

Pendekatan *Machine Learning* untuk Deteksi Dini Kanker Paru-Paru: Mengoptimalkan Sensitivitas dan Akurasi

Nia Nuraeni¹, Puji Astuti^{2*}

^{1,2} Sistem Informasi, Teknik dan Informatika, Universitas Bina Sarana Informatika, Indonesia
nia.nne@bsi.ac.id, puji.pat@bsi.ac.id

Abstrak

Kanker paru-paru merupakan penyebab utama kematian akibat kanker secara global. Deteksi dini secara akurat sangat penting untuk meningkatkan peluang kesembuhan pasien. Pendekatan machine learning semakin banyak digunakan dalam dunia medis karena kemampuannya dalam mengolah data secara cepat. Oleh karena itu, machine learning berpotensi menjadi alat bantu yang efektif dalam proses deteksi dini kanker. Penelitian ini bertujuan untuk membandingkan kinerja enam algoritma data mining *Logistic Regression* (LR), *Random Forest* (RF), *Support Vector Machine* (SVM), *Naive Bayes* (NB), *Decision Tree Classifier* (DTC) dan *K-Nearest Neighbors* (K-NN) dalam memprediksi kanker paru-paru. Data yang digunakan berasal dari dataset publik yang telah melalui proses prapengolahan. Evaluasi dilakukan menggunakan metrik akurasi, sensitivitas, dan *F1-Score*. Hasil menunjukkan bahwa *Decision Tree Classifier* memiliki sensitivitas tertinggi (0,558) dan *F1-Score* terbaik (0,543), menjadikannya paling efektif dalam mendeteksi kasus positif. *Random Forest* mencatatkan akurasi tertinggi (90%), menunjukkan kinerja seimbang secara keseluruhan. *Logistic Regression* dan SVM memiliki akurasi tinggi (89%) tetapi sensitivitasnya lebih rendah. Kesimpulannya, model dengan sensitivitas tinggi cocok untuk *skrining* awal guna meminimalkan kasus terlewat, sementara model dengan akurasi tinggi dapat digunakan untuk diagnosis lebih lanjut. Pendekatan *machine learning* menunjukkan potensi besar dalam deteksi dini kanker paru-paru.

Kata kunci: akurasi, deteksi dini, kanker paru-paru, machine learning, sensitivitas

1. Pendahuluan

Kanker paru-paru merupakan salah satu dari tiga penyakit yang paling berbahaya di dunia dan dapat berkembang dengan sangat cepat (Rofiani et al., 2024). Kanker paru-paru merupakan jenis kanker dimana bermula dari sel-sel abnormal tumbuh secara tidak terkendali di paru-paru. Penyakit kanker paru-paru dapat disebabkan dari dalam ataupun dari luar paru-paru (Shafa et al., 2024). Adapun faktor yang dapat menjadi penyebab kanker paru pada orang tidak merokok diantaranya asbestos, radon dan karna polusi udara (Buana & Harahap, 2022). Dengan munculnya penyakit kanker paru-paru di tubuh manusia ada beberapa gejala yang dapat dirasakan oleh kebanyakan pasien kanker paru-paru. Namun, gejala ini seringkali diabaikan dan tidak terdeteksi oleh praktisi medis sehingga hanya 14% dari pasien yang didiagnosa kanker paru-paru sembuh dari penyakitnya (Marzuq et al., 2023). Kanker paru termasuk dalam penyakit yang ganas dimana dapat disebabkan oleh perubahan genetika pada sel epitel saluran pernafasan, sehingga dapat terjadi proliferasi sel yang tidak dapat dikendalikan (Buana & Harahap, 2022). 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. Dan di Indonesia dengan kasus kanker paru-paru mencapai 34.78 di tahun 2020 (Naseh Khudori & Syauqi Haris, 2024). Kanker paru-paru dapat dideteksi saat kanker telah memasuki pada tahap stadium lanjut. Banyaknya data diagnosis kanker paru-paru saat ini, dapat digunakan untuk memprediksi kanker paru (Naseh Khudori & Syauqi Haris, 2024). Observasi menjadi acuan untuk analisis penyakit paru-paru bagaimana penerapan teknik data mining pada analisis prediksi analisis kanker paru (A. Y. Permana et al., 2023). Data mining adalah metode yang digunakan untuk mengungkap informasi berharga dengan menganalisis pola atau hubungan tertentu dalam data yang berskala besar (Duran et al., 2024). Dalam proses data mining, prediksi penyakit kanker paru-paru dapat dilakukan dengan menggunakan algoritma *logistic regression* dan *svm* (Rahmaeda et al., 2025).. Adapun dalam prediksi kanker paru-paru algoritma yang digunakan *naive bayes*, prediksi dengan *naive bayes* memiliki nilai *accuracy* sebesar 98,77% (Wulandari, 2022). Algoritma *random forest* juga dikenal untuk memprediksi penyakit dini pada kanker paru-paru, dalam penelitian ini didapat nilai *accuracy* dengan *random forest* 91,32% (Sandika et al., 2024). Metode *decision tree* dapat memprediksi kanker paru-paru dengan tingkat akurasi 94.66%

didapat dari 10 macam atribut (Anzori & Fatah, 2024).

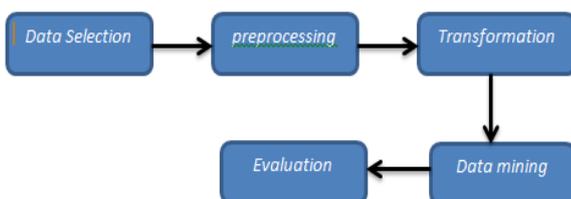
Dari beberapa hasil penelitian terdahulu dengan tingkat *accuracy* yang berbeda-beda menunjukkan bahwa setiap model algoritma baik digunakan dalam jumlah data yang kecil maupun besar dengan hasil yang memuaskan disetiap penelitian dalam hal pemecahan masalah kanker paru-paru.

Tujuan dari penelitian ini adalah untuk membandingkan dari enam model metode algoritma untuk menentukan metode mana yang paling efektif dan akurat dalam mengklasifikasikan atau memprediksi adanya kanker paru-paru pada pasien berdasarkan data yang tersedia. Setiap metode memiliki kelebihan dan kekurangan, dan tujuan dari perbandingan ini adalah untuk meningkatkan akurasi prediksi, memahami kinerja dari masing-masing metode. Mengoptimalkan proses deteksi kanker paru-paru, dan beberapa metode bekerja lebih baik dalam kondisi tertentu, dan dengan membandingkan enam metode ini, dapat dievaluasi bagaimana setiap metode menggeneralisasi hasil data. Deteksi dini kanker paru-paru sangat krusial karena pada tahap awal penyakit ini sering tidak menunjukkan gejala yang jelas. Oleh karena itu, sistem prediksi yang digunakan harus mampu mendeteksi sebanyak mungkin kasus positif (penderita kanker), sekaligus meminimalkan kesalahan dan klasifikasi. Untuk mengimplementasikan data mining menggunakan bahasa pemrograman Python, dimana memanfaatkan pustaka (*library*) yang tersedia guna mempermudah pelaksanaan *machine learning* (permana & Djamaluddin, 2023).

2. Metode

Metode *Knowledge Discovery in Database Process* (KDD) yaitu metode untuk mendapatkan sebuah informasi melalui basis data yang telah tersedia (Alghifari & Juardi, 2021).

Penelitian ini menggunakan pendekatan metode *Knowledge Discovery in Database Process* (KDD), yang mana ada beberapa tahapan, yaitu *data selection*, *preprocessing*, *transformation*, *data mining* dan *evaluation*. Berikut adalah diagram KDD.



Gambar 1. Tahapan Penelitian

1. Data Selection

Pada tahap ini adalah tahap persiapan untuk pemilihan data. Data diperoleh dari <https://www.kaggle.com/> untuk penyakit *lung cancer*. Kemudian dilakukan proses pemilihan

atribut data selanjutnya dilakukan proses *preprocessing*.

2. Preprocessing

Selanjutnya pada tahap *preprocessing* adalah data mentah dilakukan proses *handling missing value* memastikan apakah ada data yang kosong atau data yang *noise* dengan menggunakan *library pandas*.

3. Transformation

Transformasi data adalah proses mengubah data dari data asli menjadi format data yang sesuai dengan kebutuhan. Teknik ini tujuannya untuk mengubah atribut nominal menjadi numerik (Alfajr & Defiyanti, 2024).

4. Data Mining

Dalam tahap ini data siap diolah dengan menggunakan *machine learning* yaitu *google colabs* pada *library pandas*. Penulis menggunakan enam metode untuk mengevaluasi hasil yang didapat dari data kanker paru-paru yaitu *Logistic Regression (LR)*, *Random Forest (RF)*, *Support Vector Machine (SVM)*, *Naïve Bayes (NB)*, *Decision Tree Classifier (DTC)*, dan *K-Nearest Neighbors (K-NN)*. Alur dari setiap metode adalah membaca data training yang siap untuk digunakan, kemudian perhitungan setiap atribut dan hasil akhir mendapat nilai *accuracy* yang paling tinggi.

5. Evaluation

Pada tahap evaluasi merupakan tahap akhir dalam melakukan proses data mining, *evaluation* diukur dengan empat matrix, yaitu (Herjanto & Carudin, 2024).

- Accuracy* merupakan proporsi dari keseluruhan rangkaian data yang diprediksi dengan tepat.
- Recall* yaitu matrix yang dapat memperlihatkan seberapa banyak data dari kelas positif yang telah diprediksi dengan sangat akurat.
- Precision* yaitu sebuah *indicator* seberapa tepatnya model dalam prediksi data dari kelas positif.
- F-measure* yaitu nilai yang diperoleh dengan menggabungkan rata-rata dari *precision* dan *recall*.

3. Hasil dan Pembahasan

Berdasarkan metode diatas berikut hasil dan pembahasan pada penelitian ini

3.1. Data Selection

Data yang di peroleh dari <https://www.kaggle.com/> berupa dataset yang dikumpulkan dari situs web sistem prediksi penyakit kanker paru-paru dengan jumlah data sebanyak 3000 *record* dengan 16 atribut, seperti ditunjukkan pada Tabel 1. Atribut-atribut tersebut yang selanjutnya akan diolah di dalam *machine learning*.

Tabel 1. Deskripsi atribut pada dataset kanker paru-paru

Atribut	Tipe Atribut	Deskripsi
Gender	Kategorikal	M (Male), F (Female)
Age	Numerik	Age Of Patient
Smoking	Numerik	2=Yes, 1=No
Yellow Fingers	Numerik	2=Yes, 1=No
Anxiety	Numerik	2=Yes, 1=No
Peer_Presure	Numerik	2=Yes, 1=No
Chronic Disease	Numerik	2=Yes, 1=No
Fatigue	Numerik	2=Yes, 1=No
Allergy	Numerik	2=Yes, 1=No
Wheezing	Numerik	2=Yes, 1=No
Alcohol	Numerik	2=Yes, 1=No
Coughing	Numerik	2=Yes, 1=No
Shortness of Breath	Numerik	2=Yes, 1=No
Swallowing Difficulty	Numerik	2=Yes, 1=No
Chest Pain	Numerik	2=Yes, 1=No
Lung Cancer	Label	Yes, No

Tabel 1 diatas merupakan deskripsi atribut pada dataset yang di olah pada penelitian ini, adapun keterangan atribut dari tabel diatas yaitu: *gender*; merupakan jenis kelamin dari dataset yang terdiri dari Male dan Female dengan tipe data Kategorikal, *Age*; usia pasien dengan tipe data *Numeric*, *Smoking*; *yellow fingers*; *anxiety*; *peer_presure*; *chronic_disease*; *fatigue*; *allergy*; *wheezing*; *alcohol*; *coughing*; *shortness of breath*; *swallowing difficulty*; *chest pain*; merupakan keterangan kondisi medis pasien dengan tipe data *numeric* yang di isi angka 2 dan 1 dengan keterangan 2 = *yes* dan 1 = *no*, *lung cancer*; merupakan label dari penelitian berupa keterangan *Yes* dan *No*.

3.2. Preprocessing

Pada tahap ini dilakukan proses *handling missing value* yang memastikan apakah ada data yang kosong atau data yang *noise* dengan menggunakan library *pandas*. Berikut hasil dari proses *preprocessing*

Tabel 2. Hasil pengecekan data kosong

Atribut	Missing Value
GENDER	0
AGE	0
SMOKING	0
YELLOW_FINGERS	0
ANXIETY	0
PEER_PRESURE	0
CHRONIC_DISEASE	0
FATIGUE	0
ALLERGY	0
WHEEZING	0
ALCOHOL_CONSUMING	0
COUGHING	0
SHORTNESS_OF_BREATH	0
SWALLOWING_DIFFICULTY	0
CHEST_PAIN	0
LUNG_CANCER	0

Berdasarkan Tabel 2 diatas, hasil dari pengecekan menampilkan hasil 0, artinya tidak ada data yang kosong pada dataset tersebut.

3.3. Transformation

Proses transformasi ini mencakup pengubahan format data, dalam penelitian ini pengubahan format data ada pada atribut *Gender*. Dimana atribut tersebut di ubah menjadi Numerik yang asalnya berupa atribut Kategorikal. Proses transformasi atau pengubahan format data tersebut menggunakan rumus python seperti ditunjukkan dalam Gambar 2.

```
from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()
dft['GENDER'] = le.fit_transform(dft['GENDER'])
```

Gambar 2. Coding transformasi format data pada atribut Gender

Dengan ada nya transformasi data, maka atribut data yang semula berupa kategorikal telah di ubah menjadi data numerik yang dapat digunakan untuk model *data mining* yang akan di aplikasikan.

3.4. Data Mining

Pada penelitian ini metode machine learning yang digunakan berupa *data mining* yang membandingkan 6 model algorithma, yaitu *Naive Bayes*, *Random Forest*, *K-NN*, *DTC*, dan *Logistic Regression*.

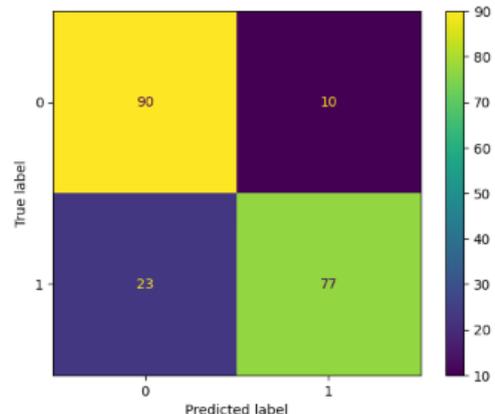
1. Naive Bayes

Naive bayes merupakan metode yang dering digunakan dalam pengolahan data web, khususnya untuk klasifikasi dan kategorisasi teks. Metode ini memiliki keunggulan dalam menangani sejumlah besar data yana banyaka di temukan di internet (Ravinder et al., 2024). Penerapan metode *Naive Bayes* pada penelitian ini dengan mengaplikasi kan ke dalam bahasa pemrograman python dimana dihasilkan akurasi seperti pada Gambar 3.

Akurasi dengan menggunakan Naive Bayes: 0.835
 Confusion Matrix:
 [[90 10]
 [23 77]]
 Classification Report:

	precision	recall	f1-score	support
0	0.80	0.90	0.85	100
1	0.89	0.77	0.82	100
accuracy			0.83	200
macro avg	0.84	0.83	0.83	200
weighted avg	0.84	0.83	0.83	200

 Precision: 0.8407588241277592
 Accuracy: 83.5 %

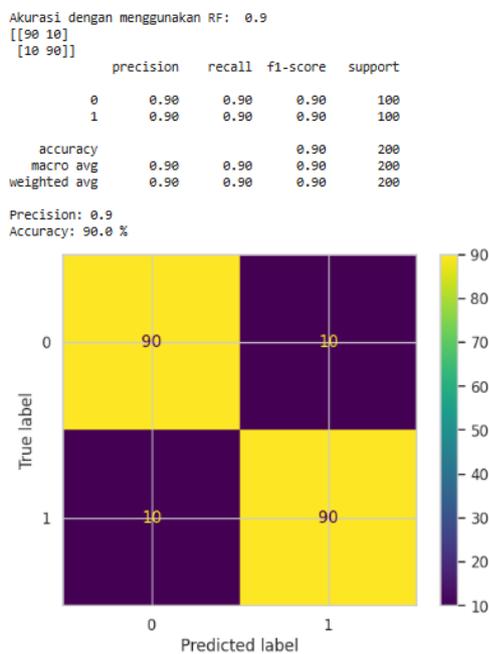


Gambar 3. Hasil penerapan metode Naive Bayes

Gambar diatas menunjukkan nilai akurasi pada metode Naïve Bayes sebesar 83,5%.

2. *Random Forest*

Random Forest merupakan model *Machine Learning* yang banyak digunakan dalam klasifikasi dan peramalan. Metode ini bekerja dengan membuat sebuah pohon keputusan yang dilatih menggunakan subset data secara acak. Hasil dari setiap pohon keputusan digabung kan untuk menghasilkan prediksi akhir. Keunggulan utama dari *Random Forest* adalah kemampuannya menangani data yang tidak seimbang dan mengatasi masalah nilai yang hilang dalam variable. (Salman et al., 2024). Setelah diterapkan dengan menggunakan python dihasilkan akurasi pada Gambar 4.



Gambar 4. Hasil penerapan metode *Random Forest*

Gambar 4 menunjukkan hasil penerapan pada metode *Random Forest* dengan akurasi sebesar 90%.

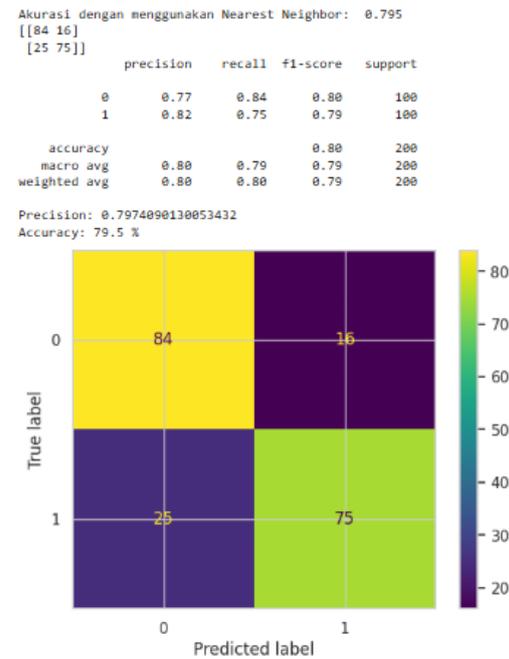
3. K-NN

K-NN diperkenalkan oleh Fix dan Hodges pada tahun 1951 dan merupakan algoritma berbasis supervised learning yang sering digunakan dalam proses klasifikasi dan pengelompokan data, KNN banyak digunakan karena kesederhanaan dan efektivitas nya dalam mengklasifikasi data (Datasets et al., 2024). Gambar 5 menunjukkan dari hasil penerapan algoritma K-NN menggunakan bahasa pemrograman python. Berdasarkan gambar diatas hasil penerapan metode K-NN menghasilkan akurasi sebesar 79.5%

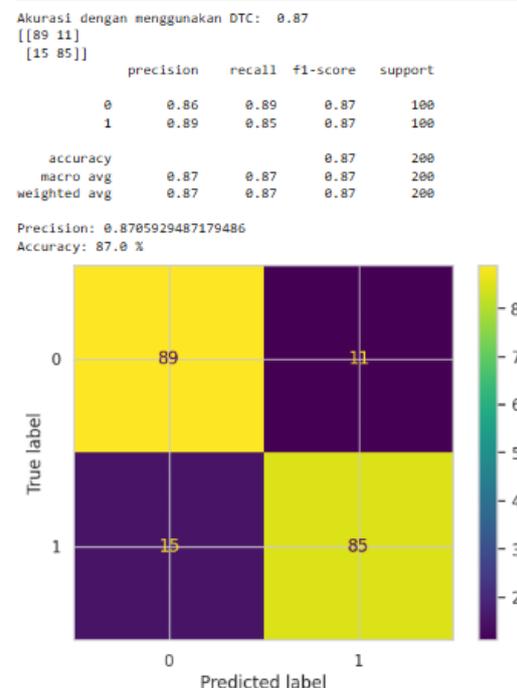
4. *Decision Tree Classifier* (DTC)

Metode *Decision Tree Classifier* (DTC) merupakan salah satu algoritma dalam machine

learning yang banyak digunakan karena kemudahan dan interprebilitasnya. Algoritma in telah menjadi alat penting dalam menyelesaikan berbagai permasalahan kompleks diberbagai bidang (Mienye & Jere, 2024). Dari hasil pengujian yang ditunjukkan pada Gambar 6, terlihat bahwa DTC memiliki tingkat akurasi sebesar 87%.



Gambar 5. Hasil penerapan metode K-NN



Gambar 6. Hasil penerapan metode DTC

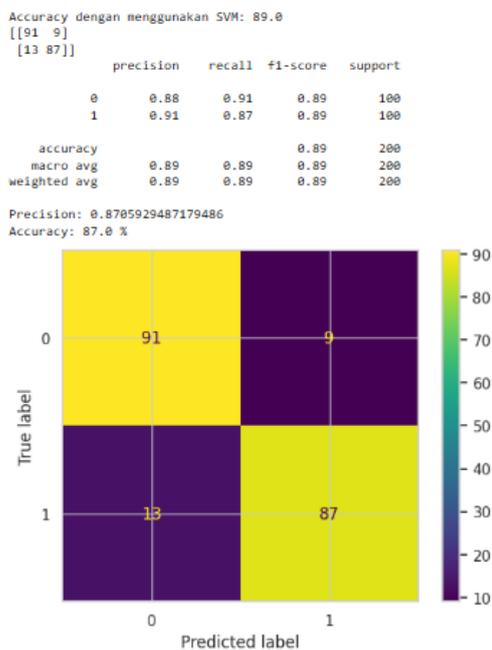
5. SVM

SVM merupakan algoritma machine learning yang digunakan untuk tugas klasifikasi dan regresi. Dalam bidang kesehatan, SVM telah

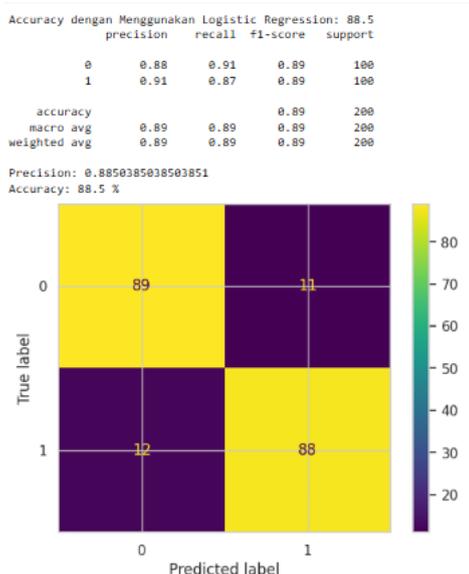
diterapkan untuk berbagai tujuan, termasuk diagnosis, prognosis dan prediksi hasil penyakit (Guido et al., 2024). Gambar 7 menunjukkan hasil penerapan algoritma SVM memperoleh akurasi sebesar 89%.

6. Logistic Regression

Logistic Regression adalah metode analisis statistik yang membangun model statistik untuk menggambarkan hubungan antara variabel hasil biner atau dikotomis (ya/tidak) sebagai variabel dependen atau respon dengan serangkaian variabel predictor atau variabel penjelas yang independen (Das, 2023). Hasil penerapan metode Logistic Regression pada penelitian ini sebesar 88.5%, sesuai dengan yang digambarkan pada Gambar 8.



Gambar 7. Hasil penerapan metode SVM



Gambar 8. Hasil penerapan metode Logistic Regression

Berdasarkan keenam model algoritma diatas hasil akurasi dari masing-masing algoritma digambarkan pada Gambar 9. Dari gambar tersebut, diketahui bahwa model dengan akurasi tertinggi adalah Random Forest (90%), diikuti oleh SVM (89,00%). Model dengan akurasi terendah adalah k-NN (79,50%). Logistic Regression dan Decision Tree memiliki nilai akurasi yang sama (88,50%). Secara keseluruhan semua model memiliki akurasi diatas 79%, yang menandakan performa yang cukup baik dengan perbedaan akurasi maksimal sebesar 10,5%.

3.5. Evaluation

Evaluasi dari model klasifikasi dilakukan dengan melihat nilai pada Akurasi, Presisi, Recall dan F1-Score. Berikut cara perhitungan Akurasi, Presisi, Recall dan F1-Score (Gargate, 2019).

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

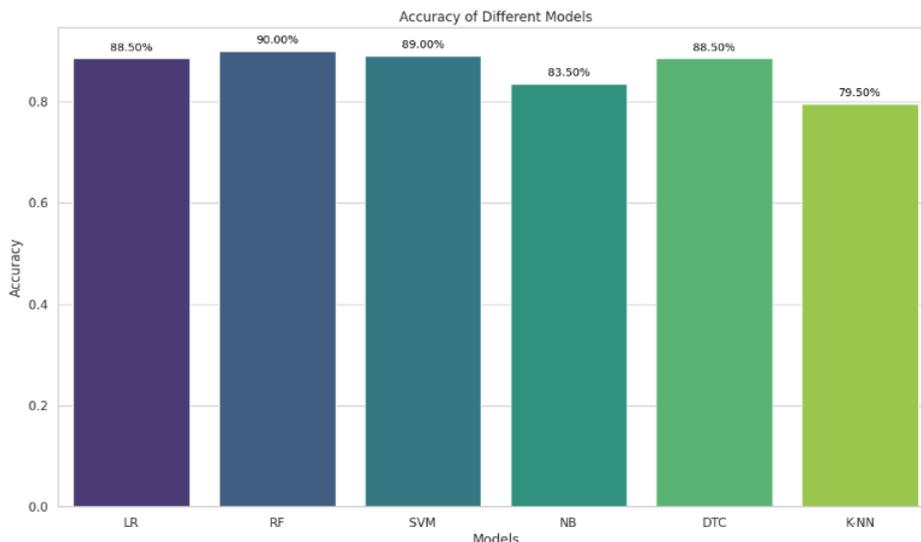
$$F1-Score = 2 \times \frac{precision \times recall}{precision+recall} \tag{4}$$

Berdasarkan uraian diatas berikut hasil dari perhitungan evaluasi dari setiap model algoritma.

Tabel 3 menunjukkan hasil evaluasi dari berbagai model machine learning dalam mendeteksi kanker paru-paru berdasarkan beberapa metric, yaitu True Positive (TP), False Negative (FN), False Positive (PF), True Negative (TN), Recall (Sensitivity), F1 Score dan Accuracy.

Algoritma Random Forest (RF), memiliki akurasi tertinggi (90%), yang berarti model ini memiliki kinerja terbaik dalam memprediksi secara keseluruhan, Decision Tree Classifier memiliki Recall (Sensitivity) tertinggi (0.5582), yang menunjukkan bahwa model ini lebih baik dalam menangkap kasus positif kanker dibandingkan dengan model lainnya. Logistis Regression dan SVM memiliki akurasi 89%, yang cukup tinggi dan hampir mendekati Random Forest, Naive Bayes memiliki akurasi rendah (83%), sedangkan K-NN memiliki akurasi paling rendah (80%), menunjukkan bahwa model ini kurang optimal dibandingkan dengan metode lainnya.

Berdasarkan tabel tersebut, Decision Tree Classifier merupakan model yang paling optimal untuk mendeteksi sebanyak mungkin kasus kanker paru-paru sedangkan Random Forest menjadi model pilihan terbaik untuk memprediksi penyakit kanker paru-paru karena memiliki nilai akurasi paling tinggi.



Gambar 9. Hasil akurasi perbandingan 6 model Algoritma Data Mining

Tabel 3. Hasil evaluasi model algoritma

Model	TP	FN	FP	TN	Recall (Sensitivity)	F1-Score	Accuracy
Naïve Bayes	246	209	235	210	0.540700	0.525600	83.00%
Random Forest	223	232	205	240	0.490100	0.505100	90.00%
Logistic Regression	239	216	232	213	0.525300	0.516200	89.00%
Decision Tree	254	201	226	219	0.558200	0.543300	85.00%
K-NN	231	224	228	217	0.507700	0.505500	80.00%
SVM	229	226	226	219	0.503300	0.503300	89.00%

4. Kesimpulan

Penelitian ini menggunakan data kanker paru-paru untuk membandingkan enam algoritma machine learning dengan evaluasi berbasis akurasi, recall dan F1-Score. Random Forest menunjukkan akurasi tertinggi (90%), sedangkan decision tree memiliki recall tertinggi, sehingga lebih efektif dalam mendeteksi kasus positif. Temuan ini menunjukkan bahwa pemilihan model harus disesuaikan dengan tujuan yaitu DTC untuk skrining awal dan RF untuk diagnosis menyeluruh. Penelitian ini bermanfaat sebagai dasar pengembangan sistem pendukung diagnosis kanker paru-paru berbasis machine learning. Namun, keterbatasan berupa cakupan data yang terbatas menjadi catatan penting. Penelitian selanjutnya disarankan untuk menggunakan data yang lebih beragam, termasuk data klinis dari rumah sakit, serta menerapkan algoritma optimasi guna meningkatkan akurasi dan keandalan model.

Daftar Pustaka:

Alfajr, N. H., & Defiyanti, S. (2024). *METODE RANDOM FOREST DAN PENERAPAN PRINCIPAL COMPONENT ANALYSIS (PCA)*. 12(3).

Alghifari, F., & Juardi, D. (2021). Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes. *Jurnal Ilmiah Informatika*, 9(02), 75–81. <https://doi.org/10.33884/jif.v9i02.3755>

Anzori, & Fatah, Z. (2024). *Jurnal Ilmiah*

Multidisiplin Nusantara Implementasi Metode Decision Tree Dalam Prediksi Kanker Paru Paru Dengan Rapidminer Jurnal Ilmiah Multidisiplin Nusantara. 2(November), 176–184.

Buana, I., & Harahap, D. A. (2022). Asbestos, Radon Dan Polusi Udara Sebagai Faktor Resiko Kanker Paru Pada Perempuan Bukan Perokok. *AVERROUS: Jurnal Kedokteran Dan Kesehatan Malikussaleh*, 8(1), 1. <https://doi.org/10.29103/averrous.v8i1.7088>

Das, A. (2023). *Logistic Regression*. Springer, Cham. https://doi.org/https://doi.org/10.1007/978-3-031-17299-1_1689

Datasets, D., Silfana, F. I., & Barata, M. A. (2024). *Using K-NN Algorithm for Evaluating Feature Selection on High*. 17(2).

Duran, P. A., Vitianingsih, A. V., Riza, M. S., Maukar, A. L., & Wati, S. F. A. (2024). Data Mining Untuk Prediksi Penjualan Menggunakan Metode Simple Linear Regression. *Teknika*, 13(1), 27–34. <https://doi.org/10.34148/teknika.v13i1.712>

Gargate, S. (2019). *Evaluating your classification model*. <https://medium.com/swlh/evaluating-your-classification-model-cb49338abb96>

Guido, R., Ferrisi, S., Lofaro, D., & Conforti, D. (2024). An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review. *Information (Switzerland)*, 15(4). <https://doi.org/10.3390/info15040235>

Herjanto, M. F. Y., & Carudin, C. (2024). Analisis

- Sentimen Ulasan Pengguna Aplikasi Sirekap Pada Play Store Menggunakan Algoritma Random Forest Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(2), 1204–1210. <https://doi.org/10.23960/jitet.v12i2.4192>
- Marzuq, R. D., Wicaksono, S. A., & Setiawan, N. Y. (2023). Prediksi Kanker Paru-Paru menggunakan Algoritme Random Forest Decision Tree. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 7(7), 3448–3456.
- Mienye, I. D., & Jere, N. (2024). A Survey of Decision Trees: Concepts, Algorithms, and Applications. *IEEE Access*, 12(May), 86716–86727. <https://doi.org/10.1109/ACCESS.2024.3416838>
- Naseh Khudori, A., & Syauqi Haris, M. (2024). Implementasi Decision tree Untuk Prediksi Kanker Paru-Paru. *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)*, 9(1), 94–106.
- Permana, A. Y., Fazri, H. N., Athoilah, M. F. N., Robi, M., & Firmansyah, R. (2023). Penerapan Data Mining Dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest. *Jurnal Ilmiah Teknik Informatika Dan Komunikasi*, 3(2), 27–41. <https://doi.org/10.55606/juitik.v3i2.472>
- Permana, B. A. C., & Djamaluddin, M. (2023). Penerapan Python Dalam Data Mining Untuk Prediksi Kanker Paru. *Infotek: Jurnal Informatika Dan Teknologi*, 6(2), 470–477. <https://doi.org/10.29408/jit.v6i2.17816>
- Rahmaeda, S., Prathivi, R., Studi, P., Informatika, T., Teknologi, F., & Semarang, U. (2025). *Komparasi Metode SVM dan Logistic Regression untuk Klasifikasi Hipotesa Penyakit Kanker Paru Paru Berdasarkan Gejala Awal Abstrak*. 6(1), 210–218.
- Ravinder, B., Seeni, S. K., Prabhu, V. S., Asha, P., Maniraj, S. ., & Srinivasan, C. (2024). Web Data Mining with Organized Contents Using Naive Bayes Algorithm. *2024 2nd International Conference on Computer, Communication and Control (IC4)*, 1–6.
- Rofiani, R., Oktaviani, L., Vernanda, D., & Hendriawan, T. (2024). Penerapan Metode Klasifikasi Decision Tree dalam Prediksi Kanker Paru-Paru Menggunakan Algoritma C4.5. *Jurnal Tekno Kompak*, 18(1), 126. <https://doi.org/10.33365/jtk.v18i1.3525>
- Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*, 2024, 69–79. <https://doi.org/10.58496/bjml/2024/007>
- Sandika, A., Ramadhan, F. R., Iman, I. N., & Jihad, J. (2024). Optimasi Prediksi Penyakit Paru-Paru dan Kanker Paru melalui Integrasi Algoritma Random Forest. 2(3), 585–591.
- Shafa, B., Handayani, H. H., Arum, S., & Lestari, P. (2024). Prediksi Kanker Paru dengan Normalisasi menggunakan Perbandingan Algoritma Random Forest, Decision Tree dan Naive Bayes. 4(3), 1057–1070.
- Wulandari, E. (2022). Klasifikasi Kanker Paru-Paru Menggunakan Metode Naive Bayes. *International Research on Big-Data and Computer Technology: I-Robot*, 6(2), 20–24. <https://doi.org/10.53514/ir.v6i2.325>

Halaman ini sengaja dikosongkan