Jurnal Ilmiah Wahana Pendidikan, Agustus 2022, 8 (12), 445-452
DOI: https://doi.org/10.5281/zenodo.6968420
p-ISSN: 2622-8327 e-ISSN: 2089-5364
Accredited by Directorate General of Strengthening for Research and Development
Available online at https://jurnal.peneliti.net/index.php/JIWP

Klasifikasi Penyakit Kangker Payudara Menggunakan Algoritma *K-Nearest Neighbor*

Widhi Ramdhani[1], David Bona[2], Rafi Bagus Musyaffa[3], Chaerur Rozikin[4]

[1,2,3,4] Teknik Informatika, Ilmu Komputer, Universitas Singaperbangsa Karawang

Received: 13 Juli 2022
Revised: 16 Juli 2022
Accepted: 19 Juli 2022

Abstract

Machine Learning (ML) merupakan salah satu cabang dari Artificial Intelligence (AI), yang berfokus pada pengembangan sistem yang dapat belajar sendiri tanpa pemrograman ulang. Machine learning ini banyak digunakan dalam berbagai bidang salah satunya di bidang kesehatan. Salah satu penyakit mematikan kedua saat ini adalah kangker. Kangker merupakan penyakit berbahaya yang bisa menimpa siapapun. Jenis kangker yang sering terjadi pada wanita salah satunya adalah kangker payudara. Dalam penelitian ini dilakukan pengklasifikasian penyakit kangker payudara dengan menerapkan metode klasifikasi menggunakan algoritma k-nearest neighbor. Hasil penelitian ini dapat memprediksi apakah terinfeksi kangker payudara jinak atau ganas. Dari algoritma KNN yang digunakan, hasil evaluasi performansi yang baik terdapat pada nilai k=21 dan k=11 dengan nilai akurasi sebesar 98%. Penelitian selanjutnya diharapkan dapat dikombinasikan dengan algoritma lainnya agar mendapatkan hasil yang lebih baik.

Keywords: *kangker payudara, machine learning, k-nearest neighbor*

(*) Corresponding Author: widhi.18127@student.unsika.ac.id,,david.bona18076@student.unsika.ac.id, rafi.bagus18202@student.unsika.ac.id, cr@staff.unsika.ac.id

How to Cite: Ramdhani, W., Bona, D., Musyaffa, R., & Rozikin, C. (2022). Klasifikasi Penyakit Kangker Payudara Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Ilmiah Wahana Pendidikan*, *8*(12), 445-452. https://doi.org/10.5281/zenodo.6968420

PENDAHULUAN

Machine Learning (ML) merupakan salah satu aplikasi *Artificial Intelligence* (AI), dan fokusnya adalah pengembangan sistem yang dapat belajar sendiri tanpa pemrograman ulang. ML membutuhkan data (*data training*) sebagai proses pembelajaran sebelum menghasilkan hasil. ML sudah dapat digunakan dalam hampir di semua bidang, salah satunya bidang Kesehatan. Pendeteksian kanker merupakan salah satu contoh penerapan ML di bidang Kesehatan.

Kangker merupakan penyakit yang banyak diderita oleh manusia dan menjadi penyebab kematian kedua dan di prediksi dapat mengalahkan penyakit jantung sebagai penyebab kematian nomor 1 saat ini (Kurniasari, 2017). Salat satu jenis kangker yang sering terjadi pada wanita yaitu kanker payudara. Kangker payudara merupakan penyakit non kulit berbahaya yang tumbuh di dalam jaringan payudara yang menjadi penyebab kematian kedua terbesar setelah kangker paru-paru. Penyakit ini disebabkan oleh berbagai faktor, mulai dari sel kelenjar dan saluran hingga jaringan penyangga payudara, kecuali kulit payudara. Menurut WHO (*World Healt Organization)*, 8 hingga 9% wanita beresiko terkena penyakit kangker payudara. Di Indonesia kurang lebih terdapat 100 penderita baru per 100.000 penduduk di setiap tahunnya (Angrainy, 2017).

Penderita kanker payudara biasanya memiliki ciri psikologis seperti gegar otak, ketakutan, depresi, dan panik. Penyebab kanker payudara masih belum pasti, namun diduga multifaktorial. Berbagai upaya telah dilakukan, seperti upaya pencegahan seperti mengedukasi masyarakat, dan pembenahan yang diperlukan untuk mengatasi masalah kanker payudara sesuai kebutuhan masing-masing. Skrinning kangker payudara sejak dini merupakan hal penting yang wajib dilakukan oleh setiap orang yang bisa dilakukan sendiri disebut juga dengan SADARI (Pemeriksaan payudara sendiri). Namun jika sudah didiagnosa terkena kangker payudara maka perlu dilakukan pemeriksaan lebih lanjut ke dokter.

Pemeriksaan untuk mendeteksi kangker selain dilakukan mandiri bisa dengan pemeriksaan mammografi atau bisa denan MRI, USG, atau biopsi (www.alodoktor.com). Cara tersebut banyak memakan waktu dalam pendeteksian, maka dari itu untuk dapat membantu dokter, pada penelitian ini akan membuat suatu pembelajaran mesin untuk mengklasifikasikan penyakit kangker payudara.

Penelitian sebelumnya yang dilakukan oleh (Athalla, 2020) mengenai Klasifikasi Penyakit Kanker Payudara dengan Metode K-NN untuk mengklasifikasi dataset. Penelitian ini menggunakan perhitungan jarak kemiripan menggunakan jarak minkowski yang telah di tentukan dari kedua jarak Euclidean dan jarak Manhattan. Penelitian ini mendapatkan hasil dengan tingkat akurasi sebesar 93%.

Penelitian lain yang dilakukan oleh (Chazar, 2020) mengenai diagnosis kanker payudara dengan menggunkanan Algoritma *Support Vector Machine* yang hanya mendapatkan hasil tinggat akurasi sebesar 88,733%. Penelitian yang lain dari (Tiana, 2020) dengan menggunakan Algoritma lain yaitu *Naïve Bayes* hanya mendapatkan hasil nilai akurasi sebesar 72,7%.

Hingga saat ini terdapat beberapa algoritma ML yang dapat digunakan dan dikembangkan untuk berbagai tujuan. Salah satunya penelitian ini yang bertujuan nuntuk membangun sebuah aplikasi ML dengan menggunakan algoritma K-NN yang dapat digunakan untuk mengklasifikasi penyakit kanker payudara. Metode KNN memiliki beberapa keunggulan, yaitu pelatihan yang sederhana, cepat, mudah dimengerti, dan efektif apabila ukuran data pelatihan besar. Namun, KNN ini juga terdapat kelemahan, yaitu nilai K yang bias (berbeda). Hasil klasifikasi dapat menghasilkan suatu prediksi penentuan jenis sel kanker payudara bersifat ganas atau jinak.

LANDASAN TEORI

Machine Learning

Machine Learning merupakan bagian dari kecerdasan buatan yang memungkinkan komputer dapat cerdas dan berperilaku seperti manusia dengan melakukan pembelajaran mesin untuk meningkakan pemahaman sistem yang dapat belajar otomatis (Restoningsih, 2020). *Machine learning* dapat belajar dari inputan data sebagai *data training* untuk melatih algoritma dalam *machine learning* dan kemudian dilakukan analisis terhadap sekumpulan data (*big data*) sehingga dapat menemukan pola tertentu.

Machine learning dapat dibedakan menjadi dua tipe teknik pembelajaran, yaitu *supervised learning* dan *unsupervised learning. Supervised learning* merupakan teknik dalam *machine learning* yang menggunakan dataset berlabel yang kemudian melakukan pembelajaran mesin dan mesin dapat mengidentifikasi

label yang diinputkan dengan fitur. Algotitma yang termasuk kedalam *supervised learning* adalah *K-Nearest Neighboor (KNN), Decision Tree, Naive Bayes, Super Vector Machine (SVM),* dan Regresi. *Unsupervised learning* merupakan salah satu teknik *machine learning* yang dimana menarik kesimpulkan berdasarkan dataset input. Algoritma yang termasuk kedalam *unsupervised learning* adalah *Fuzzy C-Means, K-Means, DBSCAN,* dan *Self Organizing Map (SOM).*

Algoritma KNN

Algoritma KNN merupakan algoritma yang banayak digunakan dalam melakukan klasifikasi. KNN merupakan algoritma yang sederhana untuk diimplementasikan tetapi menghasilkan akurasi yang baik (Wahyono, 2020). Salah satu kelemahan dari algoritma ini adalah dalam penentuan nilai k, jika nikai k terlalu besar maka akan membuat hasil klasifikasi menjadi tidak jelas atau kabur, sedangkan jika nilai k terlalu kecil atau di misalkan k=1 maka akan menyebabkan hasil klasifikasi terasa kaku karena tidak ada pilihan (Indrayanti, 2017). Maka dari itu diperlukan penelitian penentuan nilai k yang baik.

METODOLOGI

ataset

Dataset pada penelitian ini merupakan dataset tentang kanker payudara yang diambil dari kaggle. Dataset ini memiliki 569 object dari 32 variabel.

Pada 32 variabel terdapat atribut id, diagnosis, radius, tekstur, perimeter, area, kelembutan, kepadatan, kecekungan, titik cekung, simetri dan faktral dimensi yang masing-masing memiliki nilai rata-rata, nilai error dan nilai terburuk yang nilainya bervariasi.

Menghitung Jarak Kemiripan (K)

Perhitungan jarak kemiripan pada penelitian ini dilakukan dengan menggunakan perhitungan minkowski. Jarak minkowski merupakan generalisasi dari jarak Euclidean dan jarak Manhattan dalam ruang vektor (Atthalla, 2018).

Jarak minkoswki digunakan untuk menghitung jarak dua atau lebih vektor (x,y), yang dimana didalamnya terdapat variasi nilai p. Seperti pada persamaan 1.

$$d(x,y) = (\sum^{n}_{i=1}|x_i\text{-}y_i|^{p})^{1/p} \qquad (1)$$

Perhitungan jarak minkowski meliputi *Manhattan* dan *Euclidian distance* (p=1), seperti pada persamaan 2.

$$D(x,y) = \sum^{n}_{i=1}|x_i\text{-}y_1\} \qquad (2)$$

Euclidian distance (p=2), seperti persamaan 3 berikut.

$$d(x,y) = \sqrt{\sum^{n}_{k=1}(x_i - y_1)^2} \qquad (3)$$

Klasifikasi KNN

Dataset pada penelitian ini merupakan dataset tentang kanker payudara yang diambil dari kaggle. Dataset ini memiliki 569 object dari 32 variabel.

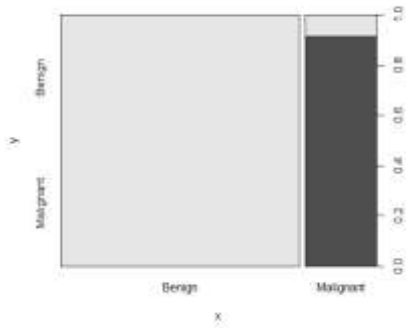
Pada 32 variabel terdapat atribut id, diagnosis, radius, tekstur, perimeter, area, kelembutan, kepadatan n, kecekungan, titik cekung, simetri dan faktral

dimensi yang masing-masing memiliki nilai rata-rata, nilai error dan nilai terburuk yang nilainya bervariasi.

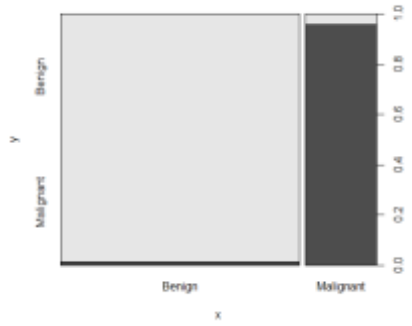
HASIL DAN PEMBAHASAN

Penelitian ini menggunakan metode klasifikasi KNN dengan software Rstudio. Setelah mendapatkan dataset, kemudian dilakukan tahapan preprocessing data dengan melakukan normalisasi data kemudian dilakukan pembagian data dengan perbandingan 80% untuk data training dan 20% untuk data testing.

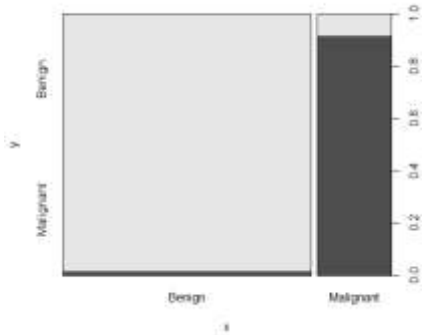
Proses data mining yang dilakukan pada penelitian ini berawal dari pemodelan menggunakan data training kemudian menerapkannya ke data testing dengan menggunakan algoritma K-Nearest Neighbor. Dari pemodelan tersebut kemudian didapatkan hasil akhir berupa nilai akurasi dari pemodelan yang digunakan sebagai berikut.



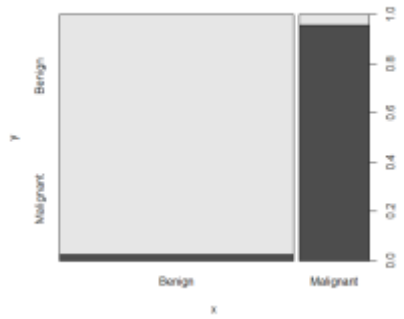
Gambar 1. Grafik pengujian performa metode KNN dengan nilai k=21



Gambar 2. Grafik pengujian performa metode KNN dengan nilai k=11



Gambar 3. Grafik pengujian performa metode KNN dengan nilai k=33



Gambar 4. Grafik pengujian performa metode KNN dengan nilai k=9

Hasil evaluasi performa dari algoritma KNN ini dilihat pada gambar 5 dengan nilai k=21 didapatkan nilai akurasi sebesar 98%.

Total Observations in Table: 100

wbcd_test_labels	wbcd_test_pred: Benign	wbcd_test_pred: Malignant	Row Total
Benign	77 1.000 0.975 0.770	0 0.000 0.000 0.000	77 0.770
Malignant	2 0.087 0.025 0.020	21 0.913 1.000 0.210	23 0.230
Column Total	79 0.790	21 0.210	100

Gambar 5.
Hasil pengujian performa KNN (k=21)

Untuk nilai k=11 didapatkan nilai akurasi sebesar 98% dapat dilihat pada tabel 6 berikut.

Total Observations in Table: 100

wbcd_test_labels	wbcd_test_pred: Benign	wbcd_test_pred: Malignant	Row Total
Benign	76 0.987 0.987 0.760	1 0.013 0.043 0.010	77 0.770
Malignant	1 0.043 0.013 0.010	22 0.957 0.957 0.220	23 0.230
Column Total	77 0.770	23 0.230	100

Gambar 6.
Hasil pengujian performa KNN (k=11)

Kemudian pada gambar 7, hasil performansi knn dengan nilai k=33 didapatkan nilai akurasinya sebesar 97%.

Total Observations in Table: 100

wbcd_test_labels	wbcd_test_pred: Benign	Malignant	Row Total
Benign	76 0.987 0.974 0.760	1 0.013 0.045 0.010	77 0.770
Malignant	2 0.087 0.026 0.020	21 0.913 0.955 0.210	23 0.230
Column Total	78 0.780	22 0.220	100

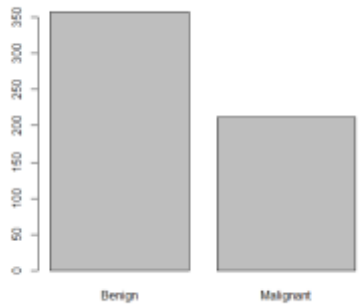
Gambar 7.
Hasil pengujian performa KNN (k=33)

Dari tabel 8 menunjukkan hasil performa algoritma KNN dengan nilai k=9 didapatkan nilai akurasinya sebesar 97% .

Total Observations in Table: 100

wbcd_test_labels	wbcd_test_pred: Benign	Malignant	Row Total
Benign	73 0.974 0.987 0.750	2 0.026 0.083 0.020	77 0.770
Malignant	1 0.043 0.013 0.010	22 0.957 0.917 0.220	23 0.230
Column Total	76 0.780	24 0.240	100

Gambar 8.
Hasil pengujian performa (k=9)



Gambar 9.
Perbandingan hasil klasifikasi kangker payudara jinak dan ganas.

Pada gambar 9 merupakan grafik jumlah dari hasil prediksi diagnosa kangker payudara dari dataset yang dianalisis, yang menunjukkan jumlah klasifikasi diagnosa kangker payudara jinak lebih banyak dengan jumlah 369, dibandingkan dengan kangker payudara ganas sebanyak 200.

KESIMPULAN DAN SARAN

Berdasarkan dari penelitian pengklasifikasian kangker payudara dengan algoritma KNN dapat ditarik kesimpulan bahwa dari penelitian ini didapatkan prediksi jenis kangker, apakah termasuk kedalam kategori kangker jinak atau

kangker ganas. Nilai persentasi pengklasifikasian penyakit kangker payudara dengan persentase 62,7% dalam kategori kangker jinak dan 37,3% dalam kategori kangker ganas. Selain itu, dilakukan evaluasi atau pengujian nilai performa dari pemodelan algoritma KNN yang digunakan dengan melakukan beberapa percobaan nilai K. Pengujian dengan nilai akurasi tertinggi diperoleh dari nilai k=21 dan k=11 dengan nilai akurasi mencapai 98%.

Saran untuk penelitian selanjutnya dapat diterapkan beberapa metode atau algoritma yang digunakan dan dikombinasikan dengan algoritma yang diterapkan pada penelitian ini agar mendapatkan hasil yang lebih baik.

DAFTAR PUSTAKA

Agarap, A. F. M. (2018, February). *On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset. In Proceedings of the 2nd International Conference on Machine Learning and Soft Computing* (pp. 5-9).

Al Bataineh, A. (2019). A comparative analysis of nonlinear machine learning algorithms for breast cancer detection. *International Journal of Machine Learning and Computing*, *9*(3), 248-254.

Angrainy, R. (2017). Hubungan Pengetahuan, Sikap Tentang Sadari Dalam Mendeteksi Dini Kanker Payudara Pada Remaja. *Jurnal Endurance, 2*(2), 232. https://doi.org/10.22216/jen.v2i2.1766

Assegie, T. A. (2020). An optimized K-Nearest Neighbor based breast cancer detection. *Journal of Robotics and Control (JRC)*, *2*(3), 115-118.

Atthalla, I. N., Jovandy, A., & Habibie, H. (2019). Klasifikasi Penyakit Kanker Payudara Menggunakan Metode K Nearest Neighbor. In Annual Research Seminar (ARS) (Vol.4, No.1, pp. 148-151).

Azis, A. I., Idris, I. S. K., Santoso, B., & Mustofa, Y. A. (2019). Pendekatan Machine Learning yang Efisien untuk Prediksi Kanker Payudara. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 3(3), 458-469.

Chazar, C., & Erawan, B. (2020). Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine. INFORMASI (Jurnal Informatika dan Sistem Informasi), 12(1), 67-80.

Indrayanti, I., Sugianti, D., & Al Karomi, A. (2017). Optimasi Parameter K Pada Algoritma K-Nearest Neighbour Untuk Klasifikasi Penyakit Diabetes Mellitus. Prosiding SNATIF, 823-829.

Kurniasari, F. N., Harti, L. B., Ariestiningsih, A. D., Wardhani, S. O., & Nugroho, S. (2017). Buku Ajar Gizi dan Kanker. Universitas Brawijaya Press.

Madyaningrum, N. A., & Sulastri. (2019). Analisa Prediksi Kekambuhan Kanker Payudara Dengan Menggunakan Algoritma K-Nearest Neighbor. Proceeding SINTAK, 180–185.

Retnoningsih, E., & Pramudita, R. (2020). Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. BINA INSANI ICT JOURNAL, 7(2), 156-165.

Tiana, E., & Wahyuni, S. (2020). Hasil Analisis Teknik Data Mining dengan Metode Naive Bayes untuk Mendiagnosa Penyakit Kanker Payudara. *Jurnal Sistem Komputer dan Informatika (JSON)*, *1*(2), 130-133.

Wahyono, W. (2020). Peningkatan Kecepatan Algoritma k-NN Untuk Sistem Pengklasifikasian Kendaraan Bermotor. Techno. Com, 19(2), 190-196.

www.alodokter.com. 2021. Mendeteksi Kangker Sejak Dini. Di https://www.alodokter.com/mendeteksi-kanker-sejak-dini (di akses 15 januari)