

Komparasi Kinerja Algoritma *Mechine Learning* Untuk Mengklasifikasi Penyakit Kanker Payudara

Raynaldo Pakan^{1*}, Evangs Mailoa²

Teknik Informatika, Universitas Kristen Satya Wacana, Salatiga, Indonesia
 e-mail *Corresponding Author*: 672019042@student.uksw.edu

Abstract

Breast cancer is a malignant neoplasm that originates from the parenchyma and produces a frequency of death that is a major cause of concern in the world. Breast cancer has a very significant impact on sufferers, including physical changes, emotional stress, and even death. In general, this disease is divided into two types, namely benign and malignant, therefore this research aims to identify how benign and malignant breast cancer is. This research processes data related to comparing the performance of Neural Network, Naive Bayes, and Decision Tree algorithms. The parameters analyzed include accuracy, precision and recall values to evaluate the performance of the three algorithms in the context of the analysis carried out. The results of the comparison of these three algorithms are that the Naive Bayes algorithm is more accurate for use in classifying breast cancer compared to the Neural Network and Decision Tree algorithms. The Naive Bayes algorithm has a higher accuracy rate of 96.4%, 96.7%, 96.6%, and 96.6%.

Keywords: *Breast cancer; Neural Network; Naive Bayes; Decesion Tree; Comparison*

Abstrak

Kanker payudara adalah suatu penyakit neoplasma ganas yang berasal dari *parenchyma* dan menghasilkan frekuensi kematian yang menjadi penyebab utama kekhawatiran di dunia. Kanker payudara membawa dampak yang sangat signifikan terhadap penderitanya termasuk perubahan fisik, stres emosional, sampai pada kematian. Secara umum Penyakit ini dibedakan menjadi dua macam yaitu jinak (*benign*) dan ganas (*malignant*), oleh karena itu penelitian ini bertujuan untuk mengidentifikasi seberapa jinak dan ganasnya kanker payudara. Penelitian ini memproses data terkait dengan perbandingan kinerja algoritma *Neural Network*, *Naive Bayes*, dan *Decision Tree*. Parameter-parameter yang dianalisis mencakup nilai keakuratan, presisi, dan recall untuk mengevaluasi performa ketiga algoritma dalam konteks analisis yang dilakukan. Adapun hasil perbandingan dari ketiga algoritma ini adalah Algoritma *Naive Bayes* lebih akurat untuk digunakan dalam melakukan klasifikasi kanker payudara dibandingkan dengan algoritma *Neural Network* dan *Decision Tree*. Algoritma *Naive Bayes* memiliki tingkat akurasi yang lebih tinggi sebanyak 96,4%, 96,7%, 96,6%, dan 96,6%.

Kata Kunci : *Kanker Payudara; Neural Network; Naive Bayes; Decesion Tree; Perbandingan.*

1. Pendahuluan

Kesehatan adalah hal yang paling utama dalam kehidupan manusia. Namun, hal itu tidak lepas dari adanya sebuah penyakit terumata penyakit kanker payudara yang merupakan penyakit neoplasma ganas yang berasal dari *parenchyma* dan menghasilkan frekuensi kematian yang menjadi penyebab utama kekhawatiran di dunia. Kanker atau sering disebut dengan tumor secara umum dibedakan menjadi dua macam yaitu jinak (*benign*) dan ganas (*malignant*) [1]. Para ahli dan dokter akan memberikan perawatan yang berbeda pada setiap pasien. Kunci untuk bertahan hidup penderita kanker payudara adalah dengan mendeteksi kanker payudara sedini mungkin, sebelum sel kanker menyebar ke bagian-bagian tubuh lainnya[2]. Penyakit kanker ini adalah salah satu penyakit yang mematikan apabila terlambat dideteseksi dan juga terlambat untuk ditangani. Penyakit ini terjadi hampir seluruhnya pada wanita, tetapi pria juga dapat mengidap penyakit kanker payudara ini. Penderita kanker payudara biasanya memiliki ciri psikologis seperti gegar otak, ketakutan, depresi, dan panik [3]. Penyebab timbulnya penyakit ini dikarenakan terlalu sering mengomsumsi makanan yang

mengandung minyak dan juga beban pikiran yang teralau banyak. Salah satu penyebab utama kematian diakibatkan oleh kanker, khususnya kanker payudara yang umumnya terjadi pada wanita. Secara umum diperkirakan kanker payudara merupakan penyebab kematian tertinggi setelah kanker paru [4]. Data statistik *American Cancer Society* yang menunjukkan bahwa kanker payudara menempati urutan pertama dengan jumlah kasus terbanyak dari seluruh jenis kasus tumor ganas (kanker) di seluruh dunia [5]. Maka dari itu, penelitian ini bertujuan untuk mengidentifikasi seberapa jinak dan ganasnya kanker payudara.

Saat ini, situasi penanganan kanker payudara masih dihadapkan pada beberapa tantangan yang menciptakan kesenjangan antara kondisi sekarang dengan kondisi ideal yang diharapkan. Meskipun telah terjadi kemajuan dalam diagnosis dan pengobatan, namun masih terdapat ketidaksetaraan akses terhadap layanan kesehatan, keterlambatan dalam deteksi dini, dan kurangnya kesadaran masyarakat. Gap ini memberikan gambaran terkait dengan masalah terukur dalam upaya menangani kanker payudara secara optimal, seperti perlunya peningkatan edukasi kesehatan, akses yang merata terhadap pelayanan medis, dan peningkatan program deteksi dini. Dengan memahami kesenjangan ini, langkah-langkah yang terukur dan dapat diimplementasikan dapat dirumuskan untuk meningkatkan upaya pencegahan, deteksi, dan penanganan kanker payudara.

Dalam merumuskan solusi untuk mengatasi tantangan penanganan kanker payudara, langkah pertama yang diperlukan adalah mengidentifikasi dan merasionalisasikan konsep-konsep solusi yang dapat efektif mengatasi kesenjangan yang telah diidentifikasi. Dalam proses ini, dilakukan klasifikasi menggunakan metode *neural network*, *naive bayes*, dan *decision tree*. Klasifikasi (pengelompokan) adalah salah satu metode dalam melakukan data mining, klasifikasi tergolong dalam *supervised learning*. Klasifikasi juga merupakan suatu cara pengelompokan yang didasarkan pada ciri-ciri tertentu [6]. Klasifikasi memiliki beberapa metode antara lain *Decision Tree*, *Random Forest*, *Naive Bayes*, *Nearest Neighbour*, *C4.5*, *Neural Network*, dan beberapa algoritma lainnya. Namun, dalam penelitian ini digunakan algoritma *neural network*, *Naive Bayes*, dan *Decision Tree*. Algoritma *Neural Network*, *Naive Bayes*, dan *Decision Tree* adalah algoritma yang cukup efektif dan paling banyak digunakan dalam klasifikasi data. Pada penelitian sebelumnya yang berjudul “Analisis Algoritma Klasifikasi *Neural Network* Untuk Diagnosis Penyakit Kanker Payudara”, mendapat sebuah hasil *accuracy* = 71,83%, *precision* = 81,08% dan *recall* = 69,17% dengan AUC sebesar = 0,806, [7].

Dalam penelitian ini dilakukan perbandingan antara algoritma *neural network*, *naive bayes*, dan juga algoritma *decision tree* dalam mengklasifikasi penyakit kanker payudara, guna untuk mendapatkan hasil yang lebih akurat dengan judul “Komparasi Kinerja Algoritma *Machine Learning* Untuk Mengklasifikasi Penyakit Kanker Payudara”. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi positif terhadap bidang kesehatan dan pengembangan teknologi machine learning dalam diagnosis penyakit.

2. Tinjauan Pustaka

Pada penelitian terdahulu yang berjudul “Analisis Klasifikasi Kanker Payudara Menggunakan Algoritma *Naive Bayes*”, penulis dalam penelitian ini adalah Hardian Oktavianto yang menjelaskan tentang dampak dari penyakit kanker payudara yang merupakan salah satu faktor penyebab kematian tertinggi di kalangan wanita. Penelitian ini menggunakan Algoritma *Naive Bayes* sebagai metode dalam mengklasifikasi penyakit kanker payudara. Hasil dari penelitian ini adalah performa algoritma *naive bayes* untuk klasifikasi kanker payudara menghasilkan nilai yang baik, dimana rata-rata persentase data yang terklasifikasi dengan benar mencapai 96.9% dan rata-rata persentase data terklasifikasi salah hanya 3.1%. Sedangkan tingkat efektivitas klasifikasi dengan *naive bayes* ini termasuk tinggi, dimana rata-rata nilai *precision* dan *recall* berada di sekitar 0.96. Nilai *precision* dan *recall* paling tinggi yaitu ketika data uji menggunakan *percentage split* 40% dengan nilai masing-masing secara berurutan mencapai 0.974 dan 0.973. [8]

Pada penelitian terdahulu yang berjudul “Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma *Machine Learning*” menjelaskan tentang dampak penyakit kanker payudara salah satu penyakit non-kulit yang paling berbahaya pada wanita, disebabkan oleh berbagai faktor, mulai dari sel dan saluran hingga jaringan pendukung payudara, selain kulit payudara, juga menjelaskan bahwa Dalam kanker payudara dapat dibedakan menjadi ganas dan jinak. Jika diketahui jenisnya, maka pencegahan dan pengobatan akan segera dilakukan sesuai dengan jenis kanker payudaranya, sehingga dapat

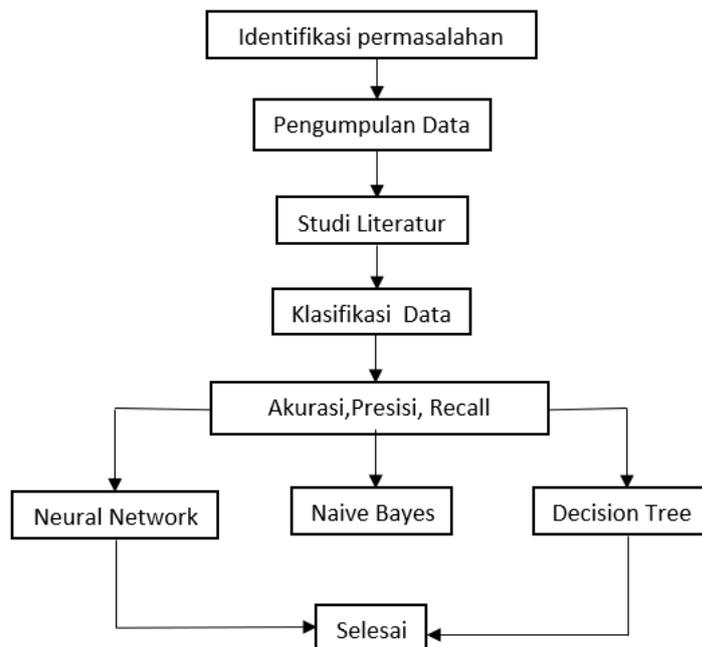
menghindari efek samping pada pasien dan bahkan kematian. Metode yang digunakan dalam penelitian ini adalah algoritma machine learning dengan pengolahan data dilakukan dengan 3 tahap yaitu data cleaning, normalisasi data, dan pembagian data sehingga mendapatkan hasil algoritma *Logistic Regression*, *Decision Tree*, *Naïve Bayes* dan *K-Nearest Neighbor* memiliki nilai akurasi yang sama tinggi yaitu sebesar 95,00%. Penerapan data mining dengan menggunakan metode algoritma *Logistic Regression*, *Decision Tree*, *Naïve Bayes* dan *K-Nearest Neighbor* dapat mempercepat pengambilan keputusan dalam memprediksi dalam klasifikasi penentuan jenis kanker payudara.[9]

Penelitian terdahulu dengan judul “Implementasi Algoritma Genetika pada Struktur Backpropagation Neural Network untuk Klasifikasi Kanker Payudara” menjelaskan tentang kanker payudara adalah jenis kanker paling umum yang diderita kaum wanita. Hingga saat ini, salah satu cara pengobatan yang umum dilakukan adalah dengan pembedahan dan jika perlu dilanjutkan dengan kemoterapi maupun radiasi. Namun pengobatan tersebut tidak akan memberikan dampak yang signifikan jika kankernya sudah mencapai stadium akhir. Penulis dalam penelitian ini adalah Adam Mizza Zamani [10]. Dalam penelitian ini menggunakan algoritma *Neural Network* untuk klasifikasi penyakit kanker payudara. Hasil dari penelitian ini adalah Metode *Neural Network* yang parameternya dioptimalkan menggunakan algoritma genetika terbukti mampu menghasilkan nilai rata-rata akurasi yang tinggi yaitu sebesar 97,00% untuk studi kasus deteksi kanker payudara. Metode ini memiliki keunggulan dalam hal tingkat akurasi jika dibandingkan metode *Naïve Bayesian* dan metode sebelumnya yaitu *Neural Network* menggunakan optimasi *Association Rules*. [10]

Berdasarkan penelitian dan penjelasan sebelumnya, maka dilakukan penelitian baru untuk mengklasifikasi penyakit kanker payudara dengan membandingkan algoritma *Neural Network*, *Naive Bayes*, dan *Decision Tree* guna untuk mendapatkan hasil yang lebih akurat. Penelitian ini dilakukan dengan menggunakan aplikasi orange untuk mendapatkan hasil Akurasi, Presisi, dan Recall dari perbandingan algoritma *Neural Network*, *Naive Bayes*, dan *Decision Tree*.

3. Metodologi

Dalam proses penelitian ini, ada beberapa tahapan yang dilakukan. Mulai dari Identifikasi permasalahan, pengumpulan data, klasifikasi data, perbandingan algoritma *Neural Network*, *Naive Bayes*, dan *Decision Tree* sampai pada selesai



Gambar 1. Bagan Metode Penelitian

Pada penelitian ini dilakukan identifikasi permasalahan dimana pada tahapan awal ini adalah tahap untuk mengetahui masalah yang terjadi dan mencari solusi yang tepat untuk sebuah permasalahan yang ada.

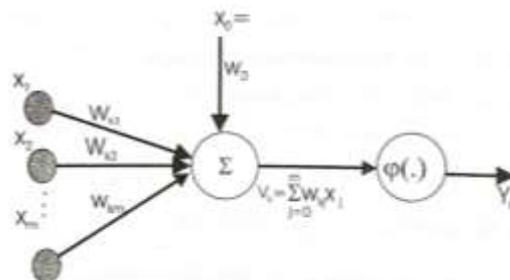
Tahapan kedua dilakukan pengumpulan data. Pada tahap pengumpulan data dilakukan secara studi pustaka dengan mengumpulkan data dari berbagai sumber seperti, Internet dan media cetak. Data yang dimasukan berupa dataset yang di ambil dari [breast-cancer-wisconsin.csv](#). Dataset ini diakses pada 15 Juni 2023, penulis dalam dataset ini adalah Fahminb33 dengan jumlah baris dalam dataset adalah 700 baris dan jumlah kolom pada dataset ini adalah 11 kolom. Berikut dibawah ini ditampilkan sebagian dari dataset *breast-cancer-wisconsin.csv*[11].

Tabel 1. Dataset *breast-cancer-wisconsin*

1	id	thickness	size	shape	adhesion	cell-size	nuclei	chromatin	nucleoli	mitoses	class
2	1000025	5	1	1	1	2	1	3	1	1	2
3	1002945	5	4	4	5	7	10	3	2	1	2
4	1015425	3	1	1	1	2	2	3	1	1	2
5	1016277	6	8	8	1	3	4	3	7	1	2
6	1017023	4	1	1	3	2	1	3	1	1	2
7	1017122	8	10	10	8	7	10	9	7	1	4
8	1018099	1	1	1	1	2	10	3	1	1	2
9	1018561	2	1	2	1	2	1	3	1	1	2
10	1033078	2	1	1	1	2	1	1	1	5	2
11	1033078	4	2	1	1	2	1	2	1	1	2

Pada tahap studi literatur digunakan penelitian sebelumnya yang menggunakan Algoritma *decision tree*, *naive bayes*, dan *K-nearest neighbour* [12], dan juga penelitian sebelumnya yang menggunakan algoritma *neural network* [7], sehingga pada kedua sumber tersebut dapat dijadikan acuan dalam melakukan perbandingan menggunakan algoritma *neural network*, *naive bayes*, dan *decision tree*.

Algoritma *Neural Network* adalah salasatu dari algoritma *mechine laearning* yang dirancang untuk mendorong mesin agar belajar mengenali pola secara mandiri. *Neural Network* (NN) adalah suatu usaha untuk meniru fungsi otak manusia [13]. *Neural network* yaitu metode pembelajaran yang biasa digunakan untuk permasalahan diskrit, real ataupun vektor, ilmu *Neural Network* sudah ada sejak tahun 1943 ketika Warren McCulloch dan Walter Pitts memperkenalkan perhitungan model *neural network* yang pertama [14]. Algoritma ini disebut sebagai *neural network* karena didesain mengikuti cara kerja otak, mengolah dan menyimpan informasi.



Gambar 2. Bentuk Algoritma *Neural Network* [15].

Algoritma *Naive Bayes* adalah algoritma klasifikasi berdasarkan teorema bayes dengan asumsi independen antar prediktor. Secara garis kecil pengklasifikasi *naive bayes* memperkirakan keberadaan fitur tertentu pada kelas tidak terkait dengan keberadaan fitur lainnya. Algoritma *Naive Bayes* dikemukakan ilmuan asal Inggris yang bernama Thomas Bayes, *naive bayes* didasarkan pada teorema bayes yang memiliki kemampuan klasifikasi serupa dengan *decesion tree* dan *neural network* [16]. *Naive Bayes* juga merupakan suatu klasifikasi yang sederhana untuk dapat menghitung seluruh kemungkinan dengan menggabungkan sejumlah kombinasi dan frekuensi suatu nilai dari basis data yang didapatkan [17].

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad \dots\dots\dots (1)$$

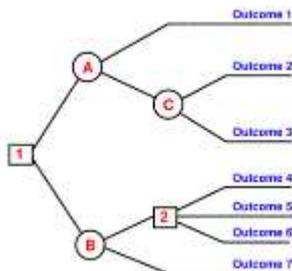
Likelihood
Class Prior Probability
Posterior Probability
Predictor Prior Probability

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

Keterangan:

- P(c|x) : probabilitas posterior kelas (c, target) yang diberikan prediktor (x, atribut).
- P(c) : adalah probabilitas kelas sebelumnya.
- P(x|c) : adalah kemungkinan yang merupakan probabilitas prediktor yang diberikan kelas.
- P(x) : adalah probabilitas prediktor sebelumnya

Algoritma *decision tree* (pohon keputusan) adalah sala satu algoritma yang ada dalam Data mining. Algoritma *decision tree* (pohon keputusan) mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target [19]. Algoritma pertama kali ditemukan oleh Al-Khawarizmi pada sekitar tahun 780M [20]. Algoritma Decision Tree juga merupakan algoritma yang dapat membantu dalam memilih salasatu dari beberapa tindakan. Ada beberapa kelebihan dan kekurangan yang ada pada algoritma ini, Sala satu kelebihanannya adalah mudah di pahami dan dianalisis. Sedangkan kekurangan yang ada pada algoritma ini adalah rentan terhadap kesalahan dalam masalah klasifikasi karena ada banyak pilihan.



Gambar 3. Bentuk Decision Tree (Pohon Keputusan) [21].

Akurasi adalah persentase dari jumlah tuple dalam data uji yang diklasifikasikan dengan benar oleh model klasifikasi. Presisi adalah ukuran kepastian, yaitu berapa persentase tuple yang dilabeli sebagai positif adalah benar pada kenyataannya, sedangkan recall adalah ukuran kelengkapan yaitu berapa persentase tuple positif yang dilabeli sebagai positif [2].

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \cdot 100\% \quad \dots\dots\dots (2)$$

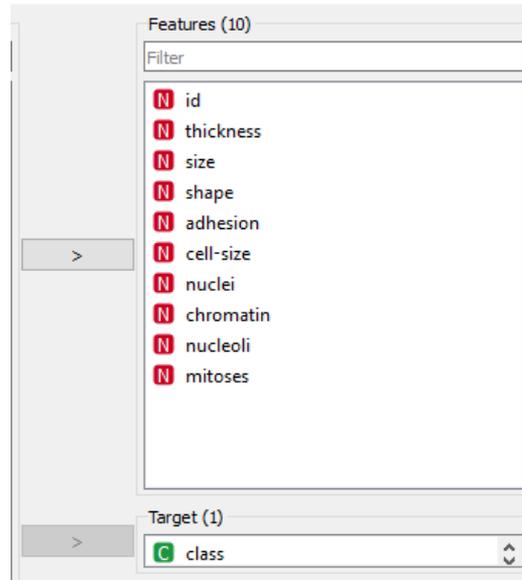
$$Presisi = \frac{TP}{FP+TP} \cdot 100\% \quad \dots\dots\dots (3)$$

$$Recall = \frac{TP}{FN+TP} \cdot 100\% \quad \dots\dots\dots (4)$$

Keterangan:

- TP : Jumlah data positif yang terklasifikasi dengan benar oleh sistem.
- TN : Jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
- FN : Jumlah data negatif namun terklasifikasi salah oleh sistem.
- FP : Jumlah data positif namun terklasifikasi salah oleh sistem.

Pada tahap klasifikasi data, digunakan aplikasi *Orange versi 3.35.0* untuk mengklasifikasikan dan membandingkan algoritma *neural network*, *naive bayes*, dan *decision tree*. Data yang belum berupa dataset dilakukan pengumpulan data dan dikerjakan pada aplikasi *Ms.Excell*. Dan apabila data sudah berupa dataset, dilakukan penginputan data pada aplikasi *Orange* untuk melakukan perbandingan dan klasifikasi data. Pada tahap ini juga dilakukan pengelompokan antara kolom target dan kolom features, yang dimana kolom class dijadikan sebagai target dan kolom yang lainnya tetap dijadikan features. Seperti yang di perlihatkan pada gambar dibawah ini.



Gambar 4. Dataset Seleksi.

Tahapan kelima sebagai tahap akhir dari penelitian ini adalah tahap perbandingan antara algoritma *neural network*, *naive bayes*, dan *decision tree*. Pada tahap ini, dataset yang sudah di input kedalam aplikasi *Orange*, selanjutnya dilakukan perbandingan antara ke-3 algoritma tersebut guna untuk mendapatkan hasil dari penelitian ini.

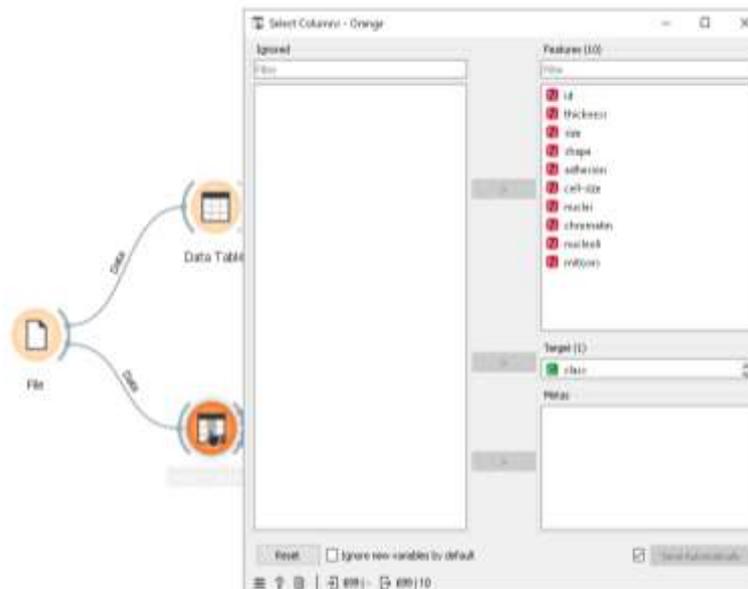
4. Hasil dan Pembahasan

1) Pemrosesan Data

Tabel 2. Sampel data yang diproses

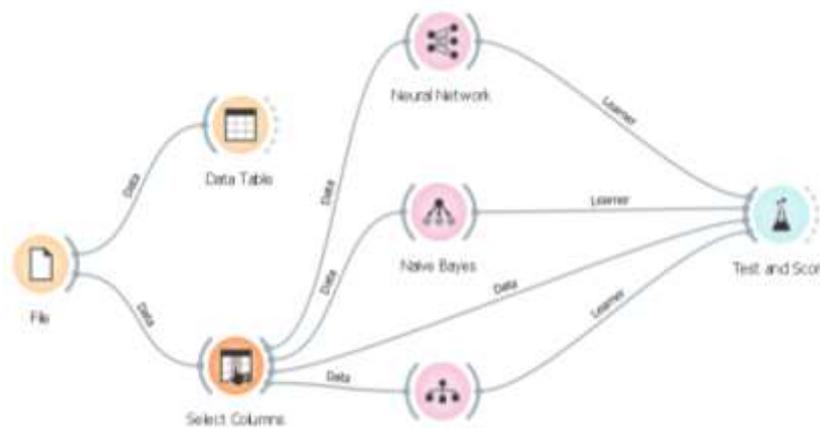
id	thickness	size	shape	adhesion	cell-size	nuclei	chromatin	nucleoli	mitoses	class
1	1000225	5	1	1	1	2	1	3	1	1
2	1002940	5	4	4	5	7	10	3	2	1
3	1005425	5	1	1	1	2	2	1	1	1
4	1006277	6	8	8	1	9	4	1	7	1
5	1007023	4	1	1	3	2	1	3	1	1
6	1007123	8	10	10	8	7	10	9	7	1
7	1008090	1	1	1	1	2	10	3	1	1
8	1008961	2	1	2	1	2	1	1	1	1
9	1009070	2	1	1	1	2	1	1	1	5
10	1009070	4	2	1	1	2	1	2	1	1
11	1009203	1	1	1	1	1	1	1	1	1
12	1006172	2	1	1	1	1	1	2	1	1
13	1041901	5	3	3	3	2	3	4	4	1
14	1043999	1	1	1	1	2	3	3	1	1
15	1044972	5	7	8	10	7	9	5	9	4
16	1047800	7	4	6	4	8	1	4	3	1
17	1048672	4	1	1	1	1	1	2	1	1
18	1049015	4	1	1	1	1	1	1	1	1
19	1050670	10	7	7	6	4	10	4	1	2
20	1060719	6	1	1	1	2	1	1	1	1
21	1064090	7	3	2	10	8	10	5	4	4
22	1064993	10	5	5	3	4	7	7	10	1
23	1066704	3	1	1	1	2	1	2	1	1
24	1067013	8	4	5	1	2	1	7	3	1
25	1069952	1	1	1	1	2	1	1	1	1
26	1080726	5	1	3	4	1	7	3	9	1
27	1080829	5	1	1	1	1	1	1	1	1

Pada Tabel 2 diperlihatkan sampel data yang akan diproses oleh metode di dalam aplikasi orange. Sampel data di atas adalah data yang masih dalam bentuk Microsoft Excel yang nantinya data tersebut akan di input kedalam aplikasi orange untuk diproses oleh metode



Gambar 5. Tampilan aplikasi orange(select columns)

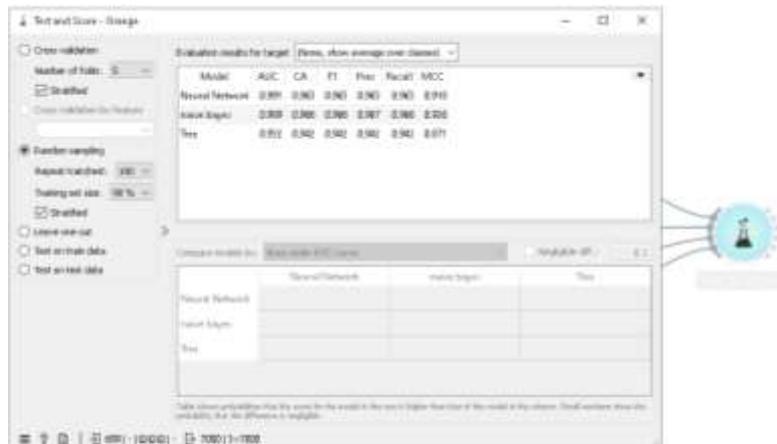
Pada gambar 5 diatas dilakukan proses select columns, pada proses ini data yang masih dalam file Ms.Excel diinput kedalam file dan selanjutnya dilakukan pengelompokan antara kolom target dan kolom features, yang dimana kolom class dijadikan sebagai target dan kolom yang lainnya tetap dijadikan features.



Gambar 6. Data diproses oleh Metode,

Pada gambar 6 diatas, adalah gambar yang memperlihatkan bahwa data sedang diproses oleh algoritma *Neural Network*, *Naive Bayes*, dan *Decision Tree*. Jika proses pada *select columns* telah selesai maka akan dilanjutkan untuk proses data menggunakan *Neural Network*, *Naive Bayes*, dan *Decision Tree*. Data yang telah diproses oleh ketiga algoritma tersebut, akan Data pada *select columns* juga akan langsung diteruskan ke bagian test and score atau bagian penyelesaian dan hasil.

Pada gambar 7, diperlihatkan hasil dari data yang telah di proses oleh metode, diperlihatkan bahwa algoritma *Naive Bayes*, memiliki tingkat keakuratan 0,966%, presisi 0,967%, dan recall 0,966%. Algoritma *Neural Network* memiliki tingkat keakuratan 0,963%, presisi 0,963%, dan recall 0,963%. Sedangkan algoritma *Decision Tree* memiliki tingkat keakuratan klasifikasi 0,942%, presisi 0,942%, recall 0,942%. Pada gambar 10 diatas, Algoritma *Naive Bayes* memiliki tingkat keakuratan klasifikasi (CA), Presisi(perc), dan Recall yang paling tinggi, dibandingkan dengan algoritma *Neural Network* dan algoritma *Decision Tree*



Gambar 7. Hasil proses dari ketiga algoritma

2) Hasil Perbandingan Algoritma Neural Network, Naive Bayes, dan Decision Tree

Pada penelitian ini mendapatkan hasil dengan 4 tabel perbandingan antara algoritma *neural network*, *naive bayes*, dan *decision tree*. Ke 4 tabel dari hasil perbandingan ini menunjukkan hasil yang berbeda pada setiap tabelnya.

Tabel 3. Perbandingan Hasil Prediksi Algoritma

Model	AUC	CA	F1	Prec	Recall	MCC
naive bayes	0.989	0.964	0.964	0.965	0.964	0.922
Neural Network	0.993	0.963	0.963	0.963	0.963	0.918
Tree	0.932	0.941	0.941	0.941	0.941	0.870

Sumber : Hasil Penelitian 2023

Berdasarkan hasil perbandingan pada tabel pertama diatas, menunjukkan bahwa algoritma *Naive Bayes* memiliki nilai keakuratan klasifikasi (CA) yang lebih tinggi dibanding dengan algoritma yang lain. Algoritma *Naive Bayes* (NB), memiliki nilai keakuratan (CA) yang paling tinggi, sebanyak 96,4%. Pada algoritma *Neural Network* (NN), memiliki nilai keakuratan sebesar 96,3%, sementara pada algoritma *Decision Tree* (DT) hanya mencapai 94,1% pada tingkat keakuratan (CA).

Presisi (*Precision*) adalah ukuran kepastian, hal ini yang pada dasarnya adalah kekuatan prediksi algoritma klasifikasi. Pada tabel diatas menunjukkan presisi algoritma *Naive Bayes* (NB) mencapai 96,5% yang lebih tinggi dibanding algoritma *Neural Network* (NN) yang mencapai nilai presisi sebesar 96,3%, sedangkan algoritma *Decision Tree* (DT) hanya mencapai 94,1% untuk hasil dari nilai presisi.

Selanjutnya, recall. Pada tabel diatas menunjukkan hasil dari algoritma *Naive Bayes* (NB) memiliki nilai dari hasil recall yang lebih baik, sebesar 96,4% dibandingkan dengan algoritma *Naural Network* (NN) dengan hasil recall sebesar 96,3%. Sedangkan algoritma *Decision Tree* hanya mencapai 94,1% dari hasil recall.

Tabel 4. Perbandingan Hasil Prediksi Algoritma

Model	AUC	CA	F1	Prec	Recall	MCC
naive bayes	0.989	0.967	0.967	0.968	0.967	0.928
Neural Network	0.994	0.967	0.967	0.967	0.967	0.927
Tree	0.931	0.941	0.940	0.940	0.941	0.868

Sumber : Hasil Penelitian 2023

Hasil Perbandinga antara ketiga algoritma yang ada pada Tabel 4, menunjukan bahwa algoritma *Naive Bayes* memiliki nilai kekuatan klasifikasi (CA) sebesar 96,7%. Dan algoritma *Neural Network* (NN) memiliki hasil nilai keakuratan yang sama dengan algoritma *Naive Bayes*, yaitu sebesar 96,7%, sementara algoritma *Decision Tree* (DT) hanya mencapai 94,1% untuk nilai keakuratan (CA). Hal ini dapat membuktikan bahwa pada tabel kedua ini algoritma *Naive Bayes* (NB) dan algoritma *Neural Network* (NN) adalah algoritma klasifikasi yang terbaik dalam menggunakan dataset *breast cancer wisconsin*.

Pada hasil Presisi (*Precision*) Algoritma Naive Bayes mencapai nilai presisi yang lebih tinggi dibanding dengan algoritma yang lain, dengan mencapai nilai presisi sebesar 96,8% sedangkan untuk Algoritma Neural Network (NN) mencapai nilai presisi sebesar 96,7%. Sementara untuk algoritma Decision Tree (DT) hanya mencapai 94.0 untuk nilai presisi.

Recall, berdasarkan tabel 2 diatas didapatkan hasil dari perbandinga ketiga algoritma, yang dimana algoritma *Naive Bayes* (NB) mencapai nilai recall sebesar 96,7%, dan algoritma *Neural Network* (NN) mencapai nilai yang sama dengan algoritma Naive Bayes, yaitu sebesar 96,7%. Sedangkan algoritma *Decision Tree* (DT) mencapai nilai recall sebesar 94,1%.

Tabel 5. Hasil Perbandingan Prediksi Algoritma

Model	AUC	CA	F1	Prec	Recall	MCC
naive bayes	0.988	0.966	0.966	0.967	0.966	0.925
Neural Network	0.991	0.963	0.964	0.964	0.963	0.919
Tree	0.934	0.941	0.940	0.940	0.941	0.868

Sumber : Hasil Penelitian 2023

Berdasarkan perbandingan antara ketiga algoritma, seperti yang diperlihatkan pada tabel 3 diatas, algoritma *Naive Bayes* mencapai nilai yang paling tinggi diantara algoritma *Naural Network* dan algoritma *Decision Tree*. Hasil dari nilai keakuratan (CA) pada algoritma *Naive Bayes* (NB) adalah sebesar 96,6%. Sedangkan algoritma *Neural Network* (NN) hanya mencapai nilai keakuratan sebesar 96,3%, sementara pada algoritma *Decision Tree* (DC) mencapai 94,1% untuk nilai keakuratan (CA).

Presisi (*Precision*), pada tabel 3 diatas diperlihatkan bahwa algoritma *Naive Bayes* (NB) memiliki nilai presisi yang lebih tinggi, yaitu sebesar 96,7% sedangkan untuk algoritma *Neural Network*, mencapai 96,4% untuk nilai presisi. Sementara algoritma *Decision Tree* hanya mencapai 94,0% nilai presisi.

Selanjutnya Recall, algoritma *Naive Bayes* mencapai nilai recall yang lebih tinggi dibandingkan dengan algoritma *Neural Network* dan algoritma *Decision Tree*. Nilai recall pada algoritma *Naive Bayes* (NB) sebesar 96.6% sedangkan untuk algoritma *Neural Nerwork* (NN) mencapai 96,3% . Sementara algoritma *Decison Tree* (DT) hanya mencapai 94,1% untuk nilai recall.

Tabel 6. Hasil Perbandingan Algoritma

Model	AUC	CA	F1	Prec	Recall	MCC
naive bayes	0.989	0.966	0.966	0.967	0.966	0.926
Neural Network	0.991	0.963	0.963	0.963	0.963	0.918
Tree	0.932	0.942	0.942	0.942	0.942	0.871

Sumber : Hasil Penelitian 2023

Berdasarkan tabel 6 diatas, Hasil perbandingan dari ketiga algoritma dapat dibuktikan bahwa *Naive Bayes* (NB) memiliki Nilai keakuratan (CA) yang lebih tinggi dengan mencapai 96,6%. Sedangkan untuk algoritma *Neural Network* (NN) mencapai tingkat keakuratan (CA) sebesar 96,3%. Sementara algoritma *Decision Tree* (DT) hanya mencapai 94,2% untuk nilai keakuratan (CA).

Pada hasil presisi (*Precision*) algoritma *Naive Bayes* (NB) mencapai 96,7% dan menjadi algoritma dengan hasil presisi yang lebih tinggi. Sedangkan algoritma *Neural Network*

(NN) mencapai 96,3%, sedangkan hasil presisi pada algoritma *Decision Tree* (DT) hanya mencapai 94,2%.

Recall. Hasil perbandingan dari tabel 4 diatas, diperlihatkan bahwa hasil recall pada algoritma *Naive Bayes* (NB) menjadi hasil yang lebih tinggi dari antara algoritma yang lain, dengan mencapai nilai recall sebesar 96,6%. Untuk algoritma *Neural Network* (NN) mencapai nilai recall sebesar 96,3%, sedangkan algoritma *Decision Tree* hanya mencapai 94,2% untuk nilai recall.

Pada penelitian sebelumnya yang berjudul “Analisis Algoritma Klasifikasi Neural Network Untuk Diagnosis Penyakit Kanker Payudara”, mendapat sebuah hasil accuracy = 71,83%, precision =81,08% dan recall =69,17% dengan AUC sebesar =0,806.[7]. Berdasarkan Tinjauan pustaka di atas juga menjelaskan, bahwa pada penelitian terdahulu yang berjudul “Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma *Machine Learning*” menjelaskan tentang dampak penyakit kanker payudara salah satu penyakit non-kulit yang paling berbahaya pada wanita, disebabkan oleh berbagai faktor, mulai dari sel dan saluran hingga jaringan pendukung payudara, selain kulit payudara, juga menjelaskan bahwa Dalam kanker payudara dapat dibedakan menjadi ganas dan jinak. Jika diketahui jenisnya, maka pencegahan dan pengobatan akan segera dilakukan sesuai dengan jenis kanker payudaranya, sehingga dapat menghindari efek samping pada pasien dan bahkan kematian. Metode yang digunakan dalam penelitian ini adalah algoritma *mechile learning* dengan pengolahan data dilakukan dengan 3 tahap yaitu data cleaning, normaliasi data, dan pembagian data sehingga mendapatkan hasil algoritma *logistic regresion, decision tree, naive bayes* dan *k-nearest neighbor* memiliki nilai akurasi yang sama tinggi yaitu sebesar 95,00%. Penerapan data mining dengan menggunakan metode Algoritma *logistic regresion, decision tree, naive bayes* dan *k-nearest neighbor* dapat mempercepat pengambilan keputusan dalam memprediksi dalam klasifikasi penentuan jenis kanker payudara [9].

5. Simpulan

Klasifikasi merupakan salah satu teknik penting dalam data mining yang dapat diaplikasikan di berbagai bidang. Penelitian ini menyajikan perbandingan antara algoritma *Neural Network, Naive Bayes, dan Decision Tree* dalam mengklasifikasi penyakit kanker payudara.

Berdasarkan penelitian yang telah dilakukan, maka diperoleh kesimpulan bahwa Algoritma *Naive Bayes* sangat baik digunakan dalam melakukan klasifikasi dibandingkan dengan algoritma *Neural Network* dan *Decision Tree*. Algoritma *Naive Bayes* memiliki tingkat akurasi yang lebih tinggi dibanding algoritma *Neural Network* dan *Decision Tree*, begitupun dengan tingkat nilai presisi dan nilai recall. Hal ini membuktikan bahwa algoritma *Naive Bayes* (NN) adalah algoritma klasifikasi terbaik dalam menggunakan dataset *breast cancer wisconsin*.

Daftar Referensi

- [1] B. Aisyah and Y. Sulisty, “Klasifikasi Kanker Payudara Menggunakan Algoritma Gain Ratio,” *J. Tek. Elektro*, vol. 8, no. 2, pp. 43–46, 2016.
- [2] J. T. Wijaya *et al.*, “Kata Kunci : Klasifikasi data, Kanker Payudara, K-Nearest Neighbor, Gaussian Naive Bayes,” vol. 1, no. 1, pp. 233–237, 2021.
- [3] 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.
- [4] I. L. Maria, A. A. Sainal, and M. Nyorong, “Risiko Gaya Hidup Terhadap Kejadian Kanker Payudara Pada Wanita - Lifestyle Risk Factors of Women with Breast Cancer,” *Media Kesehat. Masy. Indones.*, vol. 13, no. 2, pp. 157–166, 2017, [Online]. Available: <https://journal.unhas.ac.id/index.php/mkmi/article/view/1988>.
- [5] A. Jamal, A. Handayani, A. A. Septiandri, E. Ripmiatin, and Y. Effendi, “Dimensionality Reduction using PCA and K-Means Clustering for Breast Cancer Prediction,” *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 9, no. 3, pp. 192-201, 2018, doi: 10.24843/lkjit.2018.v09.i03.p08.
- [6] T. Taufiq, & Y. Yudihartanti, "Penerapan Theorema Bayes Pada Penilaian Kelayakan Angkutan Kota". *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 10, no. 1, pp. 111-122, 2021.
- [7] fitra, “Analisis Algoritma Klasifikasi Neural Network Untuk Diagnosis Penyakit Kanker

- Payudara,” *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 149–156, 2019, doi: 10.33480/pilar.v15i2.601.
- [8] H. Oktavianto and R. P. Handri, “Analisis Klasifikasi Kanker Payudara Menggunakan Algoritma Naive Bayes,” *INFORMAL Informatics J.*, vol. 4, no. 3, pp. 117-126, 2020, doi: 10.19184/isj.v4i3.14170.
- [9] N. R. Muntari and K. H. Hanif, “Klasifikasi Penyakit Kanker Payudara Menggunakan Perbandingan Algoritma Machine Learning,” *J. Ilmu Komput. dan Teknol.*, vol. 3, no. 1, pp. 1–6, 2022, doi: 10.35960/ikomti.v3i1.766.
- [10] A. M. Zamani and B. Amaliah, “Implementasi Algoritma Genetika pada Struktur Backpropagation Neural Network untuk Klasifikasi Kanker Payudara,” vol. 1, 2012.
- [11] Fahminlb33, “breast-cancer-wisconsin.csv,” 2021. <https://github.com/Kodesiana/YT-Orange-Data-Mining-Series/blob/master/datasets/breast-cancer-wisconsin.csv>.
- [12] N. Bayes, M. A. Jabbar, E. Hasmin, C. Susanto, and W. Musu, “Komparasi Algoritma Decision Tree , Naive Bayes , dan K-Nearest Neighbors dalam Klasifikasi Kanker Payudara,” vol. 14, no. 3, pp. 258–270, 2022.
- [13] B. Rifai, “Algoritma Neural Network Untuk Prediksi,” *Techno Nusa Mandiri*, vol. IX, no. 1, pp. 1–9, 2013.
- [14] C. Cristina and A. Kurniawan, “Sejarah, Penerapan, dan Analisis Resiko dari Neural Network: Sebuah Tinjauan Pustaka,” *J. Inform. J. Pengemb. IT*, vol. 3, no. 2, pp. 259–270, 2018, doi: 10.30591/jpit.v3i2.890.
- [15] A. M. Khalimi, “Bentuk Algoritma Neural Network,” *pengalaman-edukasi.com*, 2020. <https://www.pengalaman-edukasi.com/2020/06/pembahasan-dasar-algoritma-artificial.html>.
- [16] H. Pramudia and A. Nugroho, “Sistem Informasi Kerusakan Laptop Menggunakan Metode Naive Bayes,” *Teknol. Elektro, Univ. Mercu Buana*, vol. 8, no. 3, pp. 206–214, 2017, [Online]. Available: <https://publikasi.mercubuana.ac.id/files/journals/4/articles/2186/submission/original/2186-4652-1-SM.pdf>.
- [17] 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.
- [18] P. Das, “Bentuk Rumus Algoritma Naive Bayes.” <https://www.codespeedy.com/naive-bayes-algorithm-in-python/>.
- [19] F. Dwi Meliani Achmad, Budanis, Slamet, “Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree,” *J. IPTEK*, vol. 16, no. 1, pp. 18–23, 2012, [Online]. Available: <http://jurnal.itats.ac.id/wp-content/uploads/2013/06/3.-BUDANIS-FINAL-hal-17-23.pdf>.
- [20] R. Hartono, Y. Sumaryana, and A. Nurfaizi, “Analisa Perbandingan Kinerja Algoritma Klasifikasi Untuk Prediksi Penyakit Kanker Payudara,” vol. 7, no. 1, pp. 116–124, 2023.
- [21] B. Rogojan, “Gambar Bentuk Decision Tree,” *SMB Lite*. <https://medium.com/smb-lite/what-is-a-decision-tree-algorithm-4531749d2a17>.
- [22] “Gambar rumus akurasi,presisi,recall,” *adpokat.github*. <https://adpokat.github.io/informasi/post/rumus-perhitungan-akurasi-presisi-recall-f-ukur/>.