**IMPLEMENTASI ALGORITMA KNN PADA** **KLASIFIKASI PENYAKIT KANKER PAYUDARA**

**PROPOSAL PROYEK AKHIR**

**Disusun Oleh:**

**Kelompok 7**

**Anisya Veneira Monica Rengkuan (21013022)**

**Gwent Nandasyah Labada (21013038)**

**Julinda Rondonuwu (21013043)**

**Uria Dutu (21013039)**



**DATA MINING DAN WAREHOUSE**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNK**

**UNIVERSITAS KATOLIK DE LA SALLE MANADO**

**2023**

DAFTAR ISI

[DAFTAR GAMBAR 2](#_Toc148563527)

[DAFTAR TABEL 3](#_Toc148563528)

[BAB I  PENDAHULUAN 4](#_Toc148563529)

[1.1 Latar Belakang 4](#_Toc148563530)

[1.2 Rumusan Masalah 5](#_Toc148563531)

[1.3 Tujuan 5](#_Toc148563532)

[1.4 Manfaat 5](#_Toc148563533)

[1.4.1 Teoritis 5](#_Toc148563534)

[1.4.2 Praktis 5](#_Toc148563535)

[1.5 Batasan Masalah 5](#_Toc148563536)

[BAB II  TINJAUAN PUSTAKA 6](#_Toc148563537)

[2.1 Algoritma K-Nearest Neighbor (KNN) 6](#_Toc148563538)

[2.3 Klasifikasi 6](#_Toc148563539)

[Klasifikasi adalah proses dari suatu objek yang memiliki ciri khas yang serupa dan dimasukkan ke dalam beberapa bagian kelas. Tujuan dari klasifikasi, yakni untuk mengidentifikasi setiap objek yang ada dengan kesamaan atribut – atribut sebagai dasar. Hal ini dilakukan agar setiap objek dimasukkan ke dalam kelompok yang sesuai. 6](#_Toc148563540)

[2.4 Penyakit Kanker Payudara 6](#_Toc148563541)

[BAB III METODOLOGI PENELITIAN 7](#_Toc148563542)

[3.1 Data 7](#_Toc148563543)

[3.2 Metode Usulan 7](#_Toc148563544)

[3.3 Metode Pembanding 7](#_Toc148563545)

[3.4 Evaluasi Performa 7](#_Toc148563546)

[DAFTAR PUSTAKA 10](#_Toc148563547)

# DAFTAR GAMBAR

Gambar 1. Precentage Split KNN 8

Gambar 2. Precentage Split Naïve Bayes 8

# DAFTAR TABEL

Tabel 1 Dataset 7

# BAB I  PENDAHULUAN

## 1.1 Latar Belakang

Penyakit kanker merupakan salah satu penyakit yang mematikan di dunia. Jumlah kematian yang disebabkan karena penyakit kanker,kanker paru,kanker hati,perut, dan kanker payudara menjadi penyebab terbesar kematian akibat kanker setiap tahunnya (Pusdatin 2015).Penyakit kanker payudara umumnya terjadi pada wanita. WHO memperkirakan 84 juta jiwa meninggal akibat kanker dalam rentang waktu 10 tahun dari 2005-2015 (Nugraha, et al., 2019). Dalam survei yang dilakukan mengemukakan bahwa 8 – 9 persen wanita mengalami kanker payudara. Jenis kanker yang paling banyak ditemukan sesudah kanker leher rahim. Pada tahun 2020 jumlah penderita kanker payudara mencapai 68.858, 16% dari total 396.914 kasus baru yang ditemui di Indonesia. Untuk jumlah kematian mencapai lebih dari 22 ribu kasus. (Rokom, 2022)

Data mining merupakan proses pengumpulan dan pengolahan data untuk mendapatkan informasi penting dari data. Data mining dapat bermanfaat untuk mengklasifikasikan, memperkirakan, memprediksi guna mendapatkan informasi yang berguna bagi semua orang. Proses pengumpulan data dan penggalian informasi dapat dilakukan dengan perangkat lunak yang menggunakan teknologi komputasi statistik, matematika, atau kecerdasan buatan (AI). Data mining sering disebut juga *Knowledge Discovery in Database (KDD).* (Setiawan, 2021)

Algoritma *K-Nearest Neighbour* merupakan metode klasifikasi objek baru dengan tetangga terdekat (K). KNN merupakan algoritma *supervised learning* dimana hasil query instance baru akan diklasifikasikan mayoritas kategori pada KNN dan kelas yang paling banyak muncul akan menjadi kelas hasil klasifikasi (Yunitasari, et al., 2021). Dalam penelitian ini akan digunakan teknik klasifikasi data mining *K-Nearest Neighbour* untuk mengklasifikasikan penyakit kanker payudara agar dapat menekan tingginya kasus penyakit payudara dan kematian akibat kanker khususnya kanker payudara.

## 1.2 Rumusan Masalah

1. Bagaimana penerapan algoritma *KNN* pada Klasifikasi Penyakit Kanker Payudara?
2. Bagaimana pengaruh algoritma *KNN* pada Klasifikasi Penyakit Kanker Payudara ?

## 1.3 Tujuan

1. Mengidentifikasi penerapan *KNN* pada Klasifikasi Penyakit Kanker Payudara.
2. Mengidentifikasi pengaruh algortima *KNN* pada hasil Klasifikasi Penyakit Kanker Payudara.

## 1.4 Manfaat

1. Mengetahui bagaimana penerapan KNN pada Klasifikasi Penyakit Kanker Payudara
2. Mengetahui bagaimana pengaruh algoritma KNN pada Klasifikasi Penyakit Kanker Payudara

### 1.4.1 Teoritis

1. Mengidentifikasi cara penerapan metode *KNN* dalam klasifikasi tingkat kepuasan *airline passenger.*
2. Menyatakan metode *KNN* dapat diterapkan dalam pengklasifikasian penyakit kanker payudara.

### 1.4.2 Praktis

1.4.2.1 Bagi Pengguna

1. Memberikan kemudahan pada Lembaga Kesehatan dalam melakukan klasifikasi penyakit kanker payudara.
2. Menemukan kekurangan dari pengklasifikasian yang sedang diterapkan untuk pengembangan pelayana Lembaga Kesehatan.

1.4.2.2 Bagi Mahasiswa

1. Mendapatkan pengalaman dalam membuat aplikasi yang dapat memudahkan dalam melakukan Implementasi Algoritma KNN Pada Klasifikasi Penyakit Kanker Payudara.
2. Mendapatkan pengetahuan dalam Implementasi Algoritma KNN Pada Klasifikasi Penyakit Kanker Payudara

## 1.5 Batasan Masalah

1. Data yang dipakai dalam penelitian ini yaitu berdasarkan data dari UCI.edu
2. Inputan hanya berupa file arff.
3. Hanya terdiri dari 2 *class* pada dataset

# BAB II  TINJAUAN PUSTAKA

## 2.1 Algoritma K-Nearest Neighbor (KNN)

Algoritma *K-Nearest Neighbor (k-NN)* adalah salah satu metode yang terdapat dalam ilmu komputer yang membahas tentang pembelajaran mesin yang diterapkan dalam pembuatan klasifikasi dan regresi data. Prinsip yang dimiliki oleh algoritma ini adalah pengelompokan data yang menggunakan kesamaan dengan tetangga terdekat dalam ruang fitur sebagai dasar prinsip penerapan metode ini. Dalam pembuatan klasifikasi data, algoritma ini membuat sekumpulan data uji yang telah dibagi untuk setiap kategori berdasarkan mayoritas kelas dari k yang dimiliki oleh tetangga terdekatnya. (Ismail, 2019)

## 2.3 Klasifikasi

## Klasifikasi adalah proses dari suatu objek yang memiliki ciri khas yang serupa dan dimasukkan ke dalam beberapa bagian kelas. Tujuan dari klasifikasi, yakni untuk mengidentifikasi setiap objek yang ada dengan kesamaan atribut – atribut sebagai dasar. Hal ini dilakukan agar setiap objek dimasukkan ke dalam kelompok yang sesuai. (Nurcahyo, 2018)

## 2.4 Penyakit Kanker Payudara

Penyakit kanker payudara adalah jenis kanker yang terletak dalam jaringan payudara. Kanker payudara bisa dialami oleh pria maupun Wanita. Ada beberapa hal yang menjadi faktor pemicu terjadinya kanker payudara, yakni usia, hormone, gaya hidup, riwaya keluarga, usia, dan genetika. (Atthalla, et al., 2018)

# BAB III METODOLOGI PENELITIAN

## 3.1 Data

Data yang ada pada tabel dibawah ini merupakan data yang diambil dari *website* Kaggle.com.

**Tabel 1. *Dataset***

|  |  |
| --- | --- |
| **Nama *Dataset*** | *Breast Cancer* |
| **Jumlah *Record*** | 286 |
| **Pembagian *Class*** | *No reccurence events and reccurence events* |
| **Jumlah Atribut** | 10 |
| **Atribut** | *Age, menopause, tumor size, inv-nodes, node-caps, deg-malig, breast, breast-quad, irradiat, class* |
| **Sumber *Dataset*** | *UCI.edu* |

## 3.2 Metode Usulan

Metode usulan yang digunakan untuk melakukan klasifikasi tingkat kepuasan pelanggan pesawat udara adalah metode K-*Nearest Neighbour (KNN)*. Metode ini relatif lebih sederhana, fleksibel, tidak memerlukan asumsi tertentu tentang data, dapat menangani data subjektif, dan mudah diinterpretasikan. Anda juga dapat menyesuaikan tingkat sensitivitas dengan mengatur parameter K.

## 3.3 Metode Pembanding

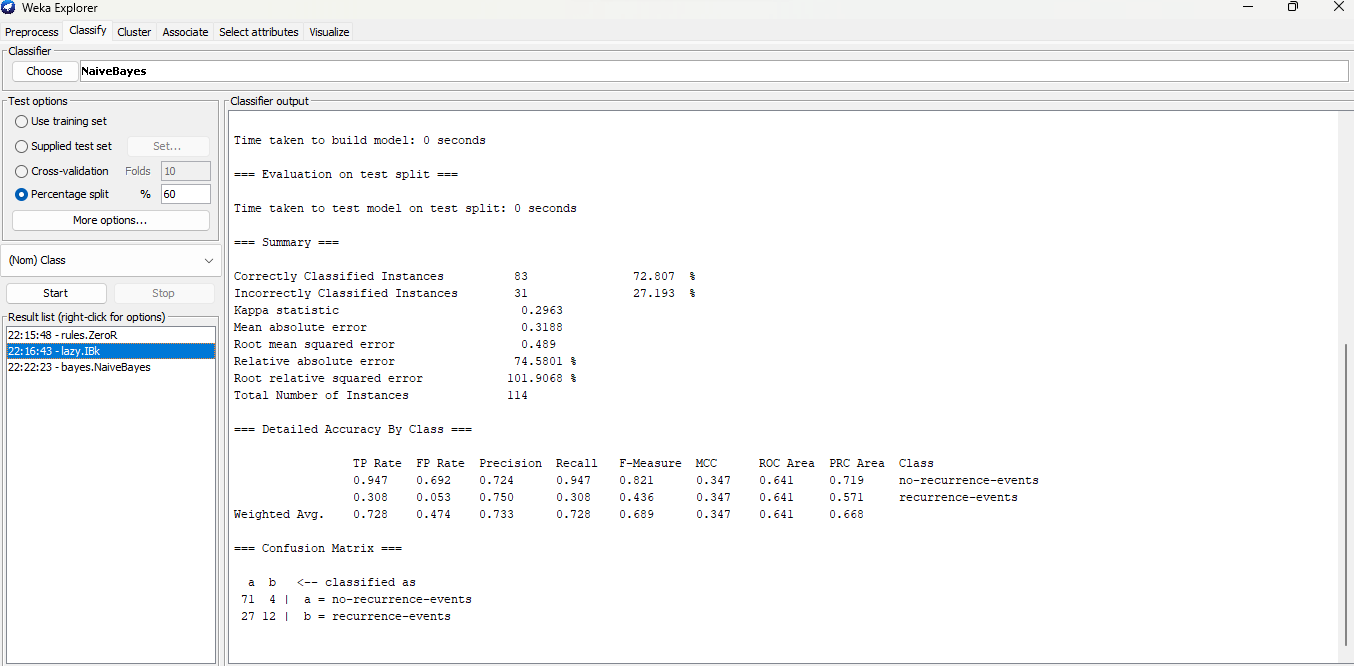
Metode pembanding yang akan digunakan pada penelitian ini adalah metode Naïve Bayes. Naïve Bayes merupakan salah satu metode yang dapat mengklasifikasikan data yang dilakukan secara statistik untuk memperkirakan probabilitas dari suatu *class.*

## 3.4 Evaluasi Performa

Untuk melakukan evaluasi performa dilakukan pengujian data menggunakan aplikasi WEKA dengan data *breast cancer* presentase data *training* sebesar 60% dan data testing sebesar 40%, selain itu digunakan 2 kriteria pembanding yakni akurasi dan waktu. Akurasi berarti mengukur sejauh mana metode KNN dapat memprediksi dengan benar. Waktu merujuk pada berapa lama metode KNN membutuhkan waktu dalam memprediksikan data. Hasil *precentage split* dari data *breast cancer* sebagai berikut:

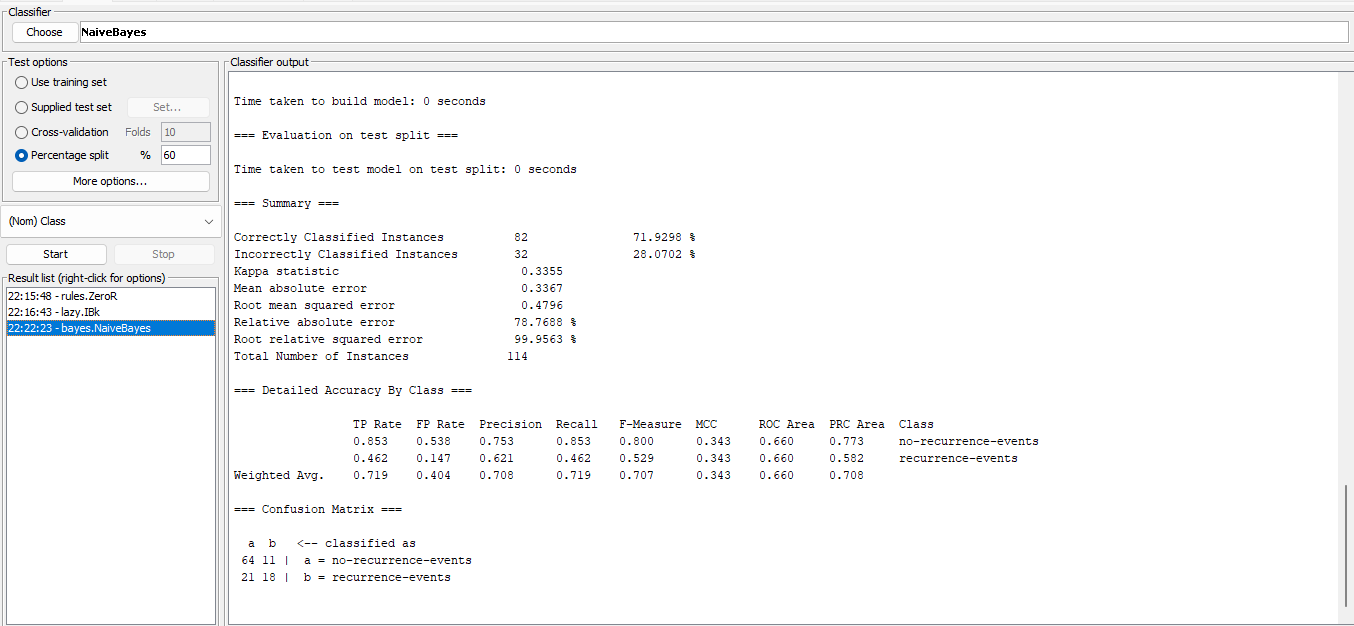
**Tabel 2. Hasil *Precentage Split***

|  |  |  |
| --- | --- | --- |
| **Kriteria Pembanding** | **KNN** | **Naïve Bayes** |
| Akurasi | 72.807 % | 71.9298% |
| Waktu | 0 *seconds* | 0 *seconds* |



**Gambar 1 Precentage Split KNN**

Gambar diatas diperoleh hasil *precentage split* kriteria pembanding akurasi pada metode KNN yang didapat sebesar 72.807% dan waktu yang didapat 0 *seconds.*



**Gambar 2 Precentage Split Naïve Bayes**

Gambar diatas diperoleh hasil percentage split Naïve Bayes sebesar 71.9289% dan waktu yang ditempuh 0 *seconds.*

**3.5 *Experimental Setup***

Penelitian ini menggunakan data trainig sebesar 60% dan data testing yang digunakan sebanyak 40%. Dari data-data tersebut akan diuji tingkat akurasi dan waktu yang didapat dari metode KNN dan Naïve Bayes. Evaluasi kinerja dari kedua metode tersebut diuji menggunakan WEKA *precentage split.*

3.5.1 Perancangan Aplikasi

Untuk pemrograman klasifikasi penyakit kanker payudara akan digunakan Bahasa pemrograman phyton.

# DAFTAR PUSTAKA

Atthalla, I. N., Jovandy, A. & Habibie, H., 2018. Klasifikasi Penyakit Kanker Payudara Menggunakan Metode K Nearest Neighbor. *Computer Science and ICT,* 4(1), pp. 148-151.

Ismail, A. M., 2019. Cara Kerja Algoritma k-Nearest Neighbor (k-NN).

Nugraha, F. S., Shidiq, J. M. & Rahayu, S., 2019. Analisis Algoritma Klasifikasi Neural Network Untuk Diagnosis Penyakit Kanker Payudara. *Pilar Nusa Mandiri,* 15(2), pp. 149-156.

Nurcahyo, A. R., 2018. Pengertian, Perubahan, dan Klasifikasi Materi.

R., 2022. *Kanker Payudara Paling Banyak Di Indonesia, Kemenkes Targetkan Pemerataan Layanan Kesehatan.* [Online]   
Available at: https://www.dicoding.com/blog/apa-itu-data-mining/  
[Accessed 17 Oktober 2023].

Setiawan, R., 2021. *Apa Itu Data Mining dan Bagaimana Metodenya.* [Online]   
Available at: https://www.dicoding.com/blog/apa-itu-data-mining/  
[Accessed 17 Oktober 2023].

Yunitasari, Hopipah, H. S. & Mayasari, R., 2021. Optimasi Backward Elimination untuk Klasifikasi Kepuasan Pelanggan Menggunakan Algoritme k-Nearest Neighbour (k-NN) dan Naive Bayes. *Technomedia Journal (TMJ),* 6(1), pp. 99-110.