediksi Kelulusan Menggunakan Orange Data Mining.” *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab* 8(1):75–81. doi: 10.36341/rabit.v8i1.3009.
- Sandag, Green Arther. 2020. “Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest.” *CogITO Smart Journal* 6(2):167–78. doi: 10.31154/cogito.v6i2.270.167-178.
- Sari, Laura, Annisa Romadloni, and Rostika Listyaningrum. 2023. “Penerapan Data Mining Dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random

# KLASIFIKASI PENYAKIT KANKER PARU MENGGUNAKAN PERBANDINGAN ALGORITMA MACHINE LEARNING

*Forest.*” *Infotekmesin* 14(1):155–62. doi: 10.35970/infotekmesin.v14i1.1751.

Septhya, Dhini, Kharisma Rahayu, Salsabila Rabbani, Vindi Fitria, Rahmaddeni Rahmaddeni, Yuda Irawan, and Regiolina Hayami. 2023. “Implementasi Algoritma Decision *Tree* Dan Support Vector Machine Untuk Klasifikasi Penyakit Kanker Paru.” *MALCOM: Indonesian Journal of Machine learning and Computer Science* 3(1):15–19. doi: 10.57152/malcom.v3i1.591.

Sowah, Robert A., Adelaide A. Bampoe-Addo, Stephen K. Armoo, Firibu K. Saalia, Francis Gatsi, and Baffour Sarkodie-Mensah. 2020. “Design and Development of Diabetes Management System Using *Machine learning*.” *International Journal of Telemedicine and Applications* 2020. doi: 10.1155/2020/8870141.

Verawati, Yulia, and Muhammad Said Hasibuan. 2021. “Perbandingan Data Set IRIS Dengan Aplikasi Rapid Miner Dan Orange Menggunakan Algoritma Klasifikasi.” *Institut Informatika Dan Bisnis Darmajaya* 158–63.

Vierisyah, Aldo, Ricky Maulana Fajri, Program Studi, Sistem Komputer, Universitas Indo, and Global Mandiri. n.d. “KLASIFIKASI KANKER PARU PARU MENGGUNAKAN CNN DENGAN 5 ARSITEKTUR.” (629):84–91.

Wiguna, Ratu Aghnia raffaidy, and Andri Irfan Rifai. 2021. “Analisis Text Clustering Masyarakat Di Twitter Mengenai Omnibus Law Menggunakan Orange Data Mining.” *Journal of Information Systems and Informatics* 3(1):1–12. doi: 10.33557/journalisi.v3i1.78.

Wulandari, Eva. 2022. “Klasifikasi Kanker Paru-Paru Menggunakan Metode Naive Bayes.” *International Research on Big-Data and Computer Technology: I-Robot* 6(2):20–24. doi: 10.53514/ir.v6i2.325.