

## Analisis Data Mining Untuk Prediksi Kanker Payudara Menggunakan Algoritma Klasifikasi

Raden Tio Putra Sudewo<sup>1\*</sup>, Yovi Pratama<sup>2</sup>, Elvi Yanti<sup>3</sup>

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dinamika Bangsa, Jambi, Indonesia

<sup>1\*</sup>radentio88@gmail.com, <sup>2</sup>yovi.pratama@gmail.com, <sup>3</sup>elvite92@gmail.com

### Abstract

*Breast cancer is one of the cancers that attack the tissue in the breast which is more often experienced by women but does not rule out also attacking men (very rare). The application of data mining in researching breast cancer patients is to make a foundation in knowing the accuracy of the model built to determine breast cancer patients with the classification of "malignant" or "benign" using naïve bayes and C4.5. The test was carried out with several experiments, namely the calculation of split data 60%, 70%, and 80 %. Testing data on breast cancer has a high accuracy at 70% split data using C4.5 algorithm and has a higher accuracy of 99.4975% compared with other testing modes, while the lowest accuracy is 80% split data on naïve bayes algorithm is 95.1648%. In the calculation that compares the algorithm C4.5 and naïve bayes algorithm, it can be said that the algorithm C4.5 as an effective algorithm either from the calculation or the final result where the test can be used as a foundation related to breast cancer considering the results of the accuracy of the algorithm C4. 5 touched 99.4975%.*

**Keywords:** Data Mining, Naïve Bayes, C4.5 Algorithm, Breast Cancer

### Abstrak

Kanker payudara merupakan salah-satu kanker yang menyerang jaringan pada payudara yang mana lebih sering dialami oleh perempuan tetapi tidak menutup kemungkinan juga menyerang laki-laki (sangat jarang terjadi). Penerapan data mining dalam meneliti pasien kanker payudara adalah untuk menjadikan sebuah landasan dalam mengetahui akurasi dari model yang di bangun guna menentukan penderita kanker payudara dengan penggolongan "ganas" atau "jinak" dengan menggunakan naïve bayes dan C4.5. Pengujian dilakukan dengan beberapa eksperimen yaitu perhitungan split data 60%, 70%, dan 80 %. Pengujian data mengenai kanker payudara memiliki akurasi yang tinggi pada split data 70 % menggunakan algoritma c4.5 dan memiliki accuracy yang lebih tinggi yaitu 99.4975 % dibandingkan dengan mode pengujian lainnya, sedangkan accuracy terendah adalah split data 80 % pada algoritma naïve bayes yaitu 95.1648%. Pada perhitungan yang membandingkan algoritma C4.5 dan algoritma naïve bayes, dapat dikatakan bahwa algoritma C4.5 sebagai algoritma yang efektif baik dari perhitungan ataupun hasil akhir yang mana pengujian tersebut dapat dijadikan suatu landasan terkait kanker payudara mengingat hasil akurasi dari algoritma C4.5 menyentuh 99.4975 % sedangkan hasil dari algoritma naïve bayes hanya menyentuh 95.8944 %.

**Kata kunci:** Data Mining, Naïve Bayes, Algoritma C4.5, Kanker Payudara

© 2023 Jurnal Pustaka Data

### 1. Pendahuluan

Kemajuan teknologi saat ini tidak terlepas dari kehidupan masyarakat. Berbagai informasi

bermunculan di berbagai belahan dunia yang kini dapat kita ketahui secara instan berkat kemajuan teknologi (globalisasi). Tentunya kemajuan teknologi tersebut telah menyebabkan perubahan yang luar

biasa dalam kehidupan masyarakat dengan segala peradaban dan budayanya. Perubahan ini juga sangat mempengaruhi transformasi nilai-nilai yang ada di masyarakat, tidak terlepas dalam aspek Kesehatan salah satunya mengenai kanker payudara [1].

### 1.1 Data Mining

Data mining didefinisikan sebagai seperangkat prosedur yang berguna untuk menemukan dan mencari nilai dalam bentuk informasi serta hubungan kompleks yang telah disimpan dalam database hingga saat ini. Ekstraksi sampel informasi data yang berguna untuk memanipulasi data menjadi informasi baru dan lebih berguna yang ditujukan untuk mengekstrak sampel berharga atau menarik yang diperoleh dari data database. [2]

Data mining adalah analisis yang memeriksa kumpulan data untuk menemukan hubungan yang tidak terduga dan merangkum data dengan cara yang berbeda dengan cara yang berbeda dari sebelumnya, yang mudah dipahami dan berguna bagi pemilik data. [3]

Data mining adalah proses menemukan hubungan baru yang bermakna, pola dan kebiasaan dengan menyortir sebagian besar data yang disimpan dalam media penyimpanan menggunakan teknologi pengenalan pola, seperti teknik statistik dan matematika. [4]

### 1.2 Kanker Payudara

Kanker payudara merupakan kanker paling mematikan setelah kanker serviks. sel kanker payudara pertama dapat tumbuh menjadi tumor 1cm dalam 8-12 tahun, Sel kanker ini berada di kelenjar susu. Sel kanker payudara ini dapat menyebar melalui aliran darah ke seluruh tubuh. [5]

Kanker adalah penyakit tidak menular yang ditandai dengan pertumbuhan sel yang tidak normal atau terus menerus dan tidak terkendali, yang dapat merusak jaringan di sekitarnya dan menyebar ke tempat yang jauh dari asal. kanker dapat ditemukan baik pada pria maupun wanita. Perubahan tubuh yang disebabkan oleh kanker mudah dilihat dengan mata telanjang. Perubahan fisik yang terjadi ini dapat berupa pengobatan seperti operasi untuk mengangkat sel kanker pada wanita penderita kanker payudara atau lebih dikenal dengan mastektomi [6]. Perubahan fisik lainnya yang disebabkan oleh kemoterapi terlihat dengan mata telanjang pada pria dan Wanita. Kanker payudara merupakan kanker nomor satu di Indonesia dan merupakan salah satu penyebab utama kematian akibat kanker. Pada tahun 2020 jumlah kasus baru kanker payudara mencapai 68.858 kasus (16,6%) dari total 396.914 kasus baru kanker di Indonesia. Sedangkan, mengenai jumlah kematiannya sudah mencapai lebih dari 22 ribu jiwa kasus. Dengan kasus yang kian banyak, penerapan sebuah metode sangat diperlukan supaya meminimalisir kesalahan dalam menentukan keputusan, Sehingga, dengan metode

yang tepat dalam proses implementasinya, dapat menggunakan sebuah metode yaitu metode naïve bayes dan C4.5. [7].

### 1.3 Algoritma Naïve Bayes

Algoritma Naïve Bayes merupakan salah satu dari algoritma yang termasuk dalam pengklasifikasi. Naïve bayes adalah pengklasifikasi dengan probabilitas sederhana dan model statistik berdasarkan teorema Bayes dengan dengan asumsi bahwa setiap atribut adalah independent [8]

$$P(H|X).P(H)$$

$$\dots\dots\dots(1)$$

$$P(X)$$

Dimana persamaan Teori Bayes Tersebut adalah sebagai berikut :

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

$P(H|X)$  : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

$P(H)$  : Probabilitas hipotesis H (prior probabilitas)

$P(X|H)$  : Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$  : Probabilitas X

Untuk menjelaskan metode Naive Bayes, perlu diperhatikan bahwa proses klasifikasi memerlukan beberapa panduan untuk menentukan kelas mana yang sesuai untuk sampel yang akan dianalisis.

Kinerja algoritma naïve bayes untuk klasifikasi kanker payudara memberikan nilai yang baik, dengan persentase rata-rata data yang terklasifikasi benar sebesar 96,9%, sedangkan persentase rata-rata data yang salah klasifikasi hanya 3,1%. Akurasi yang diperoleh dari pemodelan algoritma C4.5 dengan pembobotan atribut, yaitu sebesar 98,57%. Akurasi tersebut merupakan hasil dari kesesuaian antara prediksi klasifikasi dan hasil klasifikasi. [9]

### 1.4 Algoritma C.45

Algoritma C4.5 merupakan algoritma klasifikasi pohon keputusan yang banyak digunakan karena memiliki keunggulan utama dibanding algoritma yang lainnya. Kelebihan dari algoritma C4.5 dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima dan efisien dalam menangani dua atribut yang bertipe diskrit dan numerik. Dalam membangun pohon, algoritma C4.5 membaca semua sampel data pelatihan dari memori dan memuat nya ke dalam memori secara keseluruhan pada waktu yang bersamaan. Secara umum algoritma C4.5 untuk membangun pohon keputusan sebagai berikut: [10]

1. Pilih atribut sebagai root

2. Buat cabang untuk masing-masing nilai

3. Bagi atribut terpilih dalam cabang

4. Ulangi proses untuk masing-masing cabang sampai semua atribut terpilih pada cabang memiliki kelas yang sama.

Untuk memilih atribut dengan akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti tertera dalam persamaan 2.2 berikut: [11]

$$\text{Entropy}(S) = \sum -p_i * \log_2 p_i \quad (2)$$

Keterangan:

S: Himpunan kasus A: Atribut

N: Jumlah partisi atribut A

|S<sub>i</sub>|: Jumlah kasus pada partisi ke-i

|S|: Jumlah kasus dalam S

Dari keterangan di atas akan diperoleh nilai gain dan atribut yang paling tertinggi. Gain adalah salah satu atribut selection measure digunakan untuk memilih test atribut tiap node pada tree. Atribut dengan information gain tertinggi dipilih sebagai test atribut dari suatu node, untuk itu penghitungan nilai entropi dapat dilihat pada persamaan 2.3 berikut:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \quad (3)$$

Keterangan:

S: Himpunan kasus A: Atribut

N: Jumlah partisi S

P<sub>i</sub>: Proporsi dari S<sub>i</sub> terhadap S

Kemudian lakukan pengulangan dalam perhitungan gain dan entropy sampai semua record ter partisi. Proses partisi pohon keputusan akan berhenti saat:

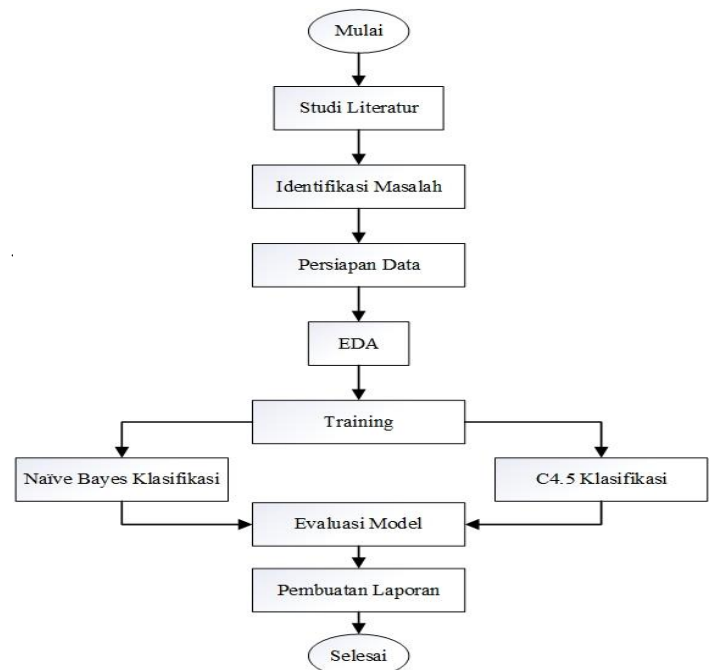
1. Semua record dalam simpul N mendapat kelas yang sama.
2. Tidak ada atribut di dalam record yang di partisi lagi dan tidak ada yang kosong.

Sehingga dari beberapa penelitian mengenai kanker payudara yang sudah dicantumkan di atas dapat disimpulkan bahwa penggunaan kedua metode tersebut akan menjadi perbandingan mengingat hasil dari kedua metode tersebut yang tidak terlalu jauh. Sehingga hal tersebut yang melandasi penulis dalam membandingkan kedua metode tersebut untuk memperoleh hasil dengan persentase yang tinggi dalam penelitian ini.

## 2. Metode Penelitian

Metodologi penelitian berguna sebagai fungsi penting untuk mengumpulkan informasi/data yang dibutuhkan untuk memecahkan berbagai jenis masalah dan berguna untuk memberikan solusi atas masalah yang ada [12].

Kerangka penelitian mencakup langkah-langkah yang diambil dalam pelaksanaan penelitian. Kerangka penelitian dirancang untuk memudahkan kita menerima hasil penelitian, sehingga dapat menyelesaikan penelitian tepat waktu, dan melakukan penelitian sesuai rencana [13].



Gambar 1. Kerangka Kerja Penelitian

### 2.1 Studi Literatur

Di tahap ini penulis melakukan pencarian terhadap landasan-landasan teori atau materi yang telah diperoleh dari berbagai sumber seperti buku, jurnal ilmiah dan juga referensi lainnya untuk melengkapi dalam penelitian baik mengenai konsep dan teori sehingga memiliki landasan yang tepat dan relevan.

### 2.2 Identifikasi Masalah

Pada tahap ini, penulis melakukan identifikasi terkait masalah yang di angkat pada penelitian ini yaitu mengenai penderita kanker payudara sehingga peneliti mampu menerapkan sebuah metode yang tepat guna terselesaikannya masalah terkait.

### 2.3 Persiapan Data

Pada tahap persiapan data, penulis melakukan beberapa metode yang akan digunakan dalam pengumpulan data yang akan dijadikan dasar dalam penelitian ini, dan berikut beberapa metode yang akan di gunakan dalam penelitian ini, yaitu : [14]

#### a. Penggunaan data dari sumber publik

Pada penelitian ini, sumber data yang digunakan adalah data yang diambil dari website Kaggle.com (<https://www.kaggle.com/datasets/yasserrh/breast-cancer-dataset?resource=download>) yang mana dengan beberapa tahapan dan pertimbangan penggunaan data terkait sangat relevan dalam berjalannya penelitian ini. [15]

#### b. Penelitian Kepustakaan (Library Research)

Untuk mendukung penelitian yang akan dilakukan, penulis melakukan penelusuran kepustakaan yaitu dengan mencari data dari buku-buku penelitian sejenis dan jurnal yang berkaitan dengan dengan metode yang digunakan oleh penulis yang digunakan dalam penelitian ini yaitu metode Naïve bayes dan C4.5.

#### c. Data Cleaning

Proses pembersihan data dari noise dan data yang tidak konsisten sehingga dalam tahap ini pengolahan bisa menjadi optimal dan memiliki hasil yang akurat dan dalam penelitiannya proses data cleaning membantu mengelola dari kesalah data error atau tidak bisa di proses

#### d. Data Transformation

Proses mengubah data menjadi data yang tepat untuk proses pengolahan. Proses konversi data kedalam format tertentu sehingga data tepat sasaran untuk data mining dan pengolahannya.

#### e. Data Integration

Proses tersebut menggabungkan data dari sejumlah sumber yang berbeda. Yang memiliki asumsi bahwa perbandingan atau penambahan data menjadi landasan dalam pengolahan

#### f. Data Selection

Proses pemilihan data dari database sangat tepat untuk keperluan analisis. Yang mana dengan tahapan ini merupakan tahap yang vital karena merupakan landasan dalam penelitian ini

### 2.4 EDA (Exploratory Data Analysis)

Exploratory Data Analysis adalah proses penelitian pertama untuk mengidentifikasi pola, menemukan outlier, menguji hipotesis dan memverifikasi asumsi. EDA membantu pengguna mendeteksi kesalahan dari awal, dapat mengidentifikasi anomali, mengetahui hubungan antar data dan mengeksplorasi faktor-

faktor penting dari data. Proses EDA ini sangat berguna dalam proses analisis statistik. [16]

Setidaknya ada empat teknik Exploratory Data Analysis yang umum digunakan oleh ahli data, antara lain:

#### a. Univariate Non-Graphical

Fokus utama dari Univariate Non-Graphical pada EDA adalah untuk mengetahui distribusi/data sampel yang mendasari dan membuat pengamatan tentang populasi. Deteksi outlier juga merupakan bagian dari analisis. Ciri-ciri distribusi populasi antara lain: Central tendency (mean, median, dan modus), spread (variabilitas, varian dan standar deviasi), bentuk distribusi (skewness dan kurtosis).

#### b. Univariate Graphical

Sementara data non-grafis tidak menunjukkan visualisasi lengkap dari data yang diolah, dalam metode grafis peneliti menampilkan data secara visual. Data yang ditunjukkan dapat berupa stem-and-leaf plots, box plots, dan histogram.

#### c. Multivariate Non-Graphical

Multivariate data memiliki banyak variabel. Metode ini menunjukkan hubungan antara dua atau lebih variabel data dengan menggunakan statistik atau tabulasi silang.

#### d. Multivariate Graphical

Teknik EDA ini menggunakan grafik untuk menunjukkan hubungan antara dua atau lebih kumpulan data. Grafis multivariate yang paling sering digunakan antara lain adalah bar chart, bar plot, heat map, bubble chart, run chart, multivariate chart, dan scatter plot.

### 2.5 Training

Pada bagian ini dilakukan proses training menggunakan algoritma klasifikasi naïve bayes untuk membuat sebuah bentuk classifier pada data pasien yang sudah ada. bentuk ini merupakan implementasi pengetahuan yang akan digunakan dalam penelitian ini.

### 2.6 Naïve bayes Klasifikasi

Pada tahap ini dilakukan klasifikasi data menggunakan tool WEKA pada data pasien yang diuji menggunakan model klasifikasi yang telah dibuat. sehingga akan menjadi perbandingan dengan metode C4.5.

## 2.7 C4.5 Klasifikasi

Pada tahap ini dilakukan klasifikasi data menggunakan tool WEKA pada data pasien yang diuji menggunakan model klasifikasi yang telah dibuat sehingga akan menjadi perbandingan dengan metode Naïve Bayes.

## 2.8 Evaluasi Model

Hasil analisis dari perancangan model akan dievaluasi keakuratan datanya dari model yang diperoleh dengan menghitung naïve bayes menggunakan Accuracy, Precesion, Recall, dan F-Mesquare

## 2.9 Penulisan Laporan

Setelah semua langkah penelitian selesai, laporan studi akan didokumentasikan sehingga dapat digunakan di masa mendatang. Baik oleh peneliti itu sendiri maupun oleh peneliti lainnya.

## 3. Hasil dan Pembahasan

### 3.1. Analisis

Peneliti mengemukakan hasil analisis dataset yang mana dengan penerapan menggunakan metode yang tepat maka akan memperoleh hasil yang sesuai dengan tujuan penelitian ini. Adapun data yang digunakan adalah data yang berhubungan dengan salah satu penyakit yang mematikan yaitu kanker lebih tepatnya dalam penelitian ini adalah kanker payudara. Maka dari itu di harapkan hasil dari analisis ini mampu menjadi landasan dalam penelitian terkait kanker payudara untuk kedepannya

### 3.2. Dataset

Pada penelitian ini, dataset yang digunakan adalah dataset yang di peroleh melalui website Kaggle.com dengan jumlah data sebanyak 569 data dan dapat diakses pada link : <https://www.kaggle.com/datasets/yasserh/breast-cancer-dataset?resource=download>. Pada dataset tersebut terdapat 32 atribut yaitu :

Tabel 1. Tabel Atribut dalam penelitian

<i>id</i>	<i>diagnosis</i>	<i>texture_mean</i>	<i>radius_mean</i>
<i>perimeter_mean</i>	<i>area_mean</i>	<i>smoothness_mean</i>	<i>compactness_mean</i>
<i>concavity_mean</i>	<i>Concave points_mean</i>	<i>symmetry_mean</i>	<i>fractal_dimension_mean</i>

<i>radius_se</i>	<i>texture_se</i>	<i>perimeter_se</i>	<i>area_se</i>
<i>smoothness_se</i>	<i>compactness_se</i>	<i>concavity_se</i>	<i>concave points_se</i>
<i>symmetry_se</i>	<i>fractal_dimension_se</i>	<i>radius_worst</i>	<i>texture_worst</i>

Setelah itu akan dilakukan pengujian dengan menggunakan metode yang akan di terapkan dalam penelitian ini dengan menggunakan weka supaya data yang digunakan memiliki potensi dalam menentukan “ganas” dan “jinak” bagi penderita kanker payudara.

Tabel 2. Tabel Atribut Correlation Analysis

Rank	Bobot	Nama Atribut
1	1	diagnosis
2	-0.73	radius_mean
3	-0.743	perimeter_mean
4	-0.709	area_mean
5	-0.597	compactness_mean
6	-0.696	concavity_mean
7	-0.777	concave point_mean
8	-0.567	radius_se
9	-0.556	perimeter_se
10	-0.548	area_se
11	-0.776	radius_worst
12	-0.783	perimeter_worst
13	-0.734	area_worst
14	-0.591	compactness_worst
15	-0.66	concavity_worst
16	-0.794	concave point_worst

Bahwa beberapa atribut memiliki korelasi yang kuat terhadap kelasnya, sehingga pada penjabaran diatas pula disimpulkan bahwa atribut yang mendekati nilai -1 dan 1 adalah atribut yang mempengaruhi hasil perhitungan penelitian (Accuracy).

Tabel 1. Perbandingan Evaluasi Akurasi Weka

Model Evaluasi	Akurasi	Jumlah Data	Persentase
<i>Algoritma Naïve Bayes (Split Data 60%)</i>	<i>Correctly Classified Instances</i>	327	<b>95.8944 %</b>
	<i>Incorrectly Classified Instances</i>	14	4.1056 %
<i>Algoritma Naïve Bayes (Split Data 70%)</i>	<i>Correctly Classified Instances</i>	380	<b>95.4774 %</b>
	<i>Incorrectly Classified Instances</i>	18	4.5226 %
<i>Algoritma Naïve Bayes (Split Data 80%)</i>	<i>Correctly Classified Instances</i>	433	<b>95.1648 %</b>
	<i>Incorrectly Classified Instances</i>	22	4.8352 %
<i>Algoritma C4.5 (Split Data 60%)</i>	<i>Correctly Classified Instances</i>	337	<b>98.827 %</b>
	<i>Incorrectly Classified Instances</i>	4	1.173 %
<i>Algoritma C4.5 (Split Data 70%)</i>	<i>Correctly Classified Instances</i>	396	<b>99.4975 %</b>
	<i>Incorrectly Classified Instances</i>	2	0.5025 %
<i>Algoritma C4.5 (Split Data 80%)</i>	<i>Correctly Classified Instances</i>	451	<b>99.1209 %</b>
	<i>Incorrectly Classified Instances</i>	4	0.8791 %

Setelah dilakukan analisis klasifikasi naïve bayes dan C4.5 pada tool WEKA menggunakan, split data 60 %, 70% dan 80 % maka diperoleh akurasi tertinggi yaitu dengan menggunakan Algoritma C4.5 (Split Data 70%) dengan persentase akurasi yaitu 99.4975 % untuk Correctly Classified Instances dan 0.5025 % untuk Incorrectly Classified Instances pada Algoritma C4.5 (Split Data 70%).

Tabel.4 Peforma Klasifikasi Menggunakan Weka

Mode Pengujian	Akurasi	Precision	Recall	F-Measure
<i>Algoritma Naïve Bayes (Split Data 60%)</i>	95.8944 %	92,63157	92,63157	92,63157
<i>Algoritma Naïve Bayes (Split Data 70%)</i>	95.4774 %	93.38842	91.86991	92.62294
<i>Algoritma Naïve Bayes (Split Data 80%)</i>	95.1648 %	92.95775	91.66667	92.3077
<i>Algoritma C4.5 (Split Data 60%)</i>	98.827 %	97.8947	97.8947	97.8947
<i>Algoritma C4.5 (Split Data 70%)</i>	<b>99.4975 %</b>	99.18699	<b>99.18699</b>	<b>99.18699</b>
<i>Algoritma C4.5 (Split Data 80%)</i>	99.1209 %	<b>99.29577</b>	97.916	98.60106

Hasil uji dari algoritma naïve bayes dan c4.5 yang menggunakan data yang diperoleh dari Kaggle.com. untuk menguji algoritma yang di pakai dalam mengklasifikasikan penderita kanker payudara, pada penelitian ini menggunakan Accuracy, Precision, Recall, dan F-Measure.

#### 4. Kesimpulan

Pengujian dilakukan dengan beberapa eksperimen yaitu perhitungan split data 60%, 70%, dan 80 %.

Pengujian data mengenai kanker payudara memiliki akurasi yang tinggi pada split data 70 % menggunakan algoritma c4.5 dan memiliki accuracy yang lebih tinggi yaitu 99.4975 % dibandingkan dengan mode pengujian lainnya, sedangkan accuracy terendah adalah split data 80 % pada algoritma naïve bayes yaitu 95.1648%.

Pengujian menggunakan perhitungan split data 60%, 70%, dan 80 %. memiliki hasil yang berbeda pada pengujian yang meliputi Precision, Recall, dan F-Measure.

Pada Precision, dapat disimpulkan bahwa split data 80 % pada algoritma c4.5 memiliki Precision yang lebih tinggi yaitu 99.29577 dibandingkan dengan mode pengujian lainnya, sedangkan Precision terendah adalah split data 60 % pada algoritma naïve bayes yaitu 92,63157.

Pada recall dapat disimpulkan bahwa split data 70 % pada algoritma c4.5 memiliki Recall yang lebih tinggi yaitu 99.18699 dibandingkan dengan mode pengujian lainnya, sedangkan Recall terendah adalah split data 80 % pada algoritma naïve bayes yaitu 91.66667.

Pada F-Measure, dapat disimpulkan bahwa split data 70 % pada algoritma c4.5 memiliki F-Measure yang lebih tinggi yaitu 99.18699 dibandingkan dengan mode pengujian lainnya, sedangkan F-Measure terendah adalah split data 80 % pada algoritma naïve bayes yaitu 91.66667.

Pada perhitungan yang membandingkan algoritma C4.5 dan algoritma naïve bayes. dapat dikatakan bahwa algoritma C4.5 sebagai algoritma yang efektif baik dari perhitungan ataupun hasil akhir yang mana pengujian tersebut dapat dijadikan suatu landasan terkait kanker payudara mengingat hasil akurasi dari algoritma C4.5 menyentuh 99.4975 % sedangkan hasil dari algoritma naïve bayes hanya menyentuh 95.8944 %

#### Daftar Rujukan

- [1] H. Oktavianto and R. P. Handri, "Analisis Klasifikasi Kanker Payudara Menggunakan Algoritma Naive Bayes," *INFORMAL Informatics J.*, vol. 4, no. 3, p. 117, 2020, doi: 10.19184/isj.v4i3.14170.

- [2] A. Wijaya, A. Mutatkin Bakti, T. Informatika, U. Bina Darma, J. A. Jendral Yani No, and P. Sumatera Selatan, "Implementasi Data Mining Menggunakan Algoritma C4," vol. 5, pp. 61–67, 2023.
- [3] D. Diy, "Analisis Data Mining Untuk Memprediksi Lama Perawatan Pasien Covid-19 Bianglala Informatika," vol. 10, no. 1, pp. 21–29, 2022.
- [4] S. Hendrian, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan," *Fakt. Exacta*, vol. 11, no. 3, pp. 266–274, 2018, doi: 10.30998/faktorexacta.v11i3.2777.
- [5] H. Dewi, "Analisis risiko kanker payudara berdasar riwayat pemakaian kontrasepsi hormonal dan usia," *J. Berk. Epidemiol.*, vol. 3, no. 1, pp. 12–23, 2015.
- [6] K. Aliwikarta *et al.*, "Prevalensi Penyakit Kanker di Indonesia Berdasarkan Pola Konsumsi Pangan dan Gaya Hidup Cancer prevalence in Indonesia Based on Food Consumption Patterns and Lifestyle Newly Weds Foods Asia Pacific 2," *J. Mutu Pangan*, vol. 3, no. 1, pp. 71–78, 2016.
- [7] B. Julia, H. Sumarti, and H. H. Kusuma, "Klasifikasi Tumor Payudara Berbasis Ciri Tekstur pada Citra Mammografi Menggunakan Metode Naive Bayes," *MIND (Multimedia Artif. Intell. Netw. Database) J.*, vol. 7, no. 2, pp. 165–176, 2022, [Online]. Available: <https://doi.org/10.26760/mindjournal.v7i2.165-176>.
- [8] R. Rachman and R. N. Handayani, "Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM," *J. Inform.*, vol. 8, no. 2, pp. 111–122, 2021, doi: 10.31294/ji.v8i2.10494.
- [9] Naomi Chatrina Siregar, Riki Ruli A. Siregar, and M. Yoga Distra Sudirman, "Implementasi Metode Naive Bayes Classifier (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ)," *J. Teknol.*, vol. 3, no. 1, pp. 102–110, 2020.
- [10] R. Haqmanullah Pambudi, B. Darma Setiawan, and Indriati, "Penerapan Algoritma C4.5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 7, pp. 2637–2643, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [11] I. M. Rahmawati, H. Wibowo, and P. Siswa, "Penerapan Algoritma C4.5 Dalam Memprediksi Peminatan Mahasiswa Program Studi Teknik Informatika," vol. 17, no. x, pp. 379–389, 1978.
- [12] H. Widyastono, "398-Article Text-1279-1-10-20161020.pdf," *Pendidikan Dan Kebudayaan*. pp. 1–17, 20017.
- [13] M. R. Fadli, "Memahami desain metode penelitian kualitatif," *Humanika*, vol. 21, no. 1, pp. 33–54, 2021, doi: 10.21831/hum.v21i1.38075.
- [14] T. Dan, I. T. Information, and T. Pada, "Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Memprediksi Penjualan Alat – Alat Telekomunikasi Dan IT (," *Jupiter*, vol. 15, pp. 597–606, 2023.
- [15] O. Solihin, "Implementasi Big Data pada Sosial Media sebagai Strategi Komunikasi Krisis Pemerintah," *J. Common*, vol. 5, no. 1, pp. 56–66, 2021.

- [16] E. D. Wahyuni, A. A. Arifiyanti, and M. Kustyani, "Exploratory Data Analysis dalam Konteks Klasifikasi Data Mining," *Pros. Nas. Rekayasa Teknol. Ind. dan Inf. XIV Tahun 2019*, vol. 2019, no. November, pp. 263–269, 2019, [Online]. Available: <http://journal.itny.ac.id/index.php/ReTII>.