Teknika 17 (2): 427 – 434

JURNAL TEKNIKA
ISSN: 0854-3143
e-ISSN: 2622-3481
Journal homepage: http://jurnal.polsri.ac.id/index.php/teknika
Journal Email: teknika@polsri.ac.id

Deteksi Dini Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor (KNN) Dan Decision Tree C-45

Fahrurrozi[1], Wasilah*[2]

[1,2]IIB Darmajaya Lampung; ,JL.ZA Pagar Alam No 93 Gedung Meneng.
Telp 0721-787214
*Email Penulis Korespondensi: fahrurrozyghalib@gmail.com

Abstrak

Kanker payudara merupakan tipe kanker yang umumya terbentuk di sel-sel payudara dan sel-sel kanker tersebut tumbuh diluar kendali. Kanker payudara dapat terjadi pada semua gender. Di Indonesia, jumlah kasus kanker payudara sampai menempati urutan pertama dibandingkan jenis kanker yang lain dan menjadi salah satu penyumbang kematian pertama. Berdasarkan jumlah kasus kematian tersebutdan mengingat kanker payudara tidak memandang gender, seharusnya baik pria maupun wanita sadar dengan kesehatan mereka dengan cara melakukan tindakan seperti deteksi dini dan menghindari risiko yang menyebabkan terjadinya kanker. Data yang digunakan dalam penelitian ini berasal dari https://www.kaggle.com/datasets/. Tujuan dari penelitian ini yaitu memanfaatkan beberapa algoritma data mining yang ada dan membandingkan dua algoritma data mining dalam melakukan klasifikasi terhadap kanker payudara. Pada penelitian ini algoritma yang digunakan dalam melakukan perbandingan adalah algoritma Decision Tree, dan K-Nearest Neighbors . penulis mengkobinasikan antara algoritma Decision Tree classifier yang memiliki kemampuan baik untuk mengolah database yang besar sebagai feature selection kemudian dengan algoritma K-Nearest Neighbors (KNN) yang layak dan relevan digunakan dalam menganalisis dan mengdiaknosa Kanker .Hasil dari pengujian menunjukkan bahwa algoritma K-Nearest Neighbors menghasilkan performa akurasi 94,73 sedangakan algoritma Decision Tree 96,49% nilai akurasi yang terbaik yaitu sama, yaitu Decision Tree mendapatkan nilai akurasi 96,49%.

Kata kunci—Kanker payudara, Decision Tree, KNN

Abstract

Breast cancer is a type of cancer that commonly forms in breast cells, and these cancer cells grow uncontrollably. Breast cancer can occur in all genders. In Indonesia, the number of breast cancer cases ranks first compared to other types of cancer and is one of the leading causes of death. Based on the number of death cases and considering that breast cancer does not discriminate by gender, both men and women should be aware of their health by taking actions such as early detection and avoiding risk factors that can lead to cancer. The data used in this research comes from https://www.kaggle.com/datasets/. The aim of this study is to utilize various data mining algorithms and compare two data mining algorithms in classifying breast cancer. In

this study, the algorithms used for comparison are the Decision Tree algorithm and the K-Nearest Neighbors algorithm. The test results show that the K-Nearest Neighbors algorithm produces an accuracy performance of 94.73%, while the Decision Tree algorithm achieves the highest accuracy score of 96.49%.

Keywords—*Breast cancer, Decision Tree, KNN*

1. PENDAHULUAN

Kanker menjadi salah satu jenis penyakit berbahaya penyebab terjadinya kematian pada manusia di seluruh dunia [1] Kanker payudara merupakan tumor dengan insiden tinggi yang mengancam kesehatan wanita secara serius [2]. Kanker payudara menempati urutan pertama terkait jumlah kanker terbanyak di Indonesia, serta menjadi salah satu penyumbang kematian pertama akibat kanker. Data Globocan tahun 2020 jumlah kasus kanker dunia mencapai total 19.292.789 kasus, untuk kasus kanker payudara berjumlah 2.261.419 (11.7%). Untuk kasus kematian yang diakibatkan oleh penyakit kanker total 9.958.133, sedangkan kasus kematian yang diakibatkan oleh kanker payudara berjumlah 684.996 menyumbang (6.9%) dari kasus kematian yang diakibatkan oleh penyakit kanker[3].

Data Globocan tahun 2020, jumlah kasus baru kanker payudara mencapai 65.858 kasus (16,6%) dari total 396.914 kasus baru penyakit kanker di Indonesia. Sementara itu, untuk jumlah kematiannya mencapai lebih dari 22 ribu jiwa kasus [4], [5]. Kanker payudara salah satu jenis kanker yang paling umum terjadi pada perempuan di seluruh dunia.

Kanker payudara ini secara umum dibagi menjadi 2,yaitu benign atau biasa disebut jinak dan malignant atau biasa disebut juga ganas, biasanya kanker payudara jinak ditandai dengan berbentuk benjolan kecil bulat, dan lembut[6]. Kanker payudara dalam tingkat jinak biasanya akan mempunyai keadaan dan pertumbuhan yang tidak bersifat kanker. Kanker ini bisa terdeteksi tetapi tidak akan menjalar dan merusak jaringan di dekatnya[6]. Pada kanker payudara dalam tingkat ganas ditandai dengan bentuk yang tidak simetris, kasar, terasa nyeri, dan lainnya[7]. Biasanya kanker payudara menjalar dan merusak jaringan dan organ lain yang ada di dekatnya[6].

Berdasarkan data Riskesdas, prevalensi tumor/kanker di Indonesia menunjukkan adanya peningkatan dari 1.4 per 1000 penduduk di tahun 2013 menjadi 1.79 per 1000 penduduk pada tahun 2018. Sedangkan angka kejadian untuk perempuan yang tertinggi adalah kanker payudara yaitu sebesar 42,1 per 100.000 penduduk dengan rata-rata kematian 17 per 100.000 penduduk. Kanker payudara merupakan salah satu jenis kanker yang paling sering terjadi pada wanita. Menurut data WHO, pada tahun 2020 terdapat sekitar 2,3 juta kasus baru kanker payudara di seluruh dunia [8]. Meskipun demikian, angka kesembuhan kanker payudara juga cukup tinggi jika dideteksi pada stadium awal dan diberikan penanganan yang tepat. Oleh karena itu, deteksi dini sangat penting dalam upaya pencegahan dan pengobatan kanker payudara.

Deteksi dini kanker payudara memiliki peran penting dalam meningkatkan prognosis dan tingkat kelangsungan hidup pasien. Hal ini bertujuan untuk mengurangi kematian akibat kanker global sebesar 2,5% per tahun, sehingga mencegah 2,5 juta kematian akibat kanker payudara di seluruh dunia antara tahun 2020 dan 2040[9] Saat ini teknologi dan metode komputasional telah berkembang pesat, yang memungkinkan penggunaan algoritma pembelajaran mesin dapat mendukung proses deteksi dini kanker payudara. Salah satu metode yang dapat digunakan untuk mendeteksi kanker payudara adalah dengan menggunakan data mining. Data mining merupakan salah satu teknik dalam bidang ilmu komputer yang memungkinkan kita untuk mengeksplorasi data secara otomatis dengan menggunakan algoritma dan teknik yang khusus

Penelitian-penelitian terdahulu dalam melakukan deteksi dan klasifikasi serta mendiagnosa kejadian kanker payudara diantaranya dengan Komparasi Algoritma Decision Tree, Naive Bayes, dan K-Nearest Neighbors dalam Klasifikasi Kanker Payudara. Hasil dari penelitian tersebut menghasilkan KNearest Neighbors dengan performa akurasi yang sangat baik dibanding algoritma Naive Bayes dan Decision Tree, yaitu 98% pada metode Hold-Out dan 96% pada

metode K-Fold [10]. Kemudian pada penelitian Data Mining Techniques in Predicting Breast Cancer juga memberikan hasil yang baik: KNN menghasilkan nilai akurasi = 92.31%, dan akurasi SVM = 95,65% [11]. Penelitian Analysis of Decision Tree and K-Nearest Neighbor Algorithmin the Classification of Breast Cancer menghasilkan performa dalam nilai akurasi Decision-Tree 91,23 % dan nilai akurasi KNN 95,61 % [12].

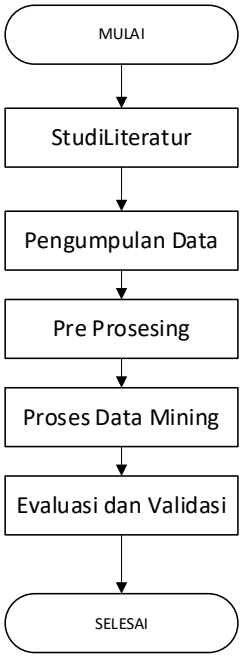
Hasil penelitian lainnya yaitu komparasi tingkat akurasi Random Forest dan KNN untuk mendiagnosis penyakit kanker payudara akurasi dari kedua model tersebut sebesar KNN 96,5% dan Random Forest sebesar 97,2% [13]. Meskipun algoritma KNN dan Decision Tree C4.5 telah berhasil diterapkan dalam berbagai domain, berdasarkan penelitian sebelumnya penelitian ini bertujuan untuk menggali potensi kedua algoritma tersebut dalam deteksi dini kanker payudara dan membandingkan performa keduanya.

2. METODE PENELITIAN

Penelitian ni dimulai dengan memperkenalkan skema penelitian dan alat yang digunakan dalam penelitian ini, yang merupakan alat Machine Learning (ML) open-source yang siap pakai. Selain itu, dijelaskan pula mengenai dataset yang digunakan dalam penelitian, Penelitian ini bertujuan untuk membandingkan performa antara algoritma K-Nearest Neighbors (KNN) dan algoritma C4.5 (C45) dalam klasifikasi kanker payudara menggunakan Rapidminer. Dalam bab ini, akan dibahas mengenai rancangan penelitian, data dan sumber data yang digunakan, teknik pengolahan data, serta analisis data untuk mendapatkan hasil penelitian yang akurat [13].

2.1 Skema Penelitian

Pada bab ini akan membahas langkah-langkah dari proses penelitian yang akan dilaksanakan. Dalam melakukan analisa dan mencari pola data untuk dijadikan sebuah dataset dalam memudahkan penelitian dan dapat berjalan dengan sistematis dan memenuhi tujuan, maka dibuat alur dalam tahapan penelitian yang akan dilakukan sebagai berikut:



Gambar 1 Alur Penelitian

2.2 Studi Literature

Pada tahap ini peneliti melakukan review jurnal yang berkaitan dengan judul penelitian yang akan diteliti oleh peneliti. Jurnal-jurnal yang menjadi rujukan tentu memiliki ruang lingkup dan metode yang sama dilakukan oleh peneliti

2.3 Pengumpulan Dataset

Data yang digunakan merupakan dataset publik berupa data yang diperoleh dari Kaggle dengan link https://www.kaggle.com/datasets/uciml/breastcancer-wisconsin-data. yang terdiri dari informasi demografis, kebiasaan, dan catatan medis historis. Beberapa pasien memutuskan untuk tidak menjawab beberapa pertanyaan karena masalah privasi (missing value). Jumlah data adalah 570 data dengan 32 atribut.

2.4 Pre-processing dan Pengolahan Data (Cleaning Data).

Preprocessing dan pengolahan data dilakukan pada Rapidminer dengan melakukan normalisasi, penghapusan fitur yang tidak berguna, dan treatment data yang hilang. Data berjumlah 570 record data dengan 32 atribut

2.5 Penerapan Data Mining

A. Algoritma K-Nearest Neighbor (K-NN)

Penerapan Algoritma K-Nearest Neighbor (K-NN) Penerapan algoritma K-Nearest Neighbor (KNN) dalam klasifikasi kanker payudara. Formula 1 merupakan rumus yang digunakan dalam perhitungan K-NN.

$$W_i = \frac{1}{d(x', x_i)}$$

$$y\ i = \arg\ max \sum(xi\ ,y2) \in Dzwi\ X\ (V = yi) \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

B. algoritma *Decision Tree*

Pada bagian ini, akan dilakukan penerapan algoritma *Decision Tree* dengan menggunakan metode C4.5 untuk menklasifikasi kanker payudara. Algoritma C4.5 merupakan salah satu algoritma yang sering digunakan dalam data mining untuk melakukan klasifikasi atau pengelompokan berdasarkan aturan-aturan yang dihasilkan dari pohon keputusan[14]. Formula 2 merupakan perhitungan Gain dan formula 3 merupakan perhitungan entropy.

$Gaint\ (S.A\ ) = Entropy(S)\frac{Si}{S}$* Entropy (Si) …… (2)

Keterangan:
S: Himpunan kasus
A: Data Atribut
n: Jumlah partisi di dalam atribut
|Si|: Jumlah kasus pada partisi ke-i
|S|: Jumlah kasus

$Entropy\ (S) = \sum_{pi}^{n}$ -pi=$\log_2$ pi …… (3)

Keterangan:
S: Himpunan kasus
n: Jumlah partisi dalam atribut
pi: Proposi dari Si terhadap S

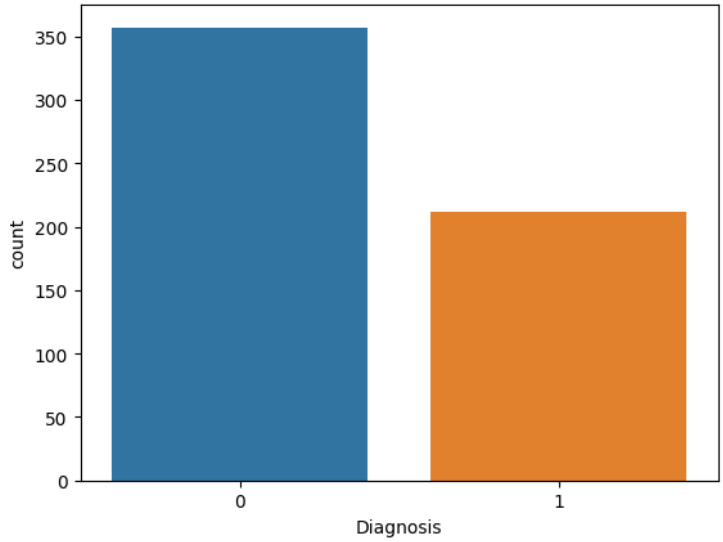
3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menerapkan algoritma KNN dan Decision Tree. Hasil dari penelitian ini berupa pengolahan data kualitatif dan data kuantitatif dengan perhitungan yang dilakukan pada sebuah dataset yang telah diperoleh. Yang sudah melalui proses preprocessing data dan analisis data.

Pada bagian hasil, penulis akan menjelaskan hasil dari pengujian terhadap model yang diusulkan. Pengujian dilakukan sesuai dengan tahapan yang telah dijelaskan dalam bagian metode penelitian. Setelah melalui proses preprocessing, dilakukan analisis perbandingan menggunakan algoritma KNN dan Decision Tree.

A. Algoritma KNN

Gambar 2 merupakan gambaran umum dari dataset. Dari grafik visualisasi yang ditampilkan,dapat dilihat dibawah ini:



Gambar 2 Grafik Diagnois Algoritma KNN.

Pada penggunaan Algoritma KNN, nilai akurasinya berada di angka 94,73% seperti pada Gambar 3.

```
Confusion Matrix:
[[68  3]
 [ 3 40]]

Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.96      0.96        71
           1       0.93      0.93      0.93        43

    accuracy                           0.95       114
   macro avg       0.94      0.94      0.94       114
weighted avg       0.95      0.95      0.95       114


Accuracy Score: 0.9473684210526315
```

Gambar 3 Akurasi Algoritma KNN

B. Decision Tree

Pada penggunaan metode Decision Tree, nilai akurasinya berada di angka 96,49% seperti pada Gambar 4.

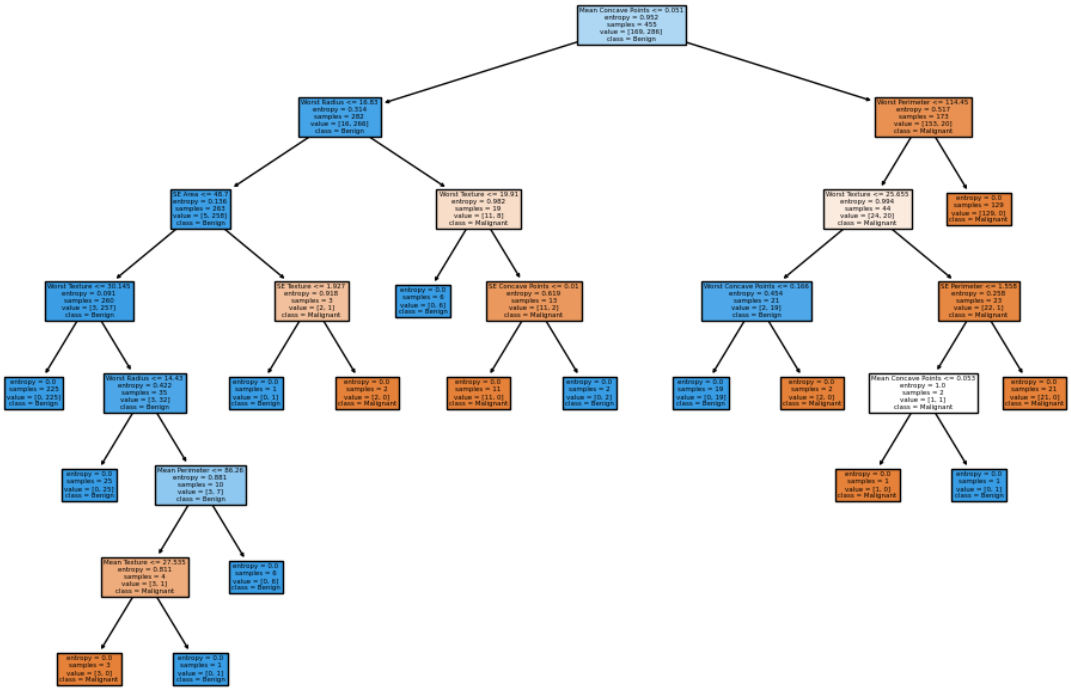
```
Accuracy: 0.9649122807017544
Confusion Matrix:
 [[39  4]
 [ 0 71]]
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.91      0.95        43
           1       0.95      1.00      0.97        71

    accuracy                           0.96       114
   macro avg       0.97      0.95      0.96       114
weighted avg       0.97      0.96      0.96       114
```

Gambar 4 Akurasi Algoritma Decision Tree

Pada penggunaan metode *Decision Tree*, Model Pohon Keputusan seperti pada Gambar 5.



Gambar 5 Model Pohon Keputusan Algoritma C4.5

C. Evaluasi dan Validasi

Penelitian ini bertujuan untuk melihat akurasi prediksi Penyakit Kanker Payudara dibandingkan dengan menggunakan algoritma C4.5,dan K-NN , kemudian menganalisa akurasi dengan membandingkan kedua metode tersebut.

4. KESIMPULAN

Berdasarkan uraian uraian padapembahasan sebelumnya, dapat ditarik kesimpulan bahwa klasifikasi kanker payudara menggunakan algoritma C4.5 yaitu sebesar 96,49% sedangkan menggunakan KNN 94,73%. Proses imputasi tidak terlalu berdampak pada hasil hasil akurasi pada algoritma C4.5, dan algoritma KNN yaitu nilai akurasi yang lebih baik mendaptkan nilai 96,49% yaitu algoritma C4.5.

5. SARAN

Pada kumpulan data, tidak terdapat kesalahan sehingga data asli yang digunakan tetap terjaga. Selain itu, masih terdapat peluang untuk memproses ulang data tersebut guna menghasilkan hasil yang lebih optimal. Selain itu, belum dilakukan perbandingan tingkat akurasi dengan algoritma lain menggunakan dataset yang serupa. Untuk meningkatkan akurasi, dapat dilakukan penyesuaian parameter sampling secara linier sesuai dengan karakteristik dataset.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Tim Redaksi Jurnal Teknika Politeknik Negeri Sriwijaya yang telah memberi memberi kesempatan, sehingga artikel ilmiah ini dapat diterbitkan.

DAFTAR PUSTAKA

[1] M. A. K. Neighbor, N. Meilani, And O. Nurdiawan, “Data Mining Untuk Klasifikasi Penderita Kanker Payudara,” Vol. 2, No. 1, Pp. 177–187, 2023.

[2] S. Tang *Et Al.*, “Clinical And Pathological Response To Neoadjuvant Chemotherapy With Different Chemotherapy Regimens Predicts The Outcome Of Locally Advanced Breast Cancer,” *Gland Surg*, Vol. 9, No. 5, Pp. 1415–1427, Jan. 2020, Doi: 10.21037/Gs-20-209.

[3] The Global Cancer Observatory, “Kasus Kanker Payudara Dunia”, Doi: 10.8.

[4] Kemenkes Ri, “Kanker Payudara Paling Banyak Di Indonesia,” 2023. Https://Www.Kemkes.Go.Id/Article/View/22020400002/Kanker-Payudara- Paling-Banyak-Di-Indonesia-Kemenkes-Targetkan-Pemerataan-Layanan-Kesehatan.Html

[5] Globocan, “Kasus Kanker Payudara Indonesia.” Accessed: Jul. 10, 2023. [Online].Available:Https://Gco.Iarc.Fr/Today/Data/Factsheets/Populations/360-Indonesia-Fact-Sheets.Pdf

[6] B. A. Farahdiba, D. Yusuf, And S. Nugroho, “Klasifikasi Kanker PayudaraMenggunakan Algoritma Gain Ratio.”

[7] Y. Ireaneus, A. Rejani, And S. T. Selvi, “Early Detection Of Breast CancerUsing Svm Classifier Technique,” 2009.

[8] “Penyakit Kanker Di Indonesia Berada Pada Urutan 8 Di Asia Tenggara DanUrutan 23 Di Asia,” 2023.

[9] Who, "Breast Cancer," 2021. [Online]. Available: Https://Www.Who.Int/News-Room/Fact-Sheets/Detail/Breast-Cancer

[10] M. A. Jabbar, E. Hasmin, Sunardi, And W. Musu, "Komparasi Algoritma Decision Tree , Naive Bayes , Dan K- Nearest Neighbors Dalam KlasifikasiKanker Payudara," *Csrid Journal*, Vol. 14, No. 3, Pp. 258–270, 2022.

[11] F. K. Nasser And S. F. Behadili, "Breast Cancer Detection Using Decision Tree And K-Nearest Neighbour Classifiers," *Iraqi Journal Of Science*, Vol.63, No. 11, Pp. 4987–5003, 2022, Doi: 10.24996/Ijs.2022.63.11.34.

[12] H. Rajaguru And S. R. Sannasi Chakravarthy, "Analysis Of Decision Tree And K-Nearest Neighbor Algorithm In The Classification Of Breast Cancer,"*Asian Pacific Journal Of Cancer Prevention*, Vol. 20, No. 12, Pp. 3777– 3781, 2019, Doi: 10.31557/Apjcp.2019.20.12.3777.

[13] Sriyanto and A. Ria Supriyatna, "Prediksi Penyakit Diabetes Menggunakan Algoritma Random Forest," *Ijccs*, vol. 17 No. 1, no. x, pp. 1–5, 2023.

[14] D. Prajarini, S. Tinggi, S. Rupa, D. Desain, and V. Indonesia, "Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Kulit," *Informatics J.*, vol. 1, no. 3, p. 137, 2016.