

**KLASIFIKASI PENYAKIT KANKER PAYUDARA
MENGUNAKAN ALGORITMA *DECISION TREE C4.5***

TUGAS AKHIR

Disusun Oleh:

Fransiscus Valentino Sanders Piay

13013066



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS KATOLIK DE LA SALLE

MANADO

2019

**KLASIFIKASI PENYAKIT KANKER PAYUDARA
MENGUNAKAN ALGORITMA *DECISION TREE C4.5***

TUGAS AKHIR

**Diajukan untuk memenuhi sebagian persyaratan untuk mendapatkan Gelar
Sarjana Teknik pada Program Studi Teknik Informatika**

Disusun Oleh:

Fransiscus Valentino Sanders Piay

13013066



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS KATOLIK DE LA SALLE

MANADO

2019



**UNIVERSITAS KATOLIK DE LA SALLE
MANADO INDONESIA**

Nama : Fransiscus Valentino Sanders Piay
Nim : 13013066
Fakultas : Teknik
Program studi : Teknik Informatika
Judul Tugas Akhir : Klasifikasi Penyakit Kanker Payudara
Menggunakan Algoritma C4.5
Dosen Pembimbing 1 : Angelia Melani Adrian, Ph.D.
Dosen Pembimbing 2 : Junaidy B. Sanger, S.Kom., M.Kom

Menyetujui,

Manado, 1 Agustus 2019

Dosen Pembimbing I

Angelia Melani Adrian, Ph.D.

Dosen Pembimbing II

Junaidy B. Sanger, S.Kom., M.Kom.

Mengetahui,

Ketua Program Studi

Vivie D. Kumenap, S.T., M.Cs.

Dekan Fakultas Teknik

Ronald A. Rachmadi, S.T., M.T.

LEMBAR PERNYATAAN

Yang bertandatangan di bawah ini:

Nama : Fransiscus Valentino Sanders Piay
NIM : 13013066
Tempat, Tanggal Lahir : Manado, 22 Februari 1996
Fakultas/Jurusan : Teknik Informatika

Menyatakan bahwa Laporan Tugas Akhir dan/atau Aplikasi atau Program berjudul "Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma *Decision Tree C4.5*" yang saya buat adalah benar hasil karya saya dan bukan karya tulis orang lain, baik sebagian atau seluruhnya kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya dan apabila pernyataan ini tidak benar maka saya bersedia menerima sanksi akademis sesuai dengan yang ditetapkan oleh Fakultas, berupa pembatalan Tugas Akhir dan hasilnya.

Manado, 1 Agustus 2019

Yang Menyatakan,



Fransiscus Valentino Sanders Piay

Menyetujui,

Dosen Pembimbing I

Angelia Melani Adrian, Ph.D.

Dosen Pembimbing II

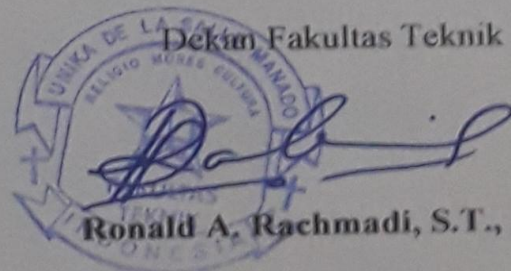
Junaidy B. Sanger, S.Kom., M.Kom.

Mengetahui,

Ketua Program Studi

Vivie D. Kumenap, S.T., M.Cs.

Dekan Fakultas Teknik



Ronald A. Rachmadi, S.T., M.T.

ABSTRACT

Breast Cancer is a group of diseases that cause cells in the body to change and grow out of control. Most types of cancer cells sometimes form lumps or masses called tumors and are named according to the part of the body where the tumor originated. Breast cancer is a malignant tumor that originates in the breast cells. This disease usually occurs in women, but men can also get it. Based on GLOBOCAN, International Agency of Research on Cancer (IARC) data in 2012 there were more than 14,067,894 new cases of cancer and 8,201,575 death caused by cancer worldwide. In the case of breast cancer, this disease is the second most common cancer in the world. Breast cancer is estimated to have 1.67 million cases diagnosed in 2012. Breast cancer, prevalent in less developed countries compared to more developed countries. It is ranked as the fifth deadly disease in the world.

In this study a data mining application will be build to classify Breast Cancer disease using the C45 Algorithm. The data used in this study were breast-cancer and breast-w-original dataset taken from UCI (University of California Irvine). Furthermore, we carried out a performance comparisons in terms of accuracy and kappa statistic for KNN, Naïve Bayes, and SVM algorithm.

Keywords: Breast Cancer, Classification, C45, Data Mining.

ABSTRAK

Penyakit Kanker Payudara adalah kelompok penyakit yang menyebabkan sel-sel di dalam tubuh berubah dan tumbuh tak terkendali. Sebagian besar jenis sel kanker kadang membentuk benjolan atau massa yang disebut tumor, dan dinamai sesuai dengan bagian tubuh dimana tumor berasal. Kanker payudara adalah tumor ganas yang berawal dari dalam sel-sel payudara. Penyakit ini terjadi hampir seluruhnya pada wanita, tetapi pria juga bisa mendapatkannya. Berdasarkan Data *GLOBOCAN, International Agency of Research on Cancer (IARC)* pada tahun 2012 terdapat lebih dari 14.067.894 kasus kanker baru dan 8.201.575 kematian akibat kanker diseluruh dunia. Dalam kasus kanker payudara, penyakit ini merupakan kanker paling umum kedua di dunia. Kanker payudara diperkirakan memiliki 1,67 juta kasus yang didiagnosis pada tahun 2012. Pada kasus kanker payudara, penyakit ini lebih banyak terjadi pada daerah yang kurang berkembang dibandingkan dengan daerah yang lebih maju. Penyakit kanker payudara menjadi urutan kelima dari penyakit yang mematikan untuk seluruh dunia.

Dalam Tugas Akhir ini akan dibangun aplikasi data mining untuk klasifikasi penyakit kanker payudara menggunakan algoritma C45. Data yang digunakan adalah *dataset breast-cancer* dan *breast-w-original* yang diambil dari UCI (*University Of California Irvine*). Dalam penelitian ini juga akan dilakukan perbandingan performa nilai akurasi dan kappa statistik antara algoritma *KNN*, *Naïve Bayes*, dan *SVM*.

Kata kunci: Kanker Payudara, Klasifikasi, C45, Penambangan Data.

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas tuntunan dan penyertaan-Nya sehingga penyusunan Laporan Tugas Akhir ini dapat berjalan dengan semestinya. Penyusunan laporan Tugas Akhir ini bertujuan untuk memenuhi salah satu persyaratan akademik program studi Teknik Informatika, Universitas Katolik De La Salle Manado. Laporan Tugas Akhir ini membahas tentang Klasifikasi Penyakit Kanker Payudara menggunakan Algoritma C4.5. Dalam proses penyusunan laporan ini, penulis banyak mendapat bantuan baik secara langsung maupun tidak langsung, karena itu pada kesempatan ini penulis ingin mengucapkan terima kasih sebesar-besarnya kepada :

1. Bapak Prof. Dr. Johanis Ohoitumur selaku Rektor Universitas Katolik De La Salle Manado.
2. Bapak Ronald A. Rachmadi, S.T., M.T selaku Dekan Fakultas Teknik Universitas Katolik De La Salle Manado.
3. Ibu Vivie D. Kumenap, S.T., M.Cs. selaku Ketua Program Studi Teknik Informatika.
4. Ibu Angelia Melani Adrian, Ph.D. sebagai Dosen Pembimbing I Tugas Akhir yang selalu memberikan nasihat dan saran dalam pengerjaan aplikasi dan laporan.
5. Bapak Junaidy B. Sanger, S.Kom., M.Kom. sebagai Pembimbing II Tugas Akhir yang selalu memberi nasihat dan saran dalam pengerjaan aplikasi dan laporan.
6. Christi, Andre, Alvin, Cecil, David, Susan, dan Giberth selaku teman-teman yang banyak membantu dan memberi semangat dalam menyelesaikan Tugas Akhir.
7. Teman-teman angkatan 2013 Fakultas Teknik, Program Studi Teknik Informatika, Universitas Katolik De La Salle Manado.
8. Mama, Papa dan *support* dari keluarga yang telah mendoakan dan memberikan dukungan yang luar biasa kepada penulis.
9. Pihak-pihak lain yang tidak dapat disebutkan satu per satu

Dalam penyusunan Tugas Akhir penulis menyadari masih banyak kekurangan, maka dari itu diharapkan saran dan masukan dari para pembaca.

Manado, Agustus 2019

Penulis

DAFTAR ISI

COVER DEPAN	I
LEMBAR PENGESAHAN	ii
LEMBAR PERNYATAAN	iii
ABSTRACT	iv
ABSTRAK	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
BAB I PENDAHULUAN	Error! Bookmark not defined.
1.1. Latar Belakang	Error! Bookmark not defined.
1.2. Rumusan Masalah	Error! Bookmark not defined.
1.3. Tujuan Tugas Akhir	Error! Bookmark not defined.
1.4. Batasan Masalah	Error! Bookmark not defined.
1.5. Manfaat Tugas Akhir	Error! Bookmark not defined.
1.6. Metode Penelitian	Error! Bookmark not defined.
1.7. Sistematika Penulisan	Error! Bookmark not defined.
BAB II STUDI PUSTAKA	Error! Bookmark not defined.
2.1. Penyakit Kanker	Error! Bookmark not defined.
2.1.1. Penyakit Kanker Payudara (<i>Breast Cancer</i>)	Error! Bookmark not defined.
2.1.2. Faktor Resiko Kanker Payudara	Error! Bookmark not defined.
2.2. <i>Data Mining</i>	Error! Bookmark not defined.
2.3. Klasifikasi	Error! Bookmark not defined.
2.4. <i>Decision Tree</i>	Error! Bookmark not defined.
2.5. Algoritma C4.5	Error! Bookmark not defined.
2.5.1. Konsep Dasar Algoritma C4.5	Error! Bookmark not defined.
2.5.2. Contoh Perhitungan menggunakan Entropi dan <i>Gain Ratio</i>	Error! Bookmark not defined.
2.5.3. Kelebihan dan Kekurangan Algoritma C4.5	Error! Bookmark not defined.
2.5.4. Implementasi Algoritma C4.5	Error! Bookmark not defined.
2.6. <i>K-Folds Cross Validation</i>	Error! Bookmark not defined.
2.7. Analisis Performa	Error! Bookmark not defined.
2.7.1. Indikator pembandingan	Error! Bookmark not defined.
2.7.1.1. Akurasi	Error! Bookmark not defined.
2.7.1.2. Kappa Statistik	Error! Bookmark not defined.
2.8. Algoritma Pembandingan	Error! Bookmark not defined.
2.8.1. SVM (<i>Support Vector Machine</i>)	Error! Bookmark not defined.
2.8.2. KNN	Error! Bookmark not defined.
2.8.3. <i>Naïve Bayes</i>	Error! Bookmark not defined.
2.9. <i>Flowchart Diagram</i>	Error! Bookmark not defined.
2.10. Bahasa Pemrograman R	Error! Bookmark not defined.
2.10.1. <i>R Studio</i>	Error! Bookmark not defined.
BAB III ANALISIS	Error! Bookmark not defined.

3.1.	Data	Error! Bookmark not defined.
3.2.	Analisis Data	Error! Bookmark not defined.
3.3.	Analisis Pengolahan Data	Error! Bookmark not defined.
3.3.1.	Seleksi Data	Error! Bookmark not defined.
3.3.2.	Pembersihan Data	Error! Bookmark not defined.
3.3.3.	Transformasi Data	Error! Bookmark not defined.
3.3.4.	<i>Data Mining</i>	Error! Bookmark not defined.
BAB IV PERANCANGAN		Error! Bookmark not defined.
4.1.	Model Fungsional Perangkat Lunak	Error! Bookmark not defined.
4.1.1.	<i>Flowchart Diagram</i> Model	Error! Bookmark not defined.
4.1.2.	Deskripsi <i>Flowchart Diagram</i>	Error! Bookmark not defined.
4.2.	Membuat <i>Storyboard</i>	Error! Bookmark not defined.
4.2.1.	Tampilan <i>Storyboard</i>	Error! Bookmark not defined.
4.3.	Perangkat Lunak dan Keras	Error! Bookmark not defined.
BAB V IMPLEMENTASI		Error! Bookmark not defined.
5.1.	Pengodean	Error! Bookmark not defined.
5.1.1.	Pengodean menggunakan <i>R Studio</i>	Error! Bookmark not defined.
5.2.	Implementasi <i>Storyboard</i>	Error! Bookmark not defined.
BAB VI PENGUJIAN		Error! Bookmark not defined.
6.1.	Hasil dan Perbandingan Algoritma	Error! Bookmark not defined.
6.1.1.	Pengujian Indikator Perbandingan	Error! Bookmark not defined.
6.1.1.1.	Akurasi	Error! Bookmark not defined.
6.1.1.2.	Kappa Statistik	Error! Bookmark not defined.
6.1.1.3.	Pohon Keputusan yang dihasilkan (<i>Tree</i>)	Error! Bookmark not defined.
6.1.2.	Perbandingan Akurasi dan Kappa Statistik dengan Metode Perbandingan.	Error! Bookmark not defined.
6.2.	Pengujian Aplikasi	Error! Bookmark not defined.
6.2.1.	Pengujian menggunakan <i>Browser</i>	Error! Bookmark not defined.
6.3.	Pembahasan	Error! Bookmark not defined.
BAB VII KESIMPULAN DAN SARAN		Error! Bookmark not defined.
7.1.	Kesimpulan	Error! Bookmark not defined.
7.2.	Saran	Error! Bookmark not defined.
DAFTAR PUSTAKA		Error! Bookmark not defined.
LAMPIRAN		Error! Bookmark not defined.

DAFTAR GAMBAR

- Gambar 2.1. Tahapan *Knowledge Discovery in Database* (KDD) **Error! Bookmark not defined.**
- Gambar 4.1. Gambar *Flowchart Diagram* **Error! Bookmark not defined.**
- Gambar 4.2. *Storyboard* Tampilan Ambil *Dataset* **Error! Bookmark not defined.**
- Gambar 4.3. *Storyboard* Aplikasi untuk Klasifikasi Penyakit Kanker Payudara menggunakan Algoritma C4.5 bagian Klasifikasi C4.5 **Error! Bookmark not defined.**
- Gambar 4.4. *Storyboard* Aplikasi untuk Klasifikasi Penyakit Kanker Payudara menggunakan Algoritma C4.5 bagian Klasifikasi Uji Data Tunggal **Error! Bookmark not defined.**
- Gambar 5.1. *Library* dan Pengodean untuk memasukan *dataset*, klasifikasi, dan menampilkan akurasi **Error! Bookmark not defined.**
- Gambar 5.2. Pengodean pada UI (*User Interface*) menggunakan *R Shiny* **Error! Bookmark not defined.**
- Gambar 5.3. Pengodean pada *Server* menggunakan *R Shiny* **Error! Bookmark not defined.**
- Gambar 5.4. Tampilan Aplikasi untuk *Dataset Breast-Cancer* **Error! Bookmark not defined.**
- Gambar 5.5. Tampilan Aplikasi untuk *Dataset Breast-W-Original* **Error! Bookmark not defined.**
- Gambar 5.6. Tampilan atribut data dari *Dataset Breast-Cancer* **Error! Bookmark not defined.**
- Gambar 5.7. Tampilan atribut data dari *Dataset Breast-W-Original* **Error! Bookmark not defined.**
- Gambar 5.8. Tampilan *Summary* hasil Klasifikasi C4.5 dengan *Dataset Breast-Cancer* **Error! Bookmark not defined.**
- Gambar 5.9. Tampilan *Summary* hasil Klasifikasi C4.5 dengan *Dataset Breast-W-Original* **Error! Bookmark not defined.**
- Gambar 5.10. Tampilan *Summary* untuk Klasifikasi Uji Data Tunggal untuk *Dataset Breast-Cancer* **Error! Bookmark not defined.**
- Gambar 5.11. Tampilan *Summary* untuk Klasifikasi Uji Data Tunggal untuk *Dataset Breast-W-Original* **Error! Bookmark not defined.**
- Gambar 6.1. Pohon Keputusan pada *Dataset Breast_Cancer* **Error! Bookmark not defined.**
- Gambar 6.2. Pohon Keputusan pada *Dataset Breast-W-Original* **Error! Bookmark not defined.**
- Gambar 6.3. Tampilan Aplikasi yang dijalankan pada *browser Mozilla Firefox* **Error! Bookmark not defined.**
- Gambar 6.4. Tampilan Aplikasi yang dijalankan pada *browser Microsoft Edge* **Error! Bookmark not defined.**
- Gambar 6.5. Tampilan Aplikasi yang dijalankan pada *browser Google Chrome* **Error! Bookmark not defined.**
- Gambar 6.6. Tampilan Aplikasi yang dijalankan pada *browser Opera Mini* **Error! Bookmark not defined.**

DAFTAR TABEL

Tabel 2.1.	Tabel <i>Confusion Matrix</i> dengan 2 Kelas	Error! Bookmark not defined.
Tabel 2.2.	Penilaian Kelayakan untuk Melakukan Kredit Pembelian Mobil	Error! Bookmark not defined.
Tabel 2.3.	Tabel dari Beberapa Simbol-Symbol dalam <i>Flowchart Diagram</i>	Error! Bookmark not defined.
Tabel 3.1.	Tabel Atribut dan Tipe Data dari <i>Dataset Breast-Cancer</i>	Error! Bookmark not defined.
Tabel 3.2.	Tabel Atribut dan Tipe Data dari <i>Dataset Breast-W-Original</i>	24
Tabel 3.3.	Tabel Deskripsi <i>Dataset Breast-Cancer</i> dan <i>Breast_Wisconsin_Original</i>	24
Tabel 3.4.	Tabel Atribut <i>Age</i> pada <i>Dataset Breast-Cancer</i>	25
Tabel 3.5.	Tabel Atribut <i>Menopause</i> pada <i>Dataset Breast-Cancer</i>	25
Tabel 3.6.	Tabel Atribut <i>Tumor-Size</i> pada <i>Dataset Breast-Cancer</i>	25
Tabel 3.7.	Tabel Atribut <i>Inv-Size</i> pada <i>Dataset Breast-Cancer</i>	26
Tabel 3.8.	Tabel Atribut <i>Node-Caps</i> pada <i>Dataset Breast-Cancer</i>	27
Tabel 3.9.	Tabel Atribut <i>Deg-Malig</i> pada <i>Dataset Breast-Cancer</i>	27
Tabel 3.10.	Tabel Atribut <i>Breast</i> pada <i>Dataset Breast-Cancer</i>	27
Tabel 3.11.	Tabel Atribut <i>Breast_Quad</i> pada <i>Dataset Breast- Cancer</i>	27
Tabel 3.12.	Tabel Atribut <i>Irradit</i> pada <i>Dataset Breast- Cancer</i>	28
Tabel 3.13.	Tabel Atribut <i>Class</i> pada <i>Dataset Breast- Cancer</i>	28
Tabel 3.14.	Tabel Atribut <i>Clump_Thickness</i> pada <i>Dataset Breast-Wisconsin-Original</i>	28
Tabel 3.15.	Tabel Atribut <i>Cell_Size_Uniformity</i> pada <i>Dataset Breast-Wisconsin-Original</i>	29
Tabel 3.16.	Tabel Atribut <i>Cell_Shape_Uniformity</i> pada <i>Dataset Breast-Wisconsin-Original</i>	29
Tabel 3.17.	Tabel Atribut <i>Marginal_Adhesion</i> pada <i>Dataset Breast-Wisconsin-Original</i>	Error! Bookmark not defined.
Tabel 3.18.	Tabel Atribut <i>Single_Epi_Cell_Size</i> pada <i>Dataset Breast-Wisconsin-Original</i>	Error! Bookmark not defined.
Tabel 3.19.	Tabel Atribut <i>Bare_Nuclei</i> pada <i>Dataset Breast-Wisconsin-Original</i>	Error! Bookmark not defined.
Tabel 3.20.	Tabel Atribut <i>Bland_Chormatin</i> pada <i>Dataset Breast-Wisconsin-Original</i>	Error! Bookmark not defined.

Tabel 3.21.	Tabel Atribut <i>Normal_Nucleioli</i> pada <i>Dataset Breast-Wisconsin-Original</i>	Error! Bookmark not defined.
Tabel 3.22.	Tabel Atribut <i>Mitose</i> pada <i>Dataset Breast-Wisconsin-Original</i>	Error! Bookmark not defined.
Tabel 3.23.	Tabel Atribut <i>Class</i> pada <i>Dataset Breast-Wisconsin-Original</i>	Error! Bookmark not defined.
Tabel 3.24.	Tabel <i>Dataset Breast_Cancer</i> dengan nilai kosong.....	Error! Bookmark not defined.
Tabel 3.25.	Tabel <i>Dataset Breast-W-Original</i> dengan nilai kosong.....	Error! Bookmark not defined.
Tabel 4.1.	Tabel Deskripsi <i>Flowchart Diagram</i>	Error! Bookmark not defined.
Tabel 4.2.	Tabel Spesifikasi Perangkat Keras.....	40
Tabel 4.3.	Tabel Spesifikasi Perangkat Lunak.....	40
Tabel 6.1.	Tabel Akurasi <i>Dataset Breast-W-Original</i>	51
Tabel 6.2.	Tabel Akurasi <i>Dataset Breast-Cancer</i>	52
Tabel 6.3.	Tabel Akurasi Tertinggi <i>Dataset Breast-Cancer</i> dan <i>Breast-W-Original</i>	Error! Bookmark not defined.
Tabel 6.4.	Tabel Kappa Statistik <i>Breast-W-Original</i>	54
Tabel 6.5.	Tabel Kappa Statistik <i>Breast-Cancer</i>	55
Tabel 6.6.	Tabel Kappa Statistik Tertinggi <i>Dataset Breast-Cancer</i> dan <i>Breast-W-Original</i>	56
Tabel 6.7.	Tabel Akurasi dan Kappa Statistik Tertinggi untuk <i>Dataset Breast-Cancer</i> dan <i>Breast-W-Original</i>	56
Tabel 6.8.	Parameter <i>KNN</i> , <i>Naïve Bayes</i> , dan <i>SVM</i>	60
Tabel 6.9.	Tabel perbandingan Akurasi dan Kappa Statistik dari <i>Dataset Breast-W-Original</i> dengan parameter $K = 5$ untuk <i>KNN</i> dan $Epsilon = 0,0005$ dan $Gamma = 0.5$ untuk <i>SVM</i>	61
Tabel 6.10.	Perbandingan Akurasi dan Kappa Statistik dari <i>Dataset Breast-W-Original</i> dengan menggunakan parameter <i>standard</i> dari <i>Weka</i>	61
Tabel 6.11.	Perbandingan Akurasi dan Kappa Statistik dari <i>Dataset Breast-Cancer</i> dengan menggunakan parameter <i>standard</i> dari <i>Weka</i>	62

