

## Klasifikasi dalam Mendeteksi Penyakit Kanker Payudara dengan Menggunakan Metode *Random Forest* dan *Adaboost*

Erisa Rizkyani<sup>1</sup>, Nur Aliffiyanti Iskandar<sup>2</sup>, Nurul Chamidah<sup>3</sup>  
Program Studi Informatika / Fakultas Ilmu Komputer  
Universitas Pembangunan Nasional Veteran Jakarta  
Jl. RS. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450, Indonesia  
erisar@upnvj.ac.id<sup>1</sup>, nurai@upnvj.ac.id<sup>2</sup>, nurul.chamidah@upnvj.ac.id<sup>3</sup>

**Abstrak.** Di Indonesia kanker payudara menempati posisi kedua dengan angka kematian mencapai 20.052 jiwa atau sebesar 1,41%. Berdasarkan data statistik GLOBOCAN 2018, tercatat 2,1 juta kasus kanker dengan jumlah meninggal 626 ribu jiwa. Angka ini akan terus meningkat apabila tidak adanya upaya penanganan secepat mungkin. Oleh karena itu diperlukan deteksi sedini mungkin dari kanker payudara, pendeteksian tersebut berguna untuk menentukan perencanaan pengobatan yang akan diambil. Adapun tujuan penelitian ini adalah memperoleh hasil akurasi yang optimal dalam mengklasifikasi dalam pendeteksian penyakit kanker payudara dengan menggunakan Metode *Random Forest* dan *Adaboost*. Berdasarkan penelitian yang telah kami lakukan, diperoleh nilai akurasi dengan menggunakan *Random Forest* sebesar 95% dan nilai akurasi dengan menggunakan menggunakan *Adaboost* sebesar 70%. Hal ini menunjukkan bahwa pendeteksian penyakit kanker payudara menggunakan *Random Forest* lebih baik digunakan dibandingkan dengan menggunakan *Adaboost*.

**Kata Kunci:** *Random Forest*, *Adaboost*, Kanker.

### 1 Pendahuluan

Frekuensi penyakit kanker payudara banyak dijumpai pada negara maju dan termasuk ke dalam salah satu jenis kanker yang banyak diderita dibandingkan dengan jenis kanker lainnya. Di Indonesia sendiri jenis kanker ini menempati posisi kedua dengan angka kematian mencapai 20.052 jiwa atau sebesar 1,41% [1]. Kanker payudara disebabkan oleh adanya pertumbuhan sel kanker yang tidak terkendali pada jaringan payudara [2]. Sel kanker tersebut yang awalnya hanya berada di payudara dapat menyebar ke jaringan tubuh lainnya [3].

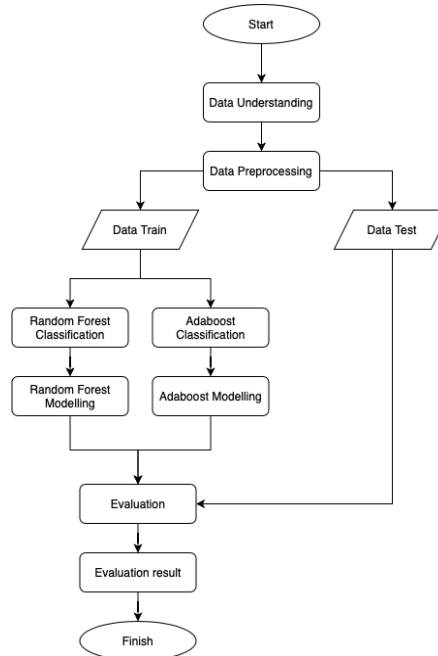
Berdasarkan data statistik GLOBOCAN 2018, tercatat 2,1 juta kasus kanker dengan jumlah meninggal 626 ribu jiwa. Angka ini akan terus meningkat apabila tidak adanya upaya penanganan secepat mungkin [2]. Oleh karena itu diperlukan deteksi sedini mungkin dari kanker payudara, pendeteksian tersebut berguna untuk menentukan perencanaan pengobatan yang akan diambil. Untuk menekan angka kematian perlu adanya sistem yang dapat membantu mendeteksi penyakit ini sedini mungkin sebelum kanker menyebar [1].

Dengan semakin berkembangnya teknologi informasi, metode *machine learning* digunakan untuk membantu pendeteksian otomatis [1]. Metode ini juga dapat membantu para ahli medis untuk mengurangi kesalahan diagnosis dalam waktu yang lebih singkat namun lebih rinci. Metode ini akan mendeteksi apakah kanker tersebut masuk ke dalam kelompok kanker ganas atau kanker jinak. Penelitian terkait klasifikasi dalam mendeteksi penyakit kanker payudara telah dilakukan sebelumnya. Penelitian pertama yaitu “deteksi penyakit kanker payudara dengan seleksi fitur berbasis *Principal Component Analysis* dan *Random Forest*”. Penelitian kedua yaitu “*ANN and Adaboost application for automatic detection of microcalcifications in breast cancer*”. Pada penelitian pertama menghasilkan akurasi sebesar 79,31% [1], sedangkan pada penelitian kedua menghasilkan akurasi sebesar 98,68% [4].

Adapun tujuan penelitian ini adalah memperoleh hasil akurasi yang optimal dalam mengklasifikasi dalam pendeteksian penyakit kanker payudara dengan menggunakan Metode *Random Forest* dan *Adaboost*.

## 2 Metode Penelitian

Berikut merupakan skema alur penelitian yang dilakukan.



**Gambar. 1.** Gambar di atas merupakan tahapan-tahapan yang dilakukan dalam penelitian ini. Dimulai dari data *understanding* hingga *evaluation result*.

### 2.1 Data

Metodologi penelitian dimulai dari pemahaman data yang dilakukan. Dalam melakukan penelitian, tahap memahami data sangat diperlukan dan harus dilakukan. Karena dalam tahapan ini perumusan dan definisi masalah akan ditemukan dan diselesaikan. Pada penelitian ini akan menggunakan dataset yang diperoleh dari *website Kaggle* ([www.kaggle.com](http://www.kaggle.com)). Dataset tersebut memiliki sembilan atribut dan satu kelas dengan jumlah 116 *record* data. Data tersebut nantinya akan dilakukan prediksi terhadap fitur-fitur yang mempengaruhi munculnya kanker payudara. Hasil dari penelitian diharapkan dapat dijadikan sebagai salah satu acuan atau sistem pendukung keputusan yang dapat membantu dalam mendeteksi kanker payudara.

### 2.2 Data Preprocessing

Setelah melakukan pengumpulan data dan data yang ada telah dipahami, tahapan selanjutnya yaitu melakukan praproses data. Praproses data merupakan tahapan untuk mengolah data agar sesuai dengan format atau struktur data yang dibutuhkan untuk ke proses selanjutnya. Langkah pertama adalah dengan melakukan *Data Cleaning*.

#### 2.2.1 Pemeriksaan *Missing Value*

Sebelum membuat model dengan metode yang digunakan, terlebih dahulu dilakukan pemeriksaan pada data. Hal tersebut berguna untuk mengetahui apakah di dalam data terdapat *missing value*, yang mana biasanya muncul pada suatu fitur pengamatan sebagai “Nan”, “?”, atau tidak ada nilai sama sekali (*blank*). Tentunya, dengan adanya *missing value* akan berpengaruh pada hasil akhir penelitian, dengan kata lain akan mengakibatkan pemodelan ataupun hasil akurasi menghasilkan hasil yang tidak optimal. Oleh sebab itu, *missing value* harus diatasi jika

ditemui pada fitur pengamatan yang digunakan. Dalam fitur yang digunakan dalam penelitian ini, didapatkan bahwa *missing value* tidak ada atau fitur-fitur yang digunakan memiliki nilai di dalamnya.

### 2.2.2 Pemeriksaan *Duplicate Data*

Selanjutnya adalah memeriksa apakah di dalam data yang digunakan terdapat duplikasi data atau salinan duplikat data berulang. Sama dengan *missing value*, adanya duplikat data akan mempengaruhi hasil akhir. Selain itu, menyebabkan ruang penyimpanan yang digunakan menjadi lebih besar dari yang seharusnya dan pemrosesan akan membutuhkan waktu yang lebih lama dibandingkan dengan tanpa adanya duplikat data. Maka, pemeriksaan ini harus dilakukan untuk menghindari hal itu semua. Pada fitur yang digunakan dalam penelitian ini, tidak ditemukan adanya duplikat data sehingga data dapat digunakan untuk proses selanjutnya.

### 2.2.3 Pemeriksaan *Outlier*

*Outlier* atau data yang memiliki nilai ekstrim, baik ekstrim tinggi maupun ekstrim rendah. Data yang memiliki *outlier* menjadi salah satu yang menandakan bahwa data tersebut memiliki struktur data yang tidak normal. Sehingga *outlier* harus dihilangkan atau dibuang dari data. Dari pemeriksaan tersebut didapatkan bahwa, tidak adanya outlier pada fitur.

Sehingga dari pemeriksaan tersebut didapatkan bahwa data yang digunakan pada penelitian ini sudah berstruktur normal, yaitu dengan tidak adanya *missing value*, duplikat data, dan *outlier*. Selepas itu, data dibagi ke dalam dua kelompok yaitu data latih (data *training*) dan data uji (data *testing*). Pengolahan data pada tahap *pre-processing* dilakukan dengan teknik *Resample*. Terdapat 116 *record* data yang akan diujikan dengan sembilan atribut dan satu kelas yang terdiri dari kontrol sehat dan pasien. Sedangkan untuk atribut terdiri dari *Age (years)*, *BMI (kg/m<sup>2</sup>)*, *Glucose (mg/dL)*, *Insulin (μU/mL)*, *HOMA*, *Leptin (ng/mL)*, *Adiponectin (μg/mL)*, *Resistin (ng/mL)*, dan *MCP-1 (pg/dL)*

## 2.3 Modelling

Kemudian setelah data telah dibagi menjadi Data *Train* dan Data *Test*, maka dari data latih (*train*) yang telah terbagi akan dilakukan proses *modelling* dengan menggunakan metode *Random Forest* dan *Adaboost*.

### 2.3.1 *Random Forest Classification*

*Random Forest* yaitu penggabungan dari beberapa *Decision Tree*. Dalam membangun *Decision Tree* diperlukan untuk menghitung nilai *entropy* yang berguna sebagai pemutus ketidakhomogenan atribut dan nilai perolehan informasi (*information gain*) [6]. Berikut ini adalah rumus persamaan dalam menghitung nilai *entropy* [5].

$$Entropy(Y) = -\sum_i p(c|Y) \log_2 p(c|Y). \quad (1)$$

Keterangan:

Y = himpunan kasus.

$p(c|Y)$  = proporsi nilai Y terhadap kelas c.

Sedangkan rumus persamaan dalam mencari nilai perolehan informasi sebagai berikut [5].

$$Information\ Gain(Y, a) = Entropy(Y) - \sum_{v \in Values} \frac{|Y_v|}{|Y|} Entropy(Y_v). \quad (2)$$

Keterangan:

*Values* (a) = seluruh nilai yang berpeluang pada himpunan kasus a.

$Y_v$  = subkelas Y serta kelas v yang memiliki keterkaitan pada kelas a.

$Y_a$  = keseluruhan nilai yang serupa dengan a.

### 2.3.2 *Adaboost Classification*

Adaptive boosting atau yang lebih dikenal dengan sebutan *Adaboost* merupakan suatu metode yang termasuk ke dalam *ensemble learning* pada algoritma boosting [6]. Selain itu, metode ini bagian dari algoritma *supervised* pada data mining yang diterapkan dalam membuat model klasifikasi. Umumnya *Adaboost* digunakan untuk model regresi, namun kini penggunaan *Adaboost* dapat digunakan untuk model statistik lainnya. Sistem kerja metode ini adalah *Adaboost* akan memberikan sejumlah bobot lebih kepada objek yang dinilai tidak tepat atau berada pada klasifikasi lemah. Berikut ini adalah tahapan dalam penerapan algoritma *Adaboost* [6].

- a. Masukkan (*input*): beberapa objek penelitian dengan label  $\{(x_i, y_i), \dots, (x_N, y_N)\}$ , suatu komponen algoritma pembelajaran, jumlah perputaran  $T$ .
- b. Inisialisasi (*initialize*): bobot objek pelatihan  $w_i^1 = 1/N$ , berlaku untuk seluruh  $i = 1, \dots, N$ .
- c. Do for  $t=1, \dots, T$ .
- d. Fungsi penggunaan komponen algoritma pembelajaran ialah melatih komponen klasifikasi,  $h_t$ , pada objek bobot pelatihan.
  - a. Menghitung kesalahan pelatihan dengan  $h_t: \varepsilon_t = \sum_{i=1}^N w_i^t, y_i \neq h_t(x_i)$
  - e. Menentukan bobot pada komponen klasifikasi  $h_t = a_t = \frac{1}{2} \ln \left( \frac{1-\varepsilon_t}{\varepsilon_t} \right)$
  - f. Memperbarui bobot objek pelatihan  $w_i^{t+1} = \frac{w_i^t \exp \{-a_t y_i h_t(x_i)\}}{C_t}, i = 1, \dots, N$   $C_t$  merupakan konstanta normalisasi.
  - g. Keluran (*output*):  $f(x) = \text{sign} \left( \sum_{t=1}^T a_t h_t(x) \right)$ .

## 2.4 Evaluation

Proses evaluasi dilakukan berdasarkan dari nilai akurasi, *ROC AUC Curves*, dan *confusion matrix*. Nantinya, hasil yang didapatkan akan disajikan ke dalam bentuk tabel dan grafik.

### 2.4.1 Confusion Matrix

*Confusion Matrix* digunakan untuk mengetahui seberapa baik ataupun besar akurasi yang diperoleh dari model klasifikasi yang telah dibuat dalam memprediksi atau mengklasifikasi kelas dari data *testing*. Berikut adalah matriks klasifikasinya:

**Tabel 1.** Penjelasan *Confusion Matrix*

		Nilai Sebenarnya	
		<i>True</i>	<i>False</i>
Nilai Prediksi	<i>True</i>	TP ( <i>True Positive</i> ) <i>Correct result</i>	FP ( <i>False Positive</i> ) <i>Unexpected result</i>
	<i>False</i>	FN ( <i>False Negative</i> ) <i>Missing result</i>	TN ( <i>True Negative</i> ) <i>Correct absence of result</i>

$$\text{precision} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{recall} = \frac{TP}{TP+FN} \quad (2)$$

$$\text{accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

*True Positive* atau TP adalah jumlah data *positive* yang terklasifikasi sebagai *positive*. *True Negative* atau TN adalah jumlah data *negative* yang terklasifikasi sebagai *negative*. *False Positive* atau FP adalah jumlah data

*negative* yang terklasifikasi sebagai *positive* atau salah prediksi. *False Negative* atau FN adalah jumlah data *positive* yang terklasifikasi sebagai *negative* atau salah prediksi. [7]

### 2.4.2 K-Fold Cross Validation

*Cross-Validation* atau estimasi rotasi, ialah suatu metode untuk memvalidasi model dalam menilai hasil statistik analisis ketika kumpulan data independen digeneralisasi [8]. Teknik ini pengembangan dari model *Split Validation* dimana validasinya mengukur *training error* dengan melakukan pengujian pada *test data*. Pengujian tersebut dilakukan sejumlah k kali yang digunakan. Langkah-langkah penggunaan *K-Folds Cross Validation* adalah membagi keseluruhan data menjadi n bagian, lalu melakukan iterasi (*fold* ke 1), yang mana bagian pertama menjadi *testing* dan bagian lainnya menjadi *training*, setelah itu menghitung *accuracy* dengan menggunakan rumus persamaan: [9].

$$Accuracy = \frac{\text{jumlah klasifikasi benar}}{\text{jumlah data uji}} \times 100\% \quad (1)$$

Tujuan dari penggunaan Cross Validation yaitu untuk mengukur performa dari model yang digunakan dalam mengolah data. Selain itu Cross Validation juga digunakan untuk membandingkan metode-metode yang digunakan dalam mengolah data, dalam penelitian ini membandingkan metode Random Forest dan Adaboost. [10]

## 3 Hasil dan Pembahasan

### 3.1 Data Preprocessing

Dataset kanker payudara yang digunakan berasal dari *website Kaggle (www.kaggle.com)*, yang terdiri dari sembilan atribut dan satu kelas dengan jumlah 116 *record* data. Sebelum dilakukan proses *modelling*, terlebih dahulu melakukan tahapan *preprocessing* yakni menyiapkan data agar menghasilkan struktur data yang dibutuhkan. Dalam proses ini, didapatkan bahwa dataset kanker payudara merupakan dataset yang memiliki struktur data yang baik, dimana tidak ditemukannya *missing value*, *duplicate data*, dan *outlier* pada data sehingga dapat melanjutkan ke proses selanjutnya.

Tabel 2. Data kanker payudara

Age	BMI	Glucose	Insulin	HOMA	Leptin	Adiponectin	Resistin	MCP.1	Classification
48	23.5	70	2.707	0.467409	8.8071	9.7024	7.99585	417.11 4	1
83	20.69	92	3.115	0.706897	8.8438	5.429285	4.06405	468.78 6	1
82	23.12	91	4.498	1.009651	17.9393	22.43204	9.27715	554.69 7	1
68	21.36	77	3.226	0.612725	9.8827	7.16956	12.766	928.22	1
86	21.11	92	3.549	0.805386	6.6994	4.81924	10.57635	773.92	1
...	...	...	...	...	...	...	...	...	...
45	26.85	92	3.33	0.755688	54.68	12.1	10.96	268.23	2
62	26.84	100	4.53	1.1174	12.45	21.42	7.32	330.16	2
65	32.05	97	5.73	1.370998	61.48	22.54	10.33	314.05	2
72	25.59	82	2.82	0.570392	24.96	33.75	3.27	392.46	2
86	27.18	138	19.91	6.777364	90.28	14.11	4.35	90.09	2
45	26.85	92	3.33	0.755688	54.68	12.1	10.96	268.23	2

### 3.2 Klasifikasi Random Forest dan Adaboost

### 3.2.1 Random Forest

**Tabel 3.** *Confusion Matrix* dan Akurasi *Random Forest*.

	True Negatif	True Positif	Akurasi
Pred. Negatif	11	1	0.916
Pred. Positif	0	12	1.0

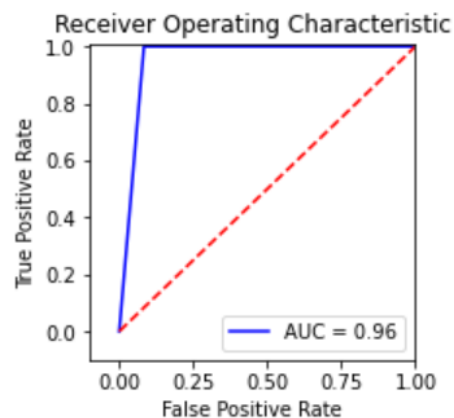
Negatif : Sehat  
 Positif : Pasien

Tabel di atas merupakan *Confusion Matrix* dari klasifikasi *Random Forest*. Berdasarkan tabel tersebut akurasi yang dihasilkan sebesar 0.958%.

**Tabel 4.** Hasil nilai *ROC AUC Curves*, *Sensitivity*, dan *Specificity* dengan *Random Forest*.

<i>ROC AUC Curves</i>	0.958
<i>Sensitivity</i>	0.916
<i>Specificity</i>	1.0

Tabel di atas merupakan hasil nilai *ROC AUC Curves*, *Sensitivity*, dan *Specificity* dari model *Random Forest* dihasilkan nilai *ROC AUC* sebesar 0.958, nilai *Sensitivity* sebesar 0.916 atau 91%. dan nilai *Specificity* sebesar 1.0 atau 100%. Nilai *ROC AUC* yang diperoleh menunjukkan bahwa penggunaan model klasifikasi bagus karena nilai *AUC* mendekati 1 yaitu 0.958. Nilai *Sensitivity* yang diperoleh menunjukkan bahwa sebesar 91% pasien diprediksi sakit dibandingkan dengan keseluruhan pasien yang sebenarnya sakit, dan nilai *Specificity* yang diperoleh menunjukkan bahwa sebesar 100% pasien benar tidak sakit dibandingkan dengan semua pasien yang sebenarnya tidak sakit.



**Gambar. 2.** Gambar berikut merupakan *ROC AUC Curve* Klasifikasi *Random Forest*.

*AUC* yang dihasilkan dari Kurva *ROC AUC* dengan klasifikasi *Random Forest* mendapatkan nilai *AUC* sebesar 0.96, sehingga dari perolehan nilai tersebut model klasifikasi *Random Forest* memiliki ukuran keterpisahan yang baik, yang mana nilai *AUC* yang dihasilkan mendekati angka 1. Selain itu, grafik di atas memperlihatkan bahwa kurva berada pada situasi yang hampir ideal, dimana model mampu membedakan antara kelas positif dan kelas negatif.

Untuk hasil dari proses validasi dengan menggunakan *K-Fold*, dihasilkan nilai akurasi terhadap model klasifikasi *Random Forest* sebesar 0.802.

### 3.2.2 Adaboost

**Tabel 5.** *Confusion Matrix* dan Akurasi *Adaboost*.

	True Negatif	True Positif	Akurasi
Pred. Negatif	8	4	0.667
Pred. Positif	3	9	0.75

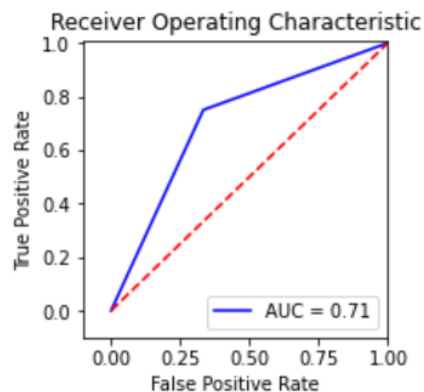
Negatif : Sehat  
 Positif : Pasien

Tabel di atas merupakan *Confusion Matrix* dari klasifikasi *Adaboost*. Berdasarkan tabel tersebut akurasi yang dihasilkan sebesar 0.708%.

**Tabel 6.** Hasil nilai *ROC AUC Curves*, *Sensitivity*, dan *Specificity* dengan *Adaboost*.

<i>ROC AUC Curves</i>	0.708
<i>Sensitivity</i>	0.666
<i>Specificity</i>	0.75

Tabel di atas merupakan hasil nilai *ROC AUC Curves*, *Sensitivity*, dan *Specificity* dari *Adaboost* dengan nilai *ROC AUC* sebesar 0.708, nilai *Sensitivity* sebesar 0.666, dan nilai *Specificity* sebesar 0.75 atau 75%. Nilai *ROC AUC* yang diperoleh menunjukkan bahwa penggunaan model klasifikasi tidak terlalu bagus karena nilai *AUC* tidak mendekati 1 yaitu 0.708. Nilai *Sensitivity* yang diperoleh menunjukkan bahwa sebesar 66% pasien diprediksi sakit dibandingkan dengan keseluruhan pasien yang sebenarnya sakit. Nilai *Specificity* yang diperoleh menunjukkan bahwa sebesar 75% pasien benar tidak sakit dibandingkan dengan semua pasien yang sebenarnya tidak sakit.



**Gambar. 3.** Gambar berikut merupakan *ROC AUC Curve* Klasifikasi *Adaboost*.



Untuk Kurva *ROC AUC* dengan klasifikasi *Adaboost*, dihasilkan nilai *AUC* sebesar 0.71. Dari perolehan nilai tersebut dapat dikatakan bahwa model dengan klasifikasi *Adaboost* memiliki peluang dalam membedakan antara kelas positif dan kelas negatif.

Proses validasi dengan *K-Fold* terhadap model klasifikasi *Adaboost* menghasilkan akurasi sebesar 0.706. Dapat disimpulkan dari proses validasi yang telah dilakukan, dihasilkan bahwa penerapan model klasifikasi *Random Forest* lebih baik dibandingkan dengan model klasifikasi *Adaboost*.

### 3.3 Tabel Perbandingan Algoritma *Random Forest* dan *Adaboost*

**Tabel 7.** Perbandingan hasil *Accuracy*, *Sensitivity*, dan *Specificity* dari klasifikasi *Random Forest* dan *Adaboost*.

	<i>Random Forest</i>	<i>Adaboost</i>
<i>Accuracy</i>	0.958	0.708
<i>Sensitivity</i>	0.916	0.666
<i>Specificity</i>	1.0	0.75
AUC	0.96	0.71

Berdasarkan tabel hasil perbandingan Klasifikasi *Random Forest* dan *Adaboost*, didapat nilai akurasi paling tinggi diperoleh ketika menggunakan metode *Random Forest* yaitu sebesar 0.958 dengan nilai AUC sebesar 0.96.

## 4 Penutup

### 4.1 Kesimpulan

Berdasarkan penelitian yang telah kami lakukan, diperoleh nilai akurasi dengan menggunakan *Random Forest* sebesar 0.958 atau 95% dan nilai akurasi dengan menggunakan menggunakan *Adaboost* sebesar 0.708 atau 70%. Hal ini menunjukkan bahwa pendeteksian penyakit kanker payudara menggunakan *Random Forest* lebih baik digunakan, yang didukung oleh perolehan nilai akurasi dan nilai AUC yang lebih tinggi dibandingkan dengan menggunakan *Adaboost*. Maka dari itu, model dengan klasifikasi *Random Forest* memiliki ukuran keterpisahan yang lebih baik dibandingkan model klasifikasi dengan *Adaboost* sehingga model *Random Forest* mampu membedakan antara pasien dengan penyakit dan pasien tanpa penyakit.

### 4.2 Saran

Dataset yang kami gunakan memiliki banyak fitur didalamnya. Hal ini membuat proses pengolahan data membutuhkan waktu yang cukup lama. Sebaiknya pada penelitian yang akan datang digunakan juga seleksi fitur pada dataset yang digunakan pada penelitian tersebut. Salah satu seleksi fitur yang cukup banyak digunakan yaitu PCA atau *Principal Component Analysis*. PCA memiliki keuntungan dalam segi akurasi serta kompleksitas waktu. Penggunaan *Adaboost* dalam penelitian ini menghasilkan akurasi yang tidak terlalu tinggi. Akan lebih baik pada penelitian selanjutnya peneliti dapat menggunakan metode lain yang lebih baik atau lebih cocok agar diperoleh nilai akurasi yang lebih tinggi dan lebih akurat.



## Referensi

- [1] Fauzi, Ahmad, Riki Supriyadi, & Nurlaelatul Maulidah. 2020. "Deteksi Penyakit Kanker Payudara Dengan Seleksi Fitur Berbasis Principal Component Analysis Dan Random Forest". *Jurnal Infotech*. Vol. 2 No. 1. Juni.
- [2] Bustamam, A., Bachtiar, A., Sarwinda, D. 2019. "Selecting Features Subsets Based On Support Vector Machine-recursive Features Elimination And One Dimensional-naïve Bayes Classifier Using Support Vector Machines For Classification Of Prostate And Breast Cancer". *Procedia Computer Science*, 157, 450–458. <https://doi.org/10.1016/j.procs.2019.08.238>
- [3] Harafani, Hani. 2019. "Forward Selection Pada Support Vector Machine Untuk Memprediksi Kanker Payudara". *Jurnal Infotech*. Vol. 1 No. 2. Desember.
- [4] Saada, Ghada, Ahmad Khadoura, & Qosai Kanafanib. 2016. "ANN and Adaboost application for automatic detection of microcalcifications in breast cancer". *The Egyptian Journal of Radiology and Nuclear Medicine*. Volume 47, Issue 4, December 2016, Pages 1803-1814. <https://doi.org/10.1016/j.ejrm.2016.08.020>
- [5] Siburian, Vanissa Wanika, & Ika Elvina Mulyana. "Prediksi Harga Ponsel Menggunakan Metode Random Forest." *Annual Research Seminar (ARS)*. Vol. 4. No. 1. 2019
- [6] Listiana, Eka, & Much Aziz Muslim. "Penerapan Adaboost Untuk Klasifikasi Support Vector Machine Guna Meningkatkan Akurasi Pada Diagnosa Chronic Kidney Disease." *Prosiding SNATIF (2017)*: 875-881.
- [7] Yoga, Taghfiruh Azhima, Prihandoko. 2018. "Penerapan Optimasi Berbasis Particle Swarm Optimization (Pso) Algoritma Naïve Bayes Dan K-nearest Neighbor Sebagai Perbandingan Untuk Mencari Kinerja Terbaik Dalam Mendeteksi Kanker Payudara". *Metik Jurnal*. Vol.2. No.2
- [8] Tempola, Firman, Miftah Muhammad, & Amal Khairan. "Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation." *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)* 5.5 (2018).
- [9] Hadikristanto, Wahyu, & Vidya Anis Fitri. "Penerapan Algoritma Naïve Bayes Berbasis Particle Swarm Optimization (Pso) Dalam Menangani Kasus Kanker Payudara." *Jurnal SIGMA* 9.1 (2019): 111-120.
- [10] Haristu, Reinardus Aji. 2019. "Penerapan Metode Random Forest Untuk Prediksi Wint Ratio Pemain Player Unknown Battleground". Diakses dari <https://repository.usd.ac.id> pada tanggal 13 Agustus 2021