

Implementasi *K-Nearest Neighbours* Dengan Google Colab Untuk Klasifikasi Penyakit Kanker Payudara

Siti Rihastuti¹, Afnan Rosyidi², Handoko³

^{1,2} Program Studi Manajemen Informatika, STMIK AMIKOM Surakarta

³ Program Studi Komputerisasi Akutansi, STMIK AMIKOM Surakarta

Email :siti@dosen.amikomsolo.ac.id¹⁾, afnan@dosen.amikomsolo.ac.id²⁾, handoko@dosen.amikomsolo.ac.id³⁾

ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasi penyakit kanker payudara berdasarkan dataset pasien penderita penyakit kanker payudara menggunakan model *K-Nearest Neighbors* dan pengujian menggunakan Google Colab. Kanker payudara adalah salah satu jenis kanker yang paling umum di kalangan wanita dan deteksi dini sangat penting untuk meningkatkan peluang kesembuhan bagi penderitanya. Sebanyak 569 record dataset penyakit kanker payudara digunakan yang diambil dari situs kaggle. Terdapat 32 variabel yang ada didalam kriteria dataset yang akan diuji. Model KNN diterapkan dengan menentukan nilai $K = 1$ hingga 20. Dari hasil pengujian menggunakan KNN dan Google Colab terhadap dataset diperoleh nilai akurasi tertinggi sebesar 96.49% dengan nilai $K=9$. Akurasi cenderung stabil dan memiliki selisih yang sedikit mulai dari $K=3$ hingga $K=20$. Nilai presisi tertinggi terdapat pada angka 97,18% dengan beberapa nilai $K=4, 6, 8, 9$, dan $K=10$. Nilai recall dan F1-score tertinggi sebesar 97.18% pada $K=9$. Perolehan nilai akurasi, presisi, recall dan F1-score yang tinggi dan cukup stabil mampu mendeteksi kasus kanker payudara dengan tingkat akurasi yang baik. Nilai F1-score yang tinggi menunjukkan bahwa model KNN memiliki keseimbangan yang baik antara precision (ketepatan dalam memprediksi secara akurat) dan recall (sensitivitas dalam mengukur kinerja model dalam memprediksi kasus kanker payudara ganas yang sebenarnya). Berdasarkan hasil pengujian menunjukkan bahwa model *K-Nearest Neighbors* memiliki kinerja yang cukup baik dalam memprediksi kanker payudara.

Kata Kunci : kanker payudara, klasifikasi knn, Google colab

ABSTRACT

This study aims to classify breast cancer based on a dataset of breast cancer patients using the *K-Nearest Neighbors* model and testing using Google Colab. Breast cancer is one of the most common types of cancer among women and early detection is very important to increase the chances of recovery for sufferers. A total of 569 records of breast cancer datasets were used which were taken from the kaggle site. There are 32 variables in the criteria dataset to be tested. The KNN model is applied by determining the value of $K=1$ to 20. From the results of testing using KNN and Google Colab on the dataset, the highest accuracy value was obtained at 96.49% with a value of $K=9$. Accuracy tends to be stable and has a slight difference starting from $K= 3$ to $K=20$. The highest precision value is at 97.18% with several values of $K=4, 6, 8, 9$, and $K=10$. The highest recall and F1-score values are 97.18% at $K=9$. The high and stable accuracy, precision, recall and F1-score values are able to detect breast cancer cases with a good level of accuracy. A high F1-score value indicates that the KNN model has a good balance between precision (accuracy in predicting accurately) and recall (sensitivity in measuring model performance in predicting actual malignant breast cancer cases). Based on the test results, it shows that the *K-Nearest Neighbors* model has a fairly good performance in predicting breast cancer.

Keywords : breast cancer, knn classification, Google colab

PENDAHULUAN

Penyakit tumor adalah benjolan atau jaringan yang tumbuh tidak normal yang bisa bersifat jinak maupun ganas (kanker). Pada bentuk tumor memiliki tepi halus dan pada perkembangan tumor lebih lambat dibandingkan kanker[1]. Kasus penyakit tumor/kanker di Indonesia berada pada urutan 8 di Asia Tenggara, sedangkan di Asia urutan ke 23, dengan kasus tertinggi untuk perempuan adalah kanker payudara dan kanker leher rahim. Berdasarkan data Riskesdas, prevalensi tumor/kanker di Indonesia menggambarkan adanya peningkatan dari 1,4 menjadi 1,79, dari tahun 2013 ke tahun 2018 dengan urutan tertinggi berada di wilayah provinsi DI Yogyakarta, Sumatera Barat dan Gorontalo[2].

Terdapat dua macam penyakit tumor yaitu tumor jinak dan tumor ganas (kanker). Tumor jinak pertumbuhannya lambat dan cenderung tidak berbahaya. Namun, ada juga sel tumor jinak yang berukuran besar dan pertumbuhannya cepat, hingga dapat mengganggu jaringan di sekitarnya. Tumor ganas (kanker) adalah pertumbuhan sel yang tidak terkendali (tidak normal) di dalam tubuh, yang disebabkan oleh perubahan (mutasi) DNA di dalam sel, sehingga dapat menyebar dan merusak sel normal di sekitarnya maupun bagian tubuh lain[3].

Kanker payudara merupakan tumor ganas dan penyakit yang banyak diderita oleh wanita, kondisi tersebut terjadi saat sel dalam payudara membelah dan tidak terkendali pertumbuhannya. Pada tahun 2020 jumlah kasus baru kanker payudara di Indonesia sebanyak 68.858 kasus dari total 396.914 kasus baru, dengan jumlah kematian lebih dari 22 ribu jiwa kasus[4]. Beberapa gejala kanker payudara yang muncul diantaranya terdapat benjolan yang tidak normal pada payudara, perubahan bentuk atau ukuran payudara dan perubahan pada puting[5]. Salah satu upaya mandiri dalam mengantisipasi kanker payudara adalah dengan melaksanakan SADARI (pemeriksaan payudara sendiri), namun jika berdasarkan pemeriksaan medis telah divonis terkena kanker payudara maka perlu penanganan lebih lanjut oleh dokter.

Untuk mengetahui apakah tumor payudara yang diderita oleh pasien termasuk tumor jinak atau ganas, diperlukan metode yang efektif

dengan gejala pasien, tes diagnostik dan riwayat medis pengecekan. Salah satu metode yang bisa digunakan adalah dengan data mining. Penerapan data mining bidang kesehatan telah mengalami kemajuan dimana data mining dapat diterapkan untuk proses klasifikasi berdasarkan penggalan pola atau hubungan pada sekumpulan data kompleks[6].

Algoritma KNN (K-Nearest Neighbor) merupakan salah satu algoritma paling sederhana yang digunakan dalam machine learning untuk regresi dan klasifikasi. Klasifikasi dalam data science adalah suatu proses memprediksi kelas atau kategori data dengan memanfaatkan nilai yang ada pada data. Proses klasifikasi pada dasarnya dilakukan agar analisis data menjadi lebih mudah dan juga memberikan hasil yang akurat[7]. KNN adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut[8]. Algoritma KNN termasuk algoritma sederhana namun akurat untuk klasifikasi data, yaitu analisis data yang membantu orang memprediksi label kelas mana yang harus diklasifikasikan ke dalam sampel[9][10]. Hasil akhir berupa tingkat akurasi, presisi, recall dan F1-score yang akan dianalisa dan dievaluasi untuk mengetahui apakah model KNN yang dipakai pada pengujian terhadap dataset merupakan algoritma yang tepat dalam memprediksi jenis kanker payudara. Akurasi didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual. Presisi diartikan sebagai rasio item relevan atau kecocokan yang dipilih terhadap semua item yang terpilih. Recall didefinisikan sebagai rasio dari item relevan yang dipilih terhadap total jumlah item relevan yang tersedia. F1-score merupakan keseimbangan antara presisi dan recall.

Penelitian ini bertujuan untuk mengetahui pengimplementasian model KNN menggunakan Google Colabotary dalam mengklasifikasi penyakit kanker payudara (jinak atau ganas) berdasarkan dataset pasien dan berapakah tingkat akurasi dengan algoritma KNN. Penelitian ini diharapkan dapat membantu bidang kedokteran dalam ketepatan penanganan dan perawatan pasien yang lebih efektif sesuai dengan kondisi kanker payudara yang diderita.

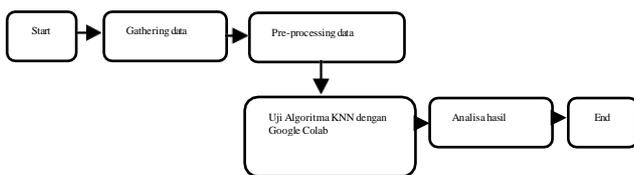
Penelitian yang dilakukan oleh [7] tentang penerapan KNN untuk klasifikasi penyakit kanker payudara menghasilkan akurasi sebesar 72,62% terhadap pasien yang dinyatakan sembuh akan terjangkit lagi atau tidak, sedangkan penelitian yang diusulkan akan menghasilkan klasifikasi kanker payudara termasuk jinak atau ganas beserta akurasinya.

Penelitian tentang deteksi tumor otak menggunakan dataset image menghasilkan akurasi yang mendekati 62%[11] dan penggunaan KNN untuk deteksi penyakit kanker payudara[12], kedua penelitian tersebut menggunakan dataset berupa image, sedangkan penelitian yang diusulkan menggunakan dataset berupa 32 variabel yang berisi nilai karakteristik visual kanker. Model KNN bisa dikombinasikan dengan SAW (Sample Adding Weight) untuk deteksi dini kanker serviks dengan tingkat akurasi mencapai 93,48% [13]. Analisis kanker paru-paru dengan model KNN dengan akurasi 80% [14].

Penelitian[15] membahas penerapan KNN dalam klasifikasi kanker payudara menggunakan software uji Rstudio menghasilkan akurasi 98% dengan k=11 dan k=21, sedangkan penelitian yang diusulkan menggunakan software uji Google colabotary. Penelitian [16] membahas penerapan KNN dalam klasifikasi kanker payudara dengan 10 variabel yang digunakna, sedangkan penelitian yang diusulkan menggunakan 32 variabel.

METODE

Tahapan penelitian digambarkan dalam Gambar 1 berikut:



Gambar 1. Tahapan Penelitian

a. Gathering data

Tahapan awal adalah dengan menentukan dataset yang akan digunakan. Pada penelitian ini digunakan 569 dataset pasien penyakit kanker payudara dengan jumlah atribut sebanyak 33 termasuk class.

b. Pre-processing data

Tidak semua data atau record yang ada pada

atau record dan atribut harus melalui beberapa tahap pengolahan awal data seperti pembersihan data (*cleaning*), *filtering*, transformasi data, dan lainnya sehingga diperoleh data yang berkualitas.

c. Uji Algoritma KNN dengan Google Colab

Penelitian ini menggunakan *K-Nearest Neighbours*. Penggunaan K-NN ditujukan untuk mendapatkan sejumlah nilai titik terdekat dengan nilai titik baru dengan nilai K = 1 hingga K=20. Berikut untuk alur perhitungan algoritma *K-Nearest Neighbor*. Pengujian akan dilakukan dengan *software Google Collaboratory* yang menggunakan bahasa pemrograman *Phyton*.

d. Analisa Hasil

Tahapan ini merupakan proses menganalisa hasil dari penerapan algoritma KNN dan hasil pengujian menggunakan Google Colab sesuai dengan tujuan penelitian. Validasi dilakukan untuk mengukur hasil klasifikasi guna mengetahui tingkat *accuracy*, *precision*, dan *recall*.

HASIL DAN PEMBAHASAN

a. Data Pre-processing

Sebanyak 569 dataset dan 32 atribut termasuk class yang digunakan dalam penelitian yang diambil dari kaggle.com. Penelitian ini dimaksudkan untuk mengetahui tingkat akurasi penerapan KNN menggunakan software uji Google Colab dalam mengklasifikasi kanker payudara.

radius_mcm	texture_mcm	perimeter_mcm	area_mcm	smoothness_mcm	compactness_mcm	concavity_mcm	convexity_mcm	points_mcm	symmetry_mcm	fractal_dimension
17.99	10.38	122.8	1000.01189	0.2776	0.3001	0.1471	0.2419	0.0751		
20.57	17.77	132.9	1326.008474	0.07864	0.0869	0.07017	0.1812	0.05667		
19.69	21.25	130	1203.01096	0.1599	0.1974	0.1279	0.2069	0.05999		
11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052	0.2597	0.09744	
20.29	14.34	155.1	1297.01003	0.1328	0.198	0.1043	0.1809	0.05883		
12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	0.08089	0.2087	0.07613	
18.25	19.98	119.6	1040.00463	0.189	0.1127	0.074	0.1794	0.05742		
13.71	20.83	90.2	577.9	0.1189	0.1645	0.09566	0.05985	0.2196	0.07451	
15.78	17.89	103.6	781.00971	0.1292	0.09954	0.06606	0.1842	0.06082		
19.17	24.8	132.4	1123.00974	0.2458	0.2065	0.1118	0.2597	0.078		
15.85	23.95	103.7	782.7	0.08401	0.1802	0.09938	0.05364	0.1847	0.05338	
13.73	22.61	93.6	578.3	0.1131	0.2293	0.2128	0.08025	0.2069	0.07682	
14.54	27.54	90.73	458.8	0.1139	0.1595	0.1639	0.07364	0.2303	0.07077	
14.68	20.13	94.74	684.5	0.09867	0.072	0.07395	0.05259	0.1586	0.05922	
16.13	20.68	108.1	798.8	0.117	0.2022	0.1722	0.1028	0.2164	0.07356	
19.81	22.15	130	1260.009831	0.1827	0.1479	0.09498	0.1582	0.05395		
13.54	14.36	87.46	466.3	0.09779	0.08129	0.06664	0.04781	0.1885	0.05566	
13.08	15.71	85.83	520.01075	0.127	0.04568	0.0311	0.1967	0.06811		

Gambar 2. Dataset Kanker Payudara

b. Penerapan KNN

- 1) Buka Google Colab
- 2) Upload File dataset

```

# Mount Google Drive
from google.colab import drive
drive.mount('/content/drive')

[20] # Mount Google Drive
from google.colab import drive
drive.mount('/content/drive')

[39] import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib.pyplot as plt
data = pd.read_csv("/content/drive/MyDrive/KNNAlgoritmaDataset.csv")
  
```

3) Memanggil dataset

```
data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 569 entries, 0 to 568
Data columns (total 33 columns):
 #   Column                Non-Null Count  Dtype
----  -----
 0   id                    569 non-null    int64
 1   diagnosis             569 non-null    object
 2   radius_mean          569 non-null    float64
 3   texture_mean         569 non-null    float64
 4   perimeter_mean      569 non-null    float64
 5   area_mean            569 non-null    float64
.....
 30  symmetry_worst      569 non-null    float64
 31  fractal_dimension_worst 569 non-null    float64
```

4) Menerapkan Model KNN

Nilai K ditentukan dari 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 dan 20.

```
# Membuat model KNN
knn = KNeighborsClassifier(n_neighbors=5)
knn.fit(X_train, y_train)

# Melakukan prediksi pada data test
y_pred = knn.predict(X_test)
```

c. Pengujian dengan Google Colab

Sebanyak 569 dataset diuji menggunakan algoritma KNN dan software yang akan digunakan untuk pengujian pada penelitian ini yaitu Google Colaboratory dengan menggunakan bahasa pemrograman pythonGoogle Colab. Dataset dipisah ke dalam data testing sebanyak 20% dan 80% merupakan data training. Nilai K=1 hingga 20.

```
# Membagi dataset menjadi training dan testing set
X_train, X_test, y_train, y_test =
train_test_split(X, y, test_size=0.2,
random_state=42)
```

d. Hasil Pengujian

Dari pengujian yang dilakukan menggunakan Algoritma KNN dan Google Colab, dibuat evaluasi algoritma KN dengan Confusion matrix seperti pada Gambar 3. Confusion matrix bertujuan untuk memetakan kinerja algoritma dalam bentuk tabulasi. Confusion matrix menunjukkan hubungan antara benar tidaknya sebuah data dikategorikan. Confusion matrix terdiri dari True positive (TP), False Positive (FP), False Negative (FN), dan True Negative (TN). True positive merepresentasikan data yang berada pada kelas positif yang diprediksi secara benar oleh algoritma. False Positive merepresentasikan data yang seharusnya berada pada kelas positif diprediksi menjadi kelas negatif oleh algoritma.

False Negative merupakan data yang seharusnya berada di kelas negatif diprediksi menjadi kelas positif oleh algoritma. True Negative merupakan data yang berada pada kelas negatif dan diprediksi secara benar oleh algoritma.

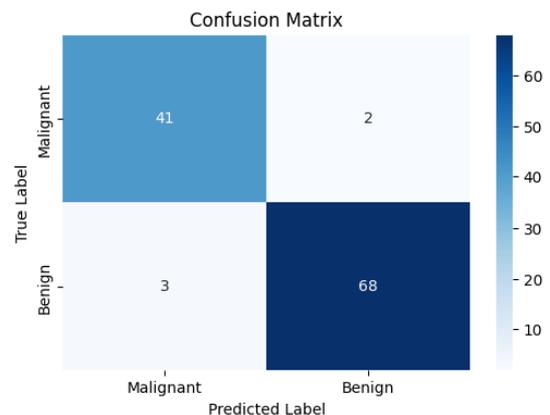
```
# Visualisasi Confusion Matrix
import seaborn as sns
plt.figure(figsize=(6,4))
sns.heatmap(cm, annot=True, fmt="d",
cmap="Blues", xticklabels=['Malignant',
'Benign'], yticklabels=['Malignant',
'Benign'])
plt.xlabel('Predicted Label')
plt.ylabel('True Label')
plt.title('Confusion Matrix')
plt.show()
```

```
Confusion Matrix:
[[41  2]
 [ 3 68]]

Classification Report:
              precision    recall  f1-score   support

     0       0.93      0.95      0.94         43
     1       0.97      0.96      0.96         71

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



Gambar 3. Confusion Matrix

Jumlah prediksi benar (True Positive) untuk kelas 0 sebanyak 43 dan jumlah prediksi salah (False Negatif) sebanyak 2. Presisi pada kelas 0 menunjukkan nilai 0,93 untuk prediksi benar, Recall 0,95 dan F1-score 0,94. Jumlah prediksi benar (True Positive) untuk kelas 1 sebanyak 68 dan jumlah prediksi salah (False Negatif) sebanyak 3. Presisi pada kelas 1 menunjukkan nilai 0,97 untuk prediksi benar, Recall 0,96 dan F1-score 0,96. Confusion matrix menunjukkan hasil bahwa model mampu memprediksi sebagian besar data dengan benar, dengan total hanya 5 kesalahan dari 114 data. Model KNN bekerja sangat baik dengan akurasi sebesar 96%. Kinerja model seimbang dan akurat dalam menangani kelas 0 dan kelas 1 terlihat dari nilai presisi, recall, dan F1-score yang tinggi.

Selanjutnya meenentukan nilai akurasi, presisi, recall dan F1-score dengan berbagai nilai K, dimulai dari 1 hingga 20, hasilnya terlihat pada Gambar 4.

```
#Menguji berbagai nilai K dan menghitung
metrik evaluasi
for k in k_values:
    knn =
    KNeighborsClassifier(n_neighbors=k)
    knn.fit(X_train, y_train)
    y_pred = knn.predict(X_test)

    accuracies.append(accuracy_score(y_test,
y_pred))

    precisions.append(precision_score(y_test,
y_pred))

    recalls.append(recall_score(y_test,
y_pred))

    f1_scores.append(f1_score(y_test,
y_pred))

# Menampilkan hasil dalam bentuk tabel
results_df = pd.DataFrame({
    'K': k_values,
    'Accuracy': accuracies,
    'Precision': precisions,
    'Recall': recalls,
    'F1-Score': f1_scores
})
```

K	Accuracy	Precision	Recall	F1-Score	
0	1	0.938596	0.944444	0.957746	0.951049
1	2	0.938596	0.957143	0.943662	0.950355
2	3	0.947368	0.957746	0.957746	0.957746
3	4	0.956140	0.971429	0.957746	0.964539
4	5	0.947368	0.957746	0.957746	0.957746
5	6	0.956140	0.971429	0.957746	0.964539
6	7	0.947368	0.957746	0.957746	0.957746
7	8	0.956140	0.971429	0.957746	0.964539
8	9	0.964912	0.971831	0.971831	0.971831
9	10	0.956140	0.971429	0.957746	0.964539
10	11	0.956140	0.958333	0.971831	0.965035
11	12	0.956140	0.958333	0.971831	0.965035
12	13	0.956140	0.958333	0.971831	0.965035
13	14	0.956140	0.958333	0.971831	0.965035
14	15	0.956140	0.958333	0.971831	0.965035
15	16	0.956140	0.958333	0.971831	0.965035
16	17	0.947368	0.945205	0.971831	0.958333
17	18	0.956140	0.958333	0.971831	0.965035
18	19	0.947368	0.945205	0.971831	0.958333
19	20	0.956140	0.958333	0.971831	0.965035

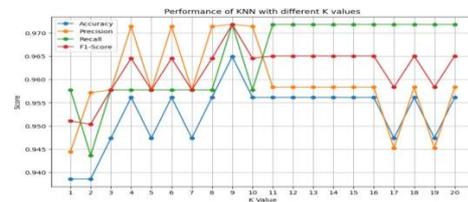
Gambar 4. Hasil Pengujian

Gambar 4 menunjukkan hasil akurasi, presisi, recall dan F1-score dengan berbagai nilai K=1 hingga 20. Nilai K=1, tingkat akurasinya 0,93, presisi 0,94, recall 0,95 dan F1-score 0,95. Nilai K=2 dengan akurasi 93%, presisi 95%, recall 94% dan F1-score 95%. Nilai K berikutnya (3 hingga 20) menunjukkan hasil yang tinggi dan tidak berbeda jauh selisihnya. Nilai akurasi tertinggi berada pada nilai K=9 yaitu 96,4%. Nilai presisi, recall dan F1-score tertinggi pada angka 97,18%. Berikutnya dibuat performa metrik untuk mengetahui grafik akurasi, presisi, recall dan F1-score dari berbagai nilai K. Hasilnya bisa dilihat pada Gambar 5.

Visualisasi performa metrik untuk berbagai

```
plt.plot(k_values, accuracies,
label='Accuracy', marker='o')
plt.plot(k_values, precisions,
label='Precision', marker='o')
plt.plot(k_values, recalls, label='Recall',
marker='o')
plt.plot(k_values, f1_scores, label='F1-
Score', marker='o')
plt.xticks(k_values)
plt.xlabel('K Value')
plt.ylabel('Score')
plt.title('Performance of KNN with different
K values')
plt.legend()
plt.grid(True)
plt.show()
```

Hasil :



nilai K

```
plt.figure(figsize=(10,6))
```

Gambar 5. Performa Metrik

Dari hasil pengujian menggunakan KNN dan Google Colab terhadap dataset menghasilkan nilai akurasi dari K=1 dan K=2. Diperoleh nilai akurasi tertinggi sebesar 96.49% pada nilai K=9. Akurasi cenderung stabil berada pada angka 94% hingga 96% untuk semua nilai K yang diuji, dan memiliki selisih yang sedikit mulai dari nilai K=3 hingga K=20. Nilai presisi (ketepatan dalam memprediksi secara akurat prediksi jenis kanker malignant atau ganas dibandingkan dengan semua prediksi malignant) tertinggi terdapat pada angka 97,18% dengan beberapa nilai K=4, 6, 8, 9, dan K=10. Nilai recall tertinggi pada K=9 sebesar 97,18% (sensitivitas dalam mengukur kinerja model dalam memprediksi kasus kanker payudara ganas yang sebenarnya), yang berarti angka 97,18% menunjukkan prediksi malignant atau ganas adalah tepat. Nilai recall stabil pada angka 95.77% hingga 97.18% pada hampir semua nilai K yang diuji, yang berarti model KNN cukup baik dalam mendeteksi kasus malignant. Nilai F1-score (merupakan nilai yang digunakan untuk menilai performa model secara keseluruhan tertinggi sebesar 97.18% pada K=9.

KESIMPULAN

Perolehan nilai akurasi, presisi, recall dan F1-score yang tinggi dan cukup stabil mampu

mendeteksi kasus kanker payudara (malignant/ganas dan benign/jinak) dengan tingkat akurasi yang baik yaitu 96,49%. Nilai F1-score yang tinggi sebesar 97,18% menunjukkan bahwa model KNN memiliki keseimbangan yang baik antara precision dan recall yang masing-masing memiliki nilai 97,18%. Nilai K=9 menghasilkan keseimbangan yang sangat baik antara akurasi dan kemampuan model (presisi, recall, f1-score) untuk mendeteksi kasus kanker (malignant/ganas) serta mengurangi false positives. Nilai K=9 merupakan pilihan yang optimal sesuai hasil metrik yang paling konsisten. Berdasarkan hasil pengujian menunjukkan bahwa model K-Nearest Neighbors memiliki kinerja yang cukup baik dalam mengklasifikasi dan memprediksi kanker payudara.

SARAN

Saran untuk penelitian berikutnya adalah sebaiknya mengkombinasikan beberapa model atau algoritma dan menambah jumlah dataset untuk semakin meningkatkan hasil akurasi dalam memprediksi kasus kanker payudara.

Daftar Pustaka

- [1] dr. R. Larasati, "Perbedaan Tumor dan Kanker - Alodokter," Alodokter. [Online]. Available: <https://www.alodokter.com/komunitas/topic/apakah-perbedaan-yang-pasti-antara-tumor-dan-kanker-itu>
- [2] K. Biro, "Kanker Payudara Paling Banyak di Indonesia, Kemenkes Targetkan Pemerataan Layanan Kesehatan," Kemenkes. [Online]. Available: <https://sehatnegeriku.kemkes.go.id/baca/umum/20220202/1639254/kanker-payudaya-paling-banyak-di-indonesia-kemenkes-targetkan-pemerataan-layanan-kesehatan/>
- [3] dr. Sienny Agustin, "Tumor Ganas, Ketahui Penyebab, Gejala, dan Pengobatannya," *Alo Dokter*. p. 25 Maret 2024, 2023. [Online]. Available: <https://www.alodokter.com/tumor-ganas-ketahui-penyebab-gejala-dan-pengobatannya>
- [4] K. Indonesia, "Kanker Payudara Paling Banyak di Indonesia, Kemenkes Targetkan Pemerataan Layanan
- [5] D. K. Yogyakarta, "KENALI BERBAGAI GEJALA KANKER PAYUDARA".
- [6] M. F. Akbarollah, W. Wiyanto, D. Ardiatma, and A. T. Zy, "Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Penyakit Jantung," *J. Comput. Syst. Informatics*, vol. 4, no. 4, pp. 850–860, 2023, doi: 10.47065/josyc.v4i4.4071.
- [7] N. Meilani and O. Nurdiawan, "Data Mining untuk Klasifikasi Penderita Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor," *J. Wahana Inform.*, vol. 2, no. 1, pp. 177–187, 2023, [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer>.
- [8] Z. Zuriati and N. Qomariyah, "Klasifikasi Penyakit Stroke Menggunakan Algoritma K-Nearest Neighbor (KNN) Classification of Stroke Using the K-Nearest Neighbor (KNN) Algorithm," vol. 1, no. 1, pp. 1–8, 2023.
- [9] J. Fahmi Idris *et al.*, "Publisher KLASIFIKASI PENYAKIT KANKER PARU MENGGUNAKAN PERBANDINGAN ALGORITMA MACHINE LEARNING," *J. MEDIA Akad.*, vol. 2, no. 2, 2024.
- [10] D. Cahyanti, A. Rahmayani, and S. A. Husniar, "Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 39–43, 2020, doi: 10.33096/ijodas.v1i2.13.
- [11] E. Najwaini, T. E. Tarigan, and F. P. Putra, "Application of the K-Nearest Neighbors (KNN) Algorithm on the Brain Tumor Dataset," ... *Artif. Intell. ...*, vol. 1, no. 1, pp. 18–26, 2023, [Online]. Available: <https://www.jurnal.yoctobrain.org/index.php/ijaimi/article/view/85>
- [12] Y. Setiawan, "Data Mining berbasis Nearest Neighbor dan Seleksi Fitur untuk Deteksi Kanker Payudara," *J. Inform. J. Pengemb. IT*, vol. 8, no. 2, pp. 89–96, 2023, doi: 10.30591/jpit.v8i2.4994.
- [13] M. F. H. Siregar, I. N. Farida, and D. M. A. Widyadara, "Penerapan Metode SAW Dan KNN Untuk Deteksi Dini Kanker Serviks," *INOTEK*, vol. 7, pp. 2549–7952, 2023.
- [14] T. Abdi Mangun, O. Nurdiawan, and A.

Using K-Nearest Neighbor Algorithm,” vol. 2, no. 2, pp. 58–61, 2023, [Online]. Available:

https://ejournal.ubibanyuwangi.ac.id/index.php/jurnal_tinsika

- [15] W. Ramdhani, D. Bona, R. B. Musyaffa, and C. Rozikin, “Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor,” *J. Ilm. Wahana Pendidik.*, vol. 8, no. 12, pp. 445–452, 2022.
- [16] M. A. Shinami and S. Bahri, “Klasifikasi Penyakit Kanker Payudara Menggunakan Metode K Nearest Neighbor,” *J. FOURIER*, vol. 12, pp. 79–85, 2023, doi: 10.14421/fourier.2023.122.79-85.