KLASIFIKASI RANDOM FOREST TERHADAP DIAGNOSA PENYAKIT KANKER PAYUDARA BERDASARKAN STATUS KEGANASAN

(RANDOM FOREST CLASSIFICATION OF BREAST CANCER DISEASE DIAGNOSIS BASED ON MALIGNANCY STATUS)

Yusrinnatul Jinana[1)], Kusrini[2)], dan Kusnawi[3)]
[1, 2,3)] PJJ Informatika Universitas Amikom Yogyakarta
Yogyakarta
e-mail: yusrinnatuljinana@students.amikom.ac.id[1)], kusrini@amikom.ac.id[2)], kusnawi@amikom.ac.id[3)]

ABSTRAK

Kanker payudara ialah salah satu penyakit dengan tingkat kematian tertinggi di dunia. Terdapat dua jenis kanker payudara, yakni ganas dan jinak. Identifikasi jenisnya memungkinkan pencegahan dan pengobatan yang tepat sebelum menyebar ke organ lain. Oleh karena itu, penelitian ini diperlukan untuk menganalisis klasifikasi data kanker payudara dalam membedakan kanker payudara jinak dan ganas menggunakan teknik data mining, seperti random forest karena mampu memberikan prediksi yang akurat dengan tingkat kesalahan rendah. Dataset yang digunakan dalam penelitian ini yaitu berjumlah 569 dan diperoleh dari data rekam medis rumah sakit ciremai. Hasil penelitian ini menunjukkan bahwa Random Forest merupakan metode yang efektif dan akurat untuk klasifikasi kanker payudara dengan akurasi 95% serta nilai AUV-ROC sebesar 0.99 dan recall 97% yang menunjukkan kemampuan model dalam membedakan kedua jenis kanker payudara dengan sangat baik sehingga dapat mengurangi resiko. Penggunaan teknik 5-Fold Cross-Validation) memastikan bahwa hasil yang diperoleh stabil dan tidak bergantung pada pembagian data tertentu, sehingga meningkatkan generalisasi model. Eksperimen terhadap berbagai parameter (n_estimators, max_depth, ukuran da-ta latih) menunjukkan bahwa konfigurasi terbaik adalah n_estimators = 100 dan max_depth = 10, yang memberikan keseimbangan optimal antara akurasi dan kompleksitas model. Model ini dapat diterapkan dalam sistem pendukung keputusan medis (Medical Decision Support System) untuk membantu dokter dalam deteksi dini kanker payudara, sehingga meningkatkan kecepatan dan keakuratan diagnosa.

***Kata Kunci**: Data Mining, kanker payudara, Random forest, dan status keganasan.*

ABSTRACT

Breast cancer is one of the diseases with the highest mortality rate in the world. There are two types of breast cancer, namely malignant and benign. Identification of the type allows for prevention and appropriate treatment before it spreads to other organs. Therefore, this study is needed to analyze the classification of breast cancer data in distinguishing benign and malignant breast cancer using data mining techniques, such as random forest because it is able to provide accurate predictions with a low error rate. The dataset used in this study amounted to 569 and was obtained from medical records of the Ciremai Hospital. The results of this study indicate that Random Forest is an effective and accurate method for breast cancer classification with an accuracy of 95% and an AUV-ROC value of 0.99 and a recall of 97% which shows the model's ability to distinguish the two types of breast cancer very well so that it can reduce risk. The use of the 5-Fold Cross-Validation technique) ensures that the results obtained are stable and do not depend on certain data divisions, thereby increasing the generalization model. Experiments on various parameters (n_estimators, max_ depth, training data size) show that the best configuration is n_estimators = 100 and max_ depth = 10, which provides an optimal balance between accuracy and model complexity. This model can be applied in a Medical Decision Support System (MDS) to assist doctors in early detection of breast cancer, thereby increasing the speed and accuracy of diagnosis.

***Keywords**: Breast cancer, data mining, Random forest, and malignancy status.*

I. PENDAHULUAN

Kanker payudara merupakan salah satu jenis kanker yang paling umum dan mematikan terutama di kalangan wanita. Penyakit ini ditandai dengan pertumbuhan sebuah sel-sel payudara yang abnormal dan tak terkendali membentuk tumor, sel kanker payudara bermula di dalam lobules penghasil susu pada payudara [1]. Menurut data WHO, kanker payudara mencapai persentase kematian yang signifikan di seluruh dunia, dengan angka kejadian yang terus meningkat setiap tahunnya [2]. Kanker payudara terdapat dua jenis yakni jinak dan ganas [3], Mengetahui jenis kanker payudara memungkinkan pencegahan dan pengobatan yang tepat, mengurangi risiko efek samping dan kematian. Diagnosis dini sangat penting untuk meningkatkan peluang kesembuhan pasien kanker payudara. Namun, banyak kasus yang terdiagnosis pada stadium lanjut, yang berkontribusi terhadap tingginya angka kematian. Oleh karena itu, diperlukan analisis untuk mengklasifikasi data kanker payudara yang besar. Teknik data mining dapat diterapkan

untuk diagnosis dini, sehingga pengobatan dapat dilakukan sebelum kanker berkembang menjadi ganas dan menyebar ke organ lain. Klasifikasi ialah metode pengelompokan data berdasarkan karakteristik serupa ke dalam kelas yang telah ditentukan [4].

Menurut penelitian [5] yang membahas mengenai *random forest*, *random forest* mempunyai keunggulan karena setiap pohon keputusan bekerja secara tim dan berkolaborasi untuk menghasilkan prediksi akhir. Metode ini cocok untuk klasifikasi biner, mampu menangani dataset dengan variabel lebih banyak daripada observasi, serta efektif dalam mengelola prediksi kontinu dan kategorikal dengan akurasi tinggi. Dengan memakai *random forest* diharapkan dapat meningkatkan keakuratan diagnosis dan mempercepat proses identifikasi penyakit kanker payudara sehingga berdampak positif pada kualitas perawatan pasien dan mengurangi kesalahan diagnosis.

Penelitian ini bertujuan untuk menganalisis performa algoritma *random forest* dalam mengklasifikasikan diagnosis kanker payudara berdasarkan status keganasan. Dengan demikian, penelitian ini diharapkan dapat membuktikan efektivitas dan akurasi model *random forest* dalam membedakan antara kanker payudara jinak dan ganas, serta memberikan kontribusi bagi pengembangan sistem pendukung keputusan medis dalam diagnosis kanker payudara. Oleh karena itu penulis mengangkat penelitian tentang "klasifikasi *random forest* terhadap diagnosa penyakit kanker payudara berdasarkan status keganasan".

II. STUDI PUSTAKA

Beberapa penelitian sebelumnya telah dilakukan untuk mengidentifikasi metode *random forest*. Penelitian [6] membandingkan 3 metode algoritma yaitu Algoritma *Naïve Bayes, Random Forest, Neural Network* untuk memprediksi pasien penyakit jantung yang menghasilkan nilai akurasi tertinggi yakni algoritma *Naïve Bayes* sebesar 83%.

Dalam penelitian [7] membandingkan 2 metode dalam mendiagnosa kanker payudara yaitu *Xgboost* (XGB) dan *Support Vector Machine Learning* (SVM) dengan menghasilkan tingkat akurasi *XGBoost* tertinggi yakni sebesar 96,26%.

Dalam penelitian [8] mengimplementasikan metode *random forest* dalam memprediksi penyakit *stroke* dengan jumlah dataset 5110 dan menghasilkan tingkat akurasi sebesar 94,61%.

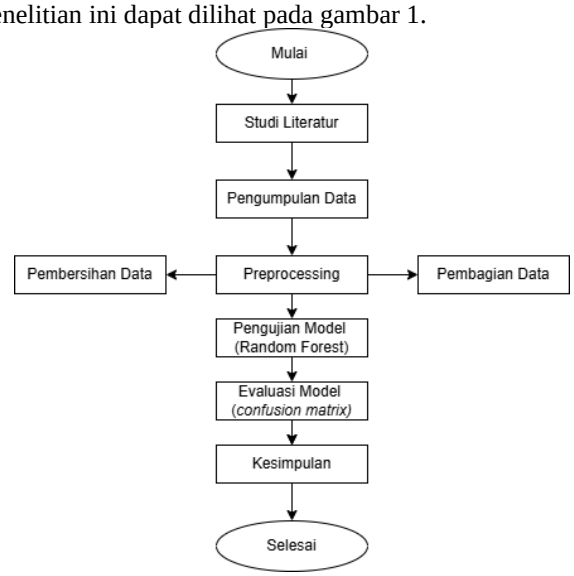
Dalam penelitian [9] implementasi algoritma *random forest* dalam klasifikasi diagnosis penyakit *stroke* dengan jumlah dataset 40910 dan menghasilkan tingkat akurasi sebesar 96%.

Dalam penelitian [10] membandingkan 5 algoritma yaitu *Decision Tree, Random Forest, Naïve Bayes,* SVM dan KNN untuk mendeteksi tanaman padi. Di studi ini algoritma yang mempunyai nilai akurasi tertinggi yakni algoritma KNN sebesar 87%

Dalam penelitian [11] memprediksi karakteristik tekstur memakai metode *Random Forest* pada citra *ultrasonografi* (USG) yang menghasilkan nilai akurasi 54%.

III. METODE PENELITIAN

Serangkaian tahapan penelitian yang dilakukan dalam penelitian ini dapat dilihat pada gambar 1.



Gambar 1 Alur Penelitian

1. Studi Literatur

Studi literatur yaitu tahapan awal penelitian ini untuk melakukan pengumpulan data dengan cara mencari informasi terkait penelitian yang relevan dan dapat dipercaya

2. Pengumpulan data dan pemrosesan data

Sumber data

Data yang digunakan dalam penelitian ini yaitu data rekam medis di rumah sakit ciremai Cirebon. Dataset ini terdiri dari 569 data dengan 10 fitur numerik yang diperoleh dari hasil biopsi dapat dilihat pada Tabel 1.

Tabel 1 Atribut Dataset.

No	Atribut	Keterangan
1	*Radius*	Ukuran rata-rata inti sel
2	*Texture*	Variasi dalam intensitas
3	Perimeter	Keliling inti sel
4	Area	Luas inti sel
5	Smoothness	Perubahan local dalam radius
6	Compactness	Seberapa cekung bagian intisel
7	Concavity	Tingkat kemiringan sel.
8	Concave Points	Jumlah bagian concave dari kontur
9	Symmetry	Simetri dari tumor
10	Fractal dimension	Dimensi fractal dari tumor

Klasifikasi target dalam dataset ini yaitu:
0 = Jinak (*Benign*)
1 = Ganas (*Malignant)*

3. Prepocessing

Preprocessing merupakan proses persiapan dan pembersihan data sebelum digunakan, sehingga model dapat belajar dengan baik dan menghasilkan prediksi yang akurat.

a. Pembersihan data

Pembersihan data pada dataset kanker payudara ini mencakup menghapus nilai yang hilang (*missing values*), normalisasi fitur untuk memastikan skala yang seragam dalam pemodelan, dan melakukan eksplorasi data untuk memahami distribusi variabel, korelasi antar fitur, dan mengidentifikasi *outlier*.

b. Pembagian data

Dalam penelitian ini data dibagi menjadi data latih (70%) dan data uji (30%) untuk menghindari overfitting dan digunakan ***5-fold cross-validation*** untuk meningkatkan realibilitas hasil.

c. Pemodelan *random forest*

Random forest merupakan algoritma *ensemble* yang menggunakan banyak pohon keputusan (*decision tree)* untuk melakukan klasifikasi. Algoritma ini bekerja dengan beberapa tahap:

1. Dataset dibagi kedalam subset yaitu data dilatih pada beberapa *decision tree*
2. Setiap *decision tree* membuat prediksi
3. Hasil dari *decision tree* digabungkan untuk membuat keputusan akhir.

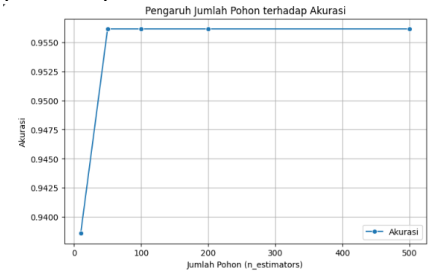
d. Evaluasi kinerja model

Model diuji menggunakan data uji dan dievaluasi menggunakan metrik utama yaitu akurasi, presisi (*precision)*, Recall (*Sensitivity), F1-Score*, AUC-ROC. Dan visualisasi evaluasi dengan *confusion matrix* dan *roc curve.*

e. Pengujian dan pencarian kinerja terbaik

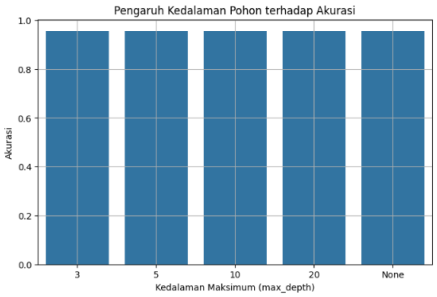
Untuk menentukan kinerja terbaik model random forest dalam klasifikasi kanker payudara, peneliti melakukan skenario uji yaitu:

1. Variasi jumlah pohon keputusan
   Hasil dari variasi jumlah pohon keputusan yaitu Akurasi meningkat seiring bertambahnya jumlah pohon, tetapi di atas 200 tidak terlalu signifikan. Gambar 2 menunjukkan hasil dari variasi jumlah pohon keputusan



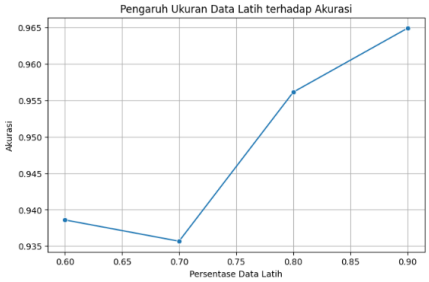
Gambar 2 variasi jumlah pohon keputusan

2. Variasi kedalaman pohon
   Hasil dari variasi kedalaman pohon yaitu terlalu dangkal (3-5) kurang akurat, tetapi terlalu dalam juga bisa menyebabkan *overfitting*. Gambar 3 menunjukkan hasil dari variasi kedalaman pohon



Gambar 3 variasi kedalaman pohon

3. Variasi jumlah data pelatihan
   Hasil dari jumlah data pelatihan yaitu Data latih lebih banyak cenderung meningkatkan akurasi, tetapi terlalu besar membuat model kurang *robust* terhadap data baru. Gambar 4 menunjukkan hasil dari variasi jumlah data pelatihan.



Gambar 4 variasi data pelatihan

f. Kesimpulan dan implikasi

Setelah melakukan pengujian dengan berbagai skenario, penelitian ini akan menyimpulkan kondisi optimal untuk random forest dalam membedakan kanker payudara jinak dan ganas berdasarkan:

- Jumlah pohon keputusan optimal
- Kedalaman pohon yang seimbang
- Teknik resampling yang efektif untuk meningkatkan deteksi kanker ganas.

IV. HASIL DAN PEMBAHASAN

Hasil dan implementasi dari penelitian ini yaitu:

1) Dataset yang digunakan.
   Penelitian ini menggunakan dataset kanker payudara dari data rekam medis rumah sakit ciremai. Dataset ini berisi 569 sampel dengan 10 atribut yang menggambarkan karakteristik sel kanker, penelitian ini membagi 2 target yaitu jinak dan ganas. Tabel 2 menunjukkan contoh detail dataset.

Tabel 2 Contoh Dataset *Rumah Sakit Ciremai*

diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean
M	17.99	10.38	122.8	1001	0.1184
M	20.57	17.77	132.9	1326	0.08474
M	19.69	21.25	130	1203	0.1096
M	11.42	20.38	77.58	386.1	0.1425
M	20.29	14.34	135.1	1297	0.1003

M	12.45	15.7	82.57	477.1	0.1278
M	18.25	19.98	119.6	1040	0.09463

2) Preprocessing data

a. Penghapusan data

Pengecekan awal dilakukan untuk melihat apakah ada data yang hilang. Dibawah ini *source code* penghapusan data. Gambar 6 menunjukkan *source code* penghapusan data.

```
[9] # Mengecek apakah ada missing values
    print("Jumlah missing values per kolom:\n", df.isnull().sum())

    # Menghapus data yang hilang (jika ada)
    df_cleaned = df.dropna()

    # Menampilkan ukuran dataset setelah penghapusan
    print("Ukuran dataset setelah pembersihan:", df_cleaned.shape)
```

Gambar 5 Penghapusan data

b. Normalisasi fitur

Agar fitur berada dalam skala yang sama, digunakan *MinMaxScaler* untuk melakukan normalisasi ke rentang 0-1.

Dibawah ini *source code* penghapusan data. Gambar 7 menunjukka normalisasi fitur

Normalisasi Fitur

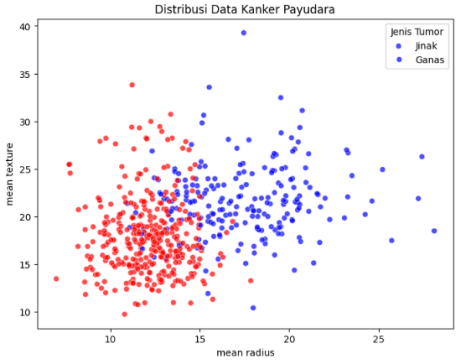
```
[17] # Memisahkan fitur dan target
     X = df_cleaned.drop(columns=['target'])
     y = df_cleaned['target']

     # Normalisasi fitur dengan MinMaxScaler
     scaler = MinMaxScaler()
     X_normalized = scaler.fit_transform(X)

     # Menampilkan data yang telah dinormalisasi (5 baris pertama)
     print(pd.DataFrame(X_normalized, columns=X.columns).head())
```

```
   mean radius  mean texture  mean perimeter  mean area  mean smoothness  \
0     0.521037      0.022658        0.545989   0.363733         0.593753
1     0.643144      0.272574        0.615783   0.501591         0.289880
2     0.601496      0.390260        0.595743   0.449417         0.514309
3     0.210090      0.360839        0.233501   0.102906         0.811321
4     0.629893      0.156578        0.630986   0.489290         0.430351

   mean compactness  mean concavity  mean concave points  mean symmetry  \
0          0.792037        0.703140             0.731113       0.686364
1          0.181768        0.203608             0.348757       0.379798
2          0.431017        0.462512             0.635686       0.509596
3          0.811361        0.565604             0.522863       0.776263
4          0.347893        0.463918             0.518390       0.378283

   mean fractal dimension  ...  worst radius  worst texture  worst perimeter  \
0                0.605518  ...      0.620776       0.141525         0.668310
1                0.141323  ...      0.606901       0.303571         0.539818
2                0.211247  ...      0.556386       0.360075         0.508442
3                1.000000  ...      0.248310       0.385928         0.241347
4                0.186816  ...      0.519744       0.123934         0.506948
```

Gambar 6 Normalisasi fitur

c. Eksplorasi data dan visualisasi

Visualisasi distribusi data kanker payudara berdasarkan dua fitur utama (*mean radius* dan *mean texture*) untuk melihat pola klasifikasi. Gambar 8 menunjukkan distribusi data kanker payudara berdasarkan dua fitur utama yaitu jinak dan ganas.



Gambar 7 Distribusi Data Kanker Payudara

d. Pembagian data

Dataset dibagi menjadi 70% data latih dan 30% data uji untuk menguji performa model. Gambar 9 menunjukkan hasil dari pembagian data latih dan data uji.

Membagi Data (Training 70% - Testing 30%)

```
# Membagi data menjadi training dan testing
X_train, X_test, y_train, y_test = train_test_split(X_normalized, y,
test_size=0.2, random_state=42, stratify=y)

# Menampilkan ukuran dataset
print("Ukuran Training Set:", X_train.shape)
print("Ukuran Testing Set:", X_test.shape)
```

```
Ukuran Training Set: (455, 30)
Ukuran Testing Set: (114, 30)
```

Gambar 8 Pembagian data

e. Implementasi *5-Fold Cross-Validation*

Untuk mendapatkan hasil yang lebih stabil, digunakan *5-Fold Cross-Validation*, yang membagi data menjadi 5 bagian, melatih model pada 4 bagian, dan mengujinya pada 1 bagian secara bergantian. Gambar 10 menunjukkan hasil dari implementasi *5 Fold cross-validation*.

5-Fold Cross-Validation dengan Random Forest

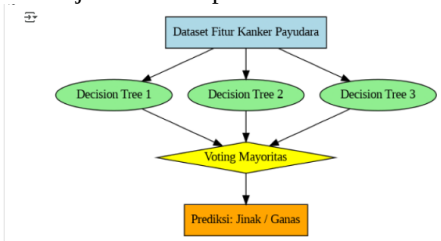
```
[20] # Inisialisasi model Random Forest
     rf_model = RandomForestClassifier(n_estimators=100, random_state=42)

     # Inisialisasi 5-Fold Cross-Validation
     kf = KFold(n_splits=5, shuffle=True, random_state=42)

     # Evaluasi model dengan cross-validation
     cv_scores = cross_val_score(rf_model, X_train, y_train, cv=kf, scoring='accuracy')

     # Menampilkan hasil cross-validation
     print("Akurasi tiap fold:", cv_scores)
     print("Rata-rata akurasi:", np.mean(cv_scores))
     print("Standar deviasi akurasi:", np.std(cv_scores))
```

```
Akurasi tiap fold: [0.93406593 0.96703297 0.96703297 0.98901099 0.92307692]
Rata-rata akurasi: 0.9560439560439561
Standar deviasi akurasi: 0.024075716813413892
```

Gambar 9 *5 Fold cross-validation*

f. Pelatihan dan evaluasi model

Model dilatih menggunakan *training set* dan dievaluasi pada *testing set*. Gambar 11 menunjukkan hasil pelatihan dan evaluasi model.

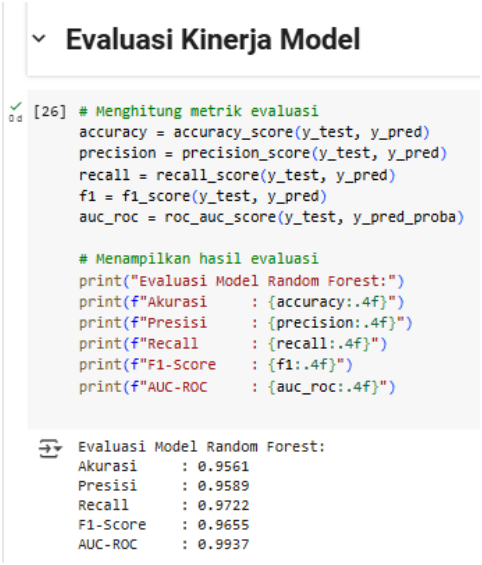



```
[25] # Inisialisasi model Random Forest
     rf_model = RandomForestClassifier(n_estimators=100, random_state=42)

     # Melatih model pada training set
     rf_model.fit(X_train, y_train)

     # Melakukan prediksi pada data testing
     y_pred = rf_model.predict(X_test)
     y_pred_proba = rf_model.predict_proba(X_test)[:, 1]  # Probabilitas kelas positif (ganas)
```

Gambar 10 Hasil pelatihan dan evaluasi model

g. Evaluasi kinerja model

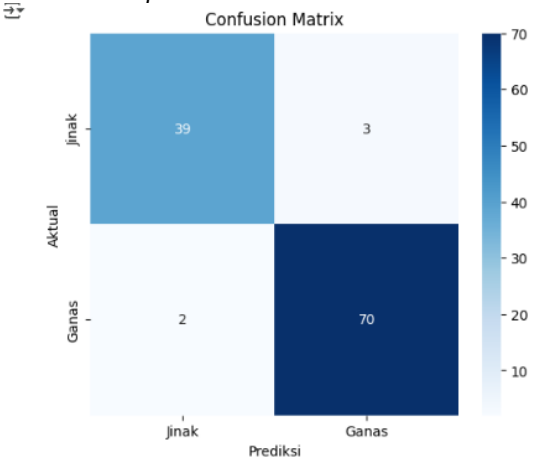
Model dievaluasi menggunakan akurasi, presisi, recall, *F1-score*, dan AUC-ROC. Gambar 12 menunjukkan hasil dari evaluasi kinerja model.



Gambar 11 hasil dari evaluasi kinerja model

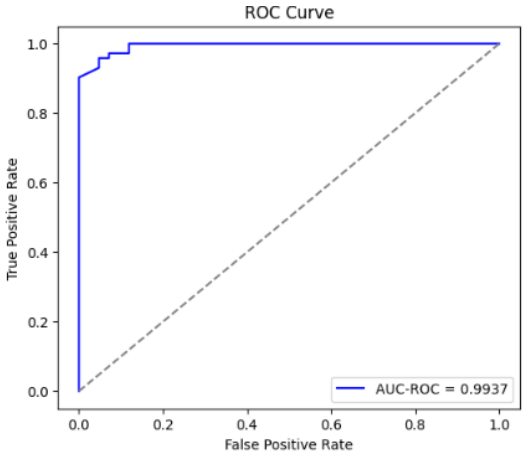
h. Visualisasi evaluasi model

Dalam penelitian ini menggunakan visualisasi evaluasi dengan *confusion matrix* dan *ROC Curve*. *Confusion matrix* merupakan tabel yang terdiri dari empat kombinasi antara nilai prediksi dan nilai aktual. Sedangkan *ROC Curve* merupakan sebuah *plot* grafis yang menggambarkan kinerja model klasifikasi biner pada berbagai nilai ambang (*threshold*). Gambar 13 menunjukkan hasil dari *confusion matrix.*



Gambar 12 *Confusion Matrix*

Gambar 14 menunjukkan hasil dari *ROC Curve*.



Gambar 13 *ROC Curve*

i. Analisis performa berdasarkan berbagai scenario. Gambar 15 merupakan pebandingan berbagai parameter dalam random forest.

Perbandingan berbagai parameter dalam Random Forest:

Parameter	Akurasi
n_estimators = 10	0.94
n_estimators = 100	0.97
n_estimators = 500	0.98
max_depth = 5	0.94
max_depth = 10	0.96
max_depth = None	0.97

Gambar 14 perbandingan parameter dalam random forest

Analisis

a) Jumlah pohon lebih banyak (*n_estimators* 100-500) cenderung meningkatkan akurasi.
b) Kedalaman pohon (*max_depth*) yang terlalu dangkal dapat menurunkan performa.

Dari hasil yang didapat peneliti membuat kesimpulan bahwa performa *random forest* dalam membedakan kanker payudara jinak dan ganas cenderung lebih baik dari beberapa metode yang dihasilkan penelitian sebelumnya. Perbedaan utama dari penelitian ini dibandingkan dengan penelitian sebelumnya adalah:

1. Akurasi *Random Forest* pada penelitian ini lebih tinggi (95%) dibandingkan penelitian sebelumnya (54%).
2. AUC-ROC sangat tinggi (0.99), menunjukkan bahwa model ini sangat mampu membedakan tumor jinak dan ganas.
3. *Random Forest* lebih *robust* terhadap fitur-fitur yang memiliki hubungan *non-linear*, sedangkan model seperti *Logistic Regression* yang digunakan oleh penelitian sebelumnya lebih terbatas dalam menangani kompleksitas data.

Random Forest lebih efektif dan akurat dibandingkan metode sebelumnya dalam membedakan tumor jinak dan

ganas. Hasil ini menunjukkan bahwa penggunaan model ini dalam system pendukung keputusan medis (*Medical Decision Support Systems*) dapat meningkatkan deteksi kanker payudara dengan lebih akurat dan andal.

V. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, model *random forest* menunjukkan kinerja yang sangat baik dalam klasifikasi kanker payudara jinak dan ganas, dengan akurasi 95% serta nilai AUV-ROC sebesar 0.99 dan recall 97% yang menunjukkan kemampuan model dalam membedakan kedua jenis kanker payudara dengan sangat baik sehingga dapat mengurangi resiko. Penggunaan teknik 5-*Fold Cross-Validation*) memastikan bahwa hasil yang diperoleh stabil dan tidak bergantung pada pembagian data tertentu, sehingga meningkatkan generalisasi model. Eksperimen terhadap berbagai parameter (*n_estimators*, *max_depth*, ukuran data latih) menunjukkan bahwa konfigurasi terbaik adalah *n_estimators* = 100 dan *max_depth* = 10, yang memberikan keseimbangan optimal antara akurasi dan kompleksitas model. Dari sisi implementasi dalam bidang medis, model *Random Forest* dapat digunakan sebagai alat bantu dalam diagnosa kanker payudara, memberikan hasil yang cepat dan akurat sehingga membantu dokter dalam pengambilan keputusan klinis.

DAFTAR PUSTAKA

[1] Sung H, F. J. (2022). *Global Cancer Observatory (GLOBOCAN)*. Retrieved from https://gco.iarc.who.int/media/globocan/factsheets/populations/360-indonesia-fact-sheet.pdf

[2] Sadikin, B. G. (2024). *Strategi Indonesia dalam Upaya Melawan Kanker* . jakarta

[3] Naomi Angela Meyana, Y. T. (2023). Klasifikasi Kanker Payudara Menggunakan Metode KNN. *SEMINARNASIONALSTATISTIKAAKTUARIA*.

[4] Jamaludin, Fajar, A. K., Mutaqin, M. Z., Mutoffar, M. M., & Setiyadi, D. (2024). KLASIFIKASIKANKER PAYUDARA MENGGUNAKAN ALGORITMA NEURAL NETWORK DAN RANDOM FOREST. *Jurnal Manajemen Informatika & Sistem Informasi (MISI*, DOI : 10.36595/misi.v5i2.

[5] D. Casadei, G. Serra, and K. Tani, "Implementation of a Direct Control Algorithm for Induction Motors Based on Discrete Space Vector Modulation," *IEEE Transactions on Power Electronics*, vol. 15, no. 4, pp.769–777, 2007.

[6] Derisma. (2020). Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung dengan Teknik Data Mining. *J. Appl. Informatics Comput*, vol. 4,no. 1, pp. 84–88, 2020, doi: 10.30871/jaic.v4i1.2152.

[7] Andryan, M. R., Fajri, M., & Sulistyowati, N. (2022). Komparasi Kinerja Algoritma Xgboost Dan Algoritma Support Vector Machine (Svm) Untuk Diagnosis Penyakit Kanker Payudara. JIKO (Jurnal Informatika Dan Komputer), 6(1), 1. https://doi.org/10.26798/jiko.v6i1.500

[8] Mitra, R. &. (2022). Efficient Prediction of Stroke Patients Using Random Forest Algorithm in Comparison to Support Vector Machine. *In Advances in Parallel Computing Algorithms, Tools and Paradigms*, (pp. 530–536). https://doi.org/10.3233/apc220075.

[9] Siregar, A. P., Purba, D. P., Pasaribu, J. P., & Bakara, K. R. (2023). Implementasi Algoritma Random Forest Dalam Klasifikasi Diagnosis Penyakit Stroke. *Jurnal Penelitian Rumpun Ilmu Teknik (JUPRIT)* .

[10] Purnamawati, A., Nugroho, W., Putri, D., & Hidayat, W. F. (2020). Deteksi Penyakit Daun pada Tanaman Padi Memakai Algoritma Decision Tree, Random Forest, Naïve Bayes, SVM dan KNN. *INFOTEKJAR:JURNAL NASIONAL INFORMATIKA DAN TEKNOLOGI JARINGAN*.

[11] Rahmawati, A., Sari, I. U., & Sumarti, H. (2024). Klasifikasi Tumor Payudara jinak dan ganas pada citra ultrasonografi (USG) berdasarkan karakteristik tekstur memakai metode Random Forest. *Teras Fisika*.