

# KLASIFIKASI PENYAKIT KANKER PAYUDARA MENGGUNAKAN *DECISION TREE C4.5*

Fransiskus V. S. Piay<sup>1</sup>, Angelia M. Adrian<sup>1\*</sup>, Junaidy B. Sanger<sup>1</sup>

<sup>1</sup>Program Studi Teknik Informatika, Fakultas Teknik, Universitas Katolik De La Salle Manado  
e-mail: madrian@unikadelasalle.ac.id

**Abstrak** – 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. Tujuan Penelitian ini adalah membangun aplikasi data mining untuk klasifikasi penyakit kanker payudara menggunakan algoritma C4.5. Data yang digunakan adalah dataset breast-cancer dan breast-w-original yang diambil dari UCI (University Of California Irvine). Dalam penelitian ini juga dilakukan perbandingan performa hasil akurasi dan kappa statistik antara algoritma KNN, Naïve Bayes, dan SVM.

**Kata Kunci** – Kanker Payudara, Klasifikasi, Algoritma C4.5, Perbandingan Performa.

## I. PENDAHULUAN

Kanker 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 [1]. 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 di seluruh 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 [2].

*Data mining* adalah proses pencarian informasi dari data-data yang berada dalam *database* yang besar. Teknik *Data Mining* sudah diaplikasikan di berbagai bidang, salah satunya pada bidang medis. Dalam data *mining* banyak algoritma yang dapat digunakan diantaranya adalah algoritma *Decision Tree*. *Decision Tree* merupakan salah satu algoritma klasifikasi yang populer. Keuntungan menggunakan algoritma *Decision Tree C4.5* adalah dapat mengolah data numerik dan kategori,

menangani atribut yang hilang serta menghasilkan aturan-aturan yang mudah diinterpretasikan [3].

Penggunaan algoritma *Decision Tree C4.5* dalam diagnosis medis sudah terbukti keandalannya, diantaranya Diagnosis Tingkat Kesehatan Pasien menggunakan metode *Decision Tree* yang bertujuan untuk mendiagnosis kesehatan pasien terhadap bahaya merokok [4], Penerapan Algoritma C4.5 Untuk Klasifikasi Data Rekam Medis Berdasarkan *International Classification Diseases (IDC-10)* yang tujuannya untuk mengklasifikasikan diagnosis penyakit berdasarkan jenis kelamin, umur, dan tanggal pasien dirawat [5], Prediksi Penyakit *Tuberculosis (TBC)* Menggunakan Algoritma C4.5 yang tujuannya untuk memprediksikan apakah pengguna menderita penyakit TBC atau tidak [6].

Tujuan dari Penelitian ini adalah untuk mengklasifikasikan penyakit kanker payudara menggunakan algoritma *Decision Tree C4.5*. Data yang digunakan adalah *dataset breast\_cancer* dan *breast-w-original* dari UCI (*University of California Irvine*). Hasil yang diperoleh akan dibandingkan dengan metode SVM (*Support Vector Machine*), KNN (*K-Nearest Neighbor*), dan *Naïve Bayes* untuk mengukur performa dari metode yang digunakan.

## II. TINJAUAN PUSTAKA

Penyakit Kanker

Kanker adalah penyakit yang terjadi akibat pertumbuhan sel secara tidak terkendali yang memiliki kemampuan untuk menyusup dan merusak sel-sel sehat di dalam tubuh. Setiap sel tubuh memiliki pusat pengendali yang bernama *nukleus*. *Nukleus* terdiri dari rantai DNA (*Deoxyribo Nucleic Acid*) yang berfungsi sebagai pengatur sifat dan cara kerja masing-masing sel termasuk untuk membelah diri. Pada fase normal sel akan membelah diri sesuai dengan waktu yang telah ditentukan dan dengan sifat genetik pada sel yang normal. Namun pada saat terjadi mutasi pada sel maka informasi genetik sel dalam rantai DNA akan mengalami kerusakan, tercetak ganda atau bahkan hilang [7]. Pada proses pembelahan pada sel, akan terjadi mutasi di mana sel-sel akan membelah diri menjadi banyak sehingga akan membentuk sebuah jaringan tumor. Berdasarkan jenis sel yang sering diserang oleh kanker, dapat dibagi menjadi 5 kelompok yaitu [7]:

1. *Karsinoma* adalah kanker yang berasal dari jaringan kulit atau jaringan lapisan luar organ dalam.

2. *Sarkoma* adalah kanker yang berasal dari jaringan ikat seperti jaringan tulang, jaringan tulang rawan, jaringan lemak, jaringan otot, dan jaringan pembuluh darah.
3. *Leukemia* adalah kanker yang berasal dari jaringan pembentuk darah seperti sumsum tulang. Kanker jenis ini menyebabkan produksi sel darah secara berlebihan yang akan kemudian dilepaskan ke aliran darah.
4. *Limfoma* dan *Mieloma* adalah kanker yang berasal dari sel-sel pada sistem kekebalan tubuh.
5. Kanker di otak dan sumsum tulang belakang adalah kanker yang berasal dari jaringan sistem saraf pusat.

Kelompok kanker di atas adalah kanker yang sering terjadi atau dialami oleh manusia pada bagian dari sel-sel dalam tubuh. Dengan kata lain kanker akan menyerang bagian yang dianggap vital bagi manusia. Selain pengelompokan tersebut, kanker juga dibagi sesuai dengan organ tempat asal jaringan kanker berkembang seperti misalnya kanker pada paru-paru atau kanker payudara [7].

### Data Mining

*Data Mining* adalah proses untuk mencari informasi yang berguna secara otomatis pada repositori data yang besar. Teknik data mining sering digunakan untuk menjelajahi atau menelusuri suatu *database* yang besar untuk menemukan suatu pola yang baru dan berguna. Selain itu dalam proses *data mining* memiliki kemampuan untuk memprediksi hasil dari pengamatan yang dilakukan di masa depan, seperti memprediksikan hal yang baru atau akan tiba. Dengan *data mining* pencarian data akan lebih mudah karena dalam *data mining* memiliki tahapan-tahapan yang membantu dalam melakukan proses tersebut. Selain itu juga dengan data mining dapat meningkatkan pencarian informasi sehingga menjadi lebih baik [10].

### Klasifikasi

Klasifikasi adalah proses untuk menemukan sebuah model atau fungsi yang menjelaskan konsep atau kelas data untuk kepentingan tertentu. Tujuan utama dari klasifikasi adalah untuk membuat data yang sebelumnya tidak terlihat dinyatakan kelasnya seakurat mungkin. Dalam klasifikasi juga memiliki 2 pemodelan yaitu secara deskriptif dan prediktif. Pemodelan Deskriptif adalah pemodelan perangkat yang digambarkan untuk membedakan objek-objek dari kelas yang berbeda. Sedangkan prediktif adalah pemodelan yang digunakan sebagai prediksi label kelas untuk data yang tidak diketahui atau dikenal [12].

### Decision Tree

*Decision Tree* (Pohon Keputusan) merupakan representasi sederhana dari teknik klasifikasi yang memiliki fungsi dengan tujuan untuk memetakan setiap himpunan atribut ke salah satu label kelas yang telah didefinisikan sebelumnya. *Decision Tree* adalah salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasikan. Dengan kemampuannya untuk membuat suatu proses menjadi lebih mudah sehingga dalam pengambilan keputusan yang kompleks dapat menjadi lebih sederhana. *Decision Tree* juga dapat menemukan hubungan

yang tersembunyi antara sejumlah data dari sebuah variabel input dengan data variabel target. Selain itu *Decision Tree* dapat memadukan antara eksplorasi data dan pemodelan, sehingga sangat baik sebagai langkah awal dalam proses pemodelan. Dengan menggunakan metode *Decision Tree*, kita dapat menghindari munculnya permasalahan dengan menggunakan kriteria yang jumlahnya lebih sedikit dari setiap node internal tanpa banyak mengurangi kualitas keputusan yang dihasilkan. Kekurangan dari *Decision Tree* diantaranya adalah terjadi overlap terutama ketika kelas-kelas dan kriteria yang digunakan jumlahnya sangat banyak dan kualitas hasil keputusan yang didapatkan sangat tergantung pada bagaimana desain dari *Decision Tree* [13].

### Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan (decision tree). Algoritma ini digunakan untuk melakukan klasifikasi atau segmentasi yang bersifat prediktif. Cabang-cabang pohon keputusan merupakan pertanyaan klasifikasi dan daun-daunnya merupakan kelas-kelas atau segmen-segmennya. C4.5 merupakan pengembangan dari algoritma ID3 [14]. Iterative Dichotomiser 3 (ID3) adalah sebuah metode dasar yang digunakan untuk membangun pohon keputusan. Algoritma yang digunakan pada metode ini menggunakan konsep dari entropy informasi. Dengan kata lain pemilihan atribut dengan menggunakan Information Gain. Gain mengukur seberapa baiknya suatu atribut dan memisahkan training example ke dalam kelas target. Atribut dengan informasi dengan nilai tertinggi yang akan dipilih [15]. Serangkaian perbaikan yang dilakukan pada ID3 mencapai puncaknya dengan menghasilkan sebuah sistem praktis dan berpengaruh untuk decision tree yaitu C4.5. Adapun perbaikannya adalah sebagai berikut [14]:

1. Algoritma C4.5 menghitung gain ratio untuk masing-masing atribut, dan atribut yang memiliki nilai yang tertinggi akan dipilih sebagai simpul atau node. Penggunaan gain ratio ini memperbaiki kelemahan dari ID3 yang menggunakan information gain.
2. Pemangkasan (pruning) dapat dilakukan pada saat pembangunan pohon(tree) ataupun pada saat proses pembangunan tree selesai.
  - a. Mampu menangani *continues attribute*.
  - b. Mampu menangani *missing data*.
  - c. Mampu membangkitkan rule dari sebuah *tree*.

### K-Fold Cross Validation

*K-Fold Cross Validation* adalah metode yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan cara melakukan perulangan dengan mengacak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut input yang acak. *K-Fold Cross Validation* menggunakan pengulangan k sama dengan berapa kali untuk membagi sebuah himpunan contoh secara acak menjadi k subset yang saling bebas, setiap ulangan disisakan satu subset untuk pengujian dan subset lainnya untuk pelatihan. Biasa digunakan dengan  $K = 5$  atau 10 fold untuk memperkirakan tingkat kesalahan yang terjadi, sebab data training pada setiap *fold* cukup berbeda dengan data training yang asli [23].

**Bahasa Pemrograman R**

R merupakan perangkat lunak dan juga bahasa pemrograman yang digunakan untuk melakukan analisis dan manipulasi data statistika dan grafik. R mudah digunakan dengan bantuan dari *user community* karena bersifat *open-software nature* dan terdapat banyak fungsi baru dari hasil kontribusi R.

**Kappa**

Kappa merupakan ukuran yang menyatakan konsistensi dari suatu pengukuran yang dilakukan dua orang penilai atau konsistensi antara dua metode pengukuran. Dengan kata lain pengukuran yang dilakukan oleh dua orang untuk menentukan nilai atau hasil serta kecocokannya. Contohnya melakukan pengecekan perhitungan pada 50 soal untuk melihat apakah nilai yang didapat adalah 100. Pengguna pertama dan kedua melakukan pengecekan untuk mengetahui apakah soal-soal tersebut bisa memperoleh nilai 100 atau tidak. Dengan melakukan pengecekan tersebut maka akan dilihat apakah kecocokan dari pengguna pertama dan kedua memiliki nilai yang saling konsisten atau mendekati [27].

Formula dari Kappa Statistik [28] yaitu:

$$Kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

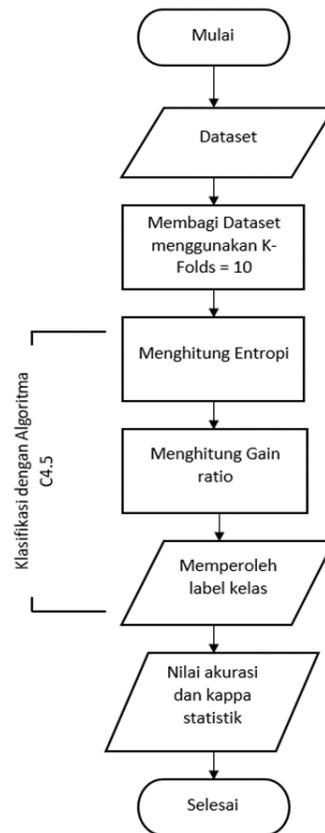
**III. METODE PENELITIAN**

Dalam penelitian ini akan dilakukan klasifikasi pada dataset yang berasal dari UCI *Machine Learning Repository* [36]. Dataset tersebut adalah dataset *Breast-Cancer* dan *Breast-W-Original*. Dalam penelitian ini data yang ada pertama *dataset breast-cancer* yang memiliki 10 atribut dan 286 *instance* dan *dataset* kedua yaitu *breast-wisconsin-original* memiliki 10 atribut dan 699 *instance*.

**IV. HASIL DAN PEMBAHASAN**

Aplikasi yang akan dibangun adalah “Klasifikasi Penyakit Kanker Menggunakan Algoritma *Decision Tree C4.5*”. Aplikasi ini dibuat untuk mengklasifikasikan penyakit kanker dengan melakukan uji akurasi dan Kappa dengan menggunakan algoritma *Decision Tree C4.5*.

Pembangunan aplikasi ini menggunakan metode analisis KDD (*Knowledge Discovery in Database*) yang disesuaikan dengan kebutuhan yang diperlukan seperti *data selection* dan *preprocessing*. Bahasa pemrograman yang digunakan dalam aplikasi ini yaitu R. Berikut ini merupakan model fungsional perangkat lunak atau *flowchart* dari sistem yang dibuat:



Gambar 1. Flowchart Klasifikasi Penyakit Kanker

Pada tahap ini akan dilakukan analisis data dengan menggunakan tahapan KDD (*Knowledge Discovery and Database*) dengan langkah-langkah sebagai berikut:

*Data Selection*

Tahap ini bertujuan untuk memilih dan meminimalkan jumlah data yang akan digunakan terdapat pada tabel daftar atribut dari data latih yang diambil:

Tabel 1. Daftar Atribut

Data	Nama Atribut	Type Data
Breast-Cancer	Age	Nominal
	Menopause	Nominal
	Tumor-size	Nominal
	Inv-nodes	Nominal
	Node-caps	Nominal
	Deg-malig	Nominal
	Breast	Nominal
	Breast-squad	Nominal
	Irradiat	Nominal
Class	Nominal	

*Preprocessing*

Pada data yang kosong diisi dengan nilai yang paling sering muncul. Pada atribut *Node\_caps* terdapat 8 nilai kosong yang akan diisi dengan nilai atribut terbanyak. Nilai atribut yang terdapat dalam atribut *node\_caps* adalah nilai *yes* dan nilai *no*. Setelah dilakukan perhitungan maka nilai terbanyak adalah nilai *no* maka nilai kosong diganti dengan *no*.

Pada atribut *Breast\_squad* terdapat 1 nilai kosong yang akan diisi dengan nilai atribut terbanyak. Pada atribut *Breast\_squad* terdapat 3 nilai atribut. Atribut itu adalah *left\_low*, *central*, dan *left\_up*. Setelah dilakukan perhitungan didapatkan nilai atribut *left\_up* lebih banyak. Dengan begitu nilai kosong diganti dengan nilai yang paling sering muncul yaitu *left\_up*.

Tabel 2. *Missing Value*

Data	Nama Atribut	Nilai Kosong	Nilai Pengganti
<i>Breast-w-original</i>	<i>Bare_nuclei</i>	14	1

Berikut ini merupakan hasil akurasi dan Kappa berdasarkan pengujian dari *dataset*:

Tabel 3. *Akurasi dari dataset Breast-W-Original*

ConfidenceFactor	MinNumObj	Akurasi
0.0100	1	0.9699988
0.0100	2	0.9685496
0.0100	3	0.9656510
0.0100	4	0.9670595
0.0100	5	0.9699580
0.1325	1	0.9728767
0.1325	2	0.9685496
0.1325	3	0.9671003
0.1325	4	0.9685087
0.1325	5	0.9742644
0.2550	1	0.9757338
0.2550	2	0.9714067
0.2550	3	0.9685289
0.2550	4	0.9685087
0.2550	5	0.9728152
0.3775	1	0.9785496
0.3775	2	0.9728146
0.3775	3	0.9728146
0.3775	4	0.9713659
0.3775	5	0.9713866
<b>0.5000</b>	<b>1</b>	<b>0.9842851</b>
0.5000	2	0.9785087
0.5000	3	0.9727939
0.5000	4	0.9756309
0.5000	5	0.9742230

Berdasarkan tabel akurasi diatas telah didapat hasil akurasi yang optimal yaitu dengan nilai *ConfidenceFactor* = 0.5000 dengan akurasi = 0.9842851.

Tabel 4. *Akurasi dari dataset Breast-Cancer*

ConfidenceFactor	MinNumObj	Akurasi
0.0100	1	0.7373947
0.0100	2	0.7373974
0.0100	3	0.7442939
0.0100	4	0.7407225
0.0100	5	0.7616585
0.1325	1	0.7901067
0.1325	2	0.7721264
0.1325	3	0.7792693

ConfidenceFactor	MinNumObj	Akurasi
0.1325	4	0.77911461
0.1325	5	0.77911461
0.2550	1	0.8114122
0.2550	2	0.8078407
0.2550	3	0.7899836
0.2550	4	0.7898604
0.2550	5	0.7934319
0.3775	1	0.8712644
0.3775	2	0.8394910
0.3775	3	0.8078407
0.3775	4	0.8004516
0.3775	5	0.7931856
<b>0.5000</b>	<b>1</b>	<b>0.8954296</b>
0.5000	2	0.8605665
0.5000	3	0.8179557
0.5000	4	0.8144901
0.5000	5	0.8036535

Tabel 5. *Kappa dari dataset Breast-W-Original*

ConfidenceFactor	MinNumObj	Kappa
0.0100	1	0.9352044
0.0100	2	0.9321023
0.0100	3	0.9255897
0.0100	4	0.9285939
0.0100	5	0.9349201
0.1325	1	0.9404076
0.1325	2	0.9311876
0.1325	3	0.9282777
0.1325	4	0.9311026
0.1325	5	0.9438908
0.2550	1	0.9470061
0.2550	2	0.9377816
0.2550	3	0.9315394
0.2550	4	0.9307867
0.2550	5	0.9407886
0.3775	1	0.9531734
0.3775	2	0.9407134
0.3775	3	0.9408319
0.3775	4	0.9369363
0.3775	5	0.9373423
<b>0.5000</b>	<b>1</b>	<b>0.9655826</b>
0.5000	2	0.9531358
0.5000	3	0.9406873
0.5000	4	0.9460811
0.5000	5	0.9433481

Tabel 6. *Kappa dari dataset Breast-Cancer*

ConfidenceFactor	MinNumObj	Kappa
0.0100	1	0.2130987
0.0100	2	0.2130987
0.0100	3	0.2369470
0.0100	4	0.2308045
0.0100	5	0.2960676
0.1325	1	0.4185370
0.1325	2	0.3706677
0.1325	3	0.4008769

ConfidenceFactor	MinNumObj	Kappa
0.1325	4	0.3877822
0.1325	5	0.3877822
0.2550	1	0.4844080
0.2550	2	0.4740444
0.2550	3	0.4361874
0.2550	4	0.4278618
0.2550	5	0.4353228
0.3775	1	0.6551209
0.3775	2	0.5595870
0.3775	3	0.4917853
0.3775	4	0.4660190
0.3775	5	0.4398414
<b>0.5000</b>	<b>1</b>	<b>0.7316793</b>
0.5000	2	0.6428806
0.5000	3	0.5321072
0.5000	4	0.5174434
0.5000	5	0.4845972

Tabel 7. Perbandingan Akurasi dan Kappa Statistik dari Dataset Breast-W-Original dengan dengan parameter  $K = 5$  untuk KNN dan  $\text{Epsilon} = 0,0005$  dan  $\text{Gamma} = 0.5$  untuk SVM.

Algoritma	Akurasi	Kappa Statistik
Decision Tree C4.5	0.783217	0.3995
SVM	0.723776	0.1732
KNN	0.730769	0.2454
Naïve Bayes	0.959943	0.9127

Dari hasil yang didapat maka dapat dilihat setelah melakukan perubahan nilai parameter, *Decision Tree C4.5* memiliki nilai akurasi dan kappa yang lebih tinggi dari algoritma lainnya. Berikut adalah perbandingan nilai akurasi dan kappa dengan menggunakan settingan *standard* pada *weka*:

Tabel 8. Perbandingan Akurasi dan Kappa Statistik dari Dataset Breast-W-Original dengan menggunakan parameter *standard* dari Weka.

Algoritma	Akurasi	Kappa Statistik
Decision Tree C4.5	0.972818	0.9406
SVM	0.959943	0.9134
KNN	0.95279	0.8948
Naïve Bayes	0.959943	0.9127

Dari hasil di atas dapat dilihat Akurasi dan Kappa Statistik dengan nilai tertinggi diperoleh algoritma *Decision Tree C4.5*. Selanjutnya algoritma SVM dan Naïve Bayes tetapi nilai kappa SVM lebih besar dari Naïve Bayes. Sedangkan KNN memiliki nilai Akurasi dan Kappa Statistik terendah.

Tabel 9. Perbandingan Akurasi dan Kappa Statistik dari Dataset Breast-Cancer dengan menggunakan parameter *standard* dari Weka.

Algoritma	Akurasi	Kappa Statistik
Decision Tree C4.5	0.783217	0.3995
SVM	0.699301	0.0118
KNN	0.685315	0.2147

Algoritma	Akurasi	Kappa Statistik
Naïve Bayes	0.716783	0.2857

Dengan hasil di atas dapat dilihat Akurasi dan Kappa Statistik dengan nilai tertinggi didapat oleh algoritma *Decision Tree C4.5*. Selanjutnya algoritma Naïve Bayes. Diikuti dengan SVM dan KNN. Tetapi nilai Akurasi pada SVM lebih besar dari KNN sedangkan pada SVM nilai kappa statistik lebih besar dari KNN.

## V. KESIMPULAN DAN SARAN

Berdasarkan hasil dari penelitian yang dilakukan maka dapat disimpulkan bahwa :

1. Aplikasi Klasifikasi Penyakit Kanker Payudara menggunakan Algoritma *Decision Tree C4.5* telah berhasil dibangun dan telah dilakukan perbandingan dengan Algoritma KNN, Naïve Bayes, dan SVM.
2. Berdasarkan hasil pengujian pada dua *dataset*, Algoritma *Decision Tree* lebih unggul dari metode pembandingan baik dalam uji akurasi maupun nilai Kappa Statistik. Nilai akurasi dari *Dataset Breast-Cancer* adalah 0,78 % dan Kappa Statistiknya 0,39 dan pada *Dataset Breast-W-Original* nilai Akurasinya sebesar 0,97 % dan Kappa Statistiknya adalah 0,94.
3. Dari hasil perbandingan Algoritma *Decision Tree C4.5* dengan Algoritma SVM, KNN, dan Naïve Bayes diperoleh Algoritma *Decision Tree C4.5* sebagai Algoritma yang paling unggul baik dari akurasi maupun kappa statistik.
4. Aplikasi dapat berjalan dengan baik pada ke empat *browser*, tetapi *browser Google Chrome* lebih disarankan untuk menjalankan aplikasi karena proses dalam membaca *link address* dari *browser* tersebut paling cepat.

Sebagai saran untuk penelitian selanjutnya adalah :

1. Diharapkan dapat menggunakan lebih banyak *dataset* untuk pengujian.
2. Algoritma C4.5 dapat digabungkan dengan algoritma optimasi untuk hasil yang lebih baik.

## DAFTAR PUSTAKA

- [1] A. Sainal, M. Nyorong, I. Maria, "Risiko Gaya Hidup Terhadap Kejadian Kanker Payudara Pada Wanita," *Jurnal MKMI*, vol. 13, pp. 157-166, 2017.
- [2] S. Prof. Dr. dr. Nila Farid Moeloek, *Kanker Payudara*, Jakarta: Kementerian Kesehatan Republik Indonesia, 2016.
- [3] Tukino, "Penerapan Algoritma C4.5 untuk memprediksi keuntungan pada PT SMOE Indonesia," *Jurnal Sistem Informasi*, vol. 9, pp. 1-8, 2018.
- [4] S. Oliver, T. Agus, H. Hidayatulah, "Diagnosis Tingkat Kesehatan Pasien Menggunakan Metode Decision Tree C4.5," *Seminar Nasional Informatika*, vol. 1, pp. 1-4, 2015.
- [5] S. Defit, Yuhandra. Y. A. Fiandra, "Penerapan Algoritma C4.5 untuk Klasifikasi Data Rekam Medis berdasarkan International Classification Diseases (ICD-10),"

- Rekayasa Sistem dan Teknologi Informasi , vol. 1, pp. 1-9, 2017.
- [6] Y. Kurniawan, D. Febriyanto, "Prediksi Penyakit Tuberculosis (TBC) Menggunakan Algoritma C4.5," *Jurnal Ilmiah Sinus*, vol. 16, pp. 1-14, 2018.
- [7] L. Y. Hendrati, G. A. T. Dewi, "Analisis Resiko Kanker Payudara Berdasar Riwayat Pemakaian Kontrasepsi Hormonal Dan Usia Menarche," *Departemen Epidemiologi FKM UA*, vol. 1, pp. 1-12, 5 Maret 2017.
- [8] N. L. P. Suariyani, G. A. R. Dyanti, "Faktor-Faktor Keterlambatan Penderita Kanker Payudara Dalam Melakukan Pemeriksaan Awal Ke Pelayanan Kesehatan," *Jurnal Kesehatan Masyarakat*, vol. 1, pp. 1-9, 25 9 2016.
- [9] H. Setyawan, D. Sutiningshi, I. Yulianti, "Faktor-Faktor Resiko Kanker Payudara (Studi Kasus Pada Rumah Sakit Ken Saras Semarang)," *Jurnal Kesehatan Masyarakat (E-Journal)*, vol. 4, pp. 1-9, 2016.
- [10] M. Steinbach, V. Kumar, P. N. Tan, *Introduction to Data Mining*, United States: Pearson Education, Inc., 2006.
- [11] J. Pei, J. Han, M. Kamber, *Data Mining: Concepts and Techniques*, 3rd Edition, USA: Morgan Kaufmann, 2011, pp. 7-8.
- [12] M. Jajuli, S. Defiyanti, "Integrasi Metode Klasifikasi Dan Clustering dalam Data Mining," *Konferensi Nasional Informatika*, vol. 1, pp. 1-6, 2015.
- [13] M. Dahria, Z. Azmi, "Decission Tree Berbasis Algoritma Untuk Pengambilan Keputusan," *Jurnal Ilmiah Sains Dan Komputer*, vol. 3, pp. 157-164, 2013.
- [14] Rismayanti, "Implementasi Algoritma C4.5 Untuk Menentukan Penerima Beasiswa Di STT HARAPAN MEDAN," *Jurnal Media Infomata*, vol. 12, pp. 1-5, 2016.
- [15] O. Kristanto, "Penerapan Algoritma Klasifikasi Data Mining ID3 untuk Menentukan Penjurusan Siswa SMAN 6 Semarang," *Jurusan Teknik Informatika, FASILKOMUDINUS*, vol. 1, pp. 1-9, 2014.
- [16] W. Nengsih, "Entropy Reduction Analysis dalam Penentuan Pohon Keputusan untuk Klasifikasi Data Carry Over," *Politeknik Caltex Riau Pekanbaru*, vol. 1, pp. 1-5, 2017.
- [17] S. Hansun, D. H. Kamagi, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *Universitas Multimedia Nusantara*, vol. 6, pp. 1-7, 2014.
- [18] D. Ayu, Nursela, "Penerapan Algoritma C4.5 untuk Klasifikasi Tingkat Keganasan Kanker Payudara," *Fakultas Ilmu Komputer*, vol. 1, pp. 1-5, 2010.
- [19] Mardji, E. Santoso, I. U. Maulina, "Klasifikasi Kanker Payudara Menggunakan Decission Tree Dengan Algoritma Iterative Dichotomizer-3 (ID-3)," *Program Teknologi Informasi dan Ilmu Komputer*, vol. 1, pp. 1-8, 2017.
- [20] I. Yulianti, R. Wajhillah, "Penerapan Algoritma C4.5 untuk Prediksi Penggunaan Jenis Kontrasepsi Berbasis Web," *Kumpulan jurnal Ilmu Komputer*, vol. 4, pp. 1-14, 2017.
- [21] Liliana, Swastina, "Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa," *Jurnal Gema Aktualita*, vol. 2, pp. 1-6, 2013.
- [22] M. Rahman, "Algoritma C4.5 untuk Menentukan Mahasiswa Penerima Beasiswa," *Jurnal Tim Darmajaya*, vol. 1, pp. 1-11, 2015.
- [23] I. Budiman, A. Farmadi, M. A. Banjarsari, "Penerapan K-Optimal Pada Algoritma KNN untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Ilmu Komputer Fmipa Unlam Berdasarkan IP Sampai Dengan Semester 4," *Kumpulan Jurnal Ilmu Komputer*, vol. 2, pp. 1-15, 2015.
- [24] H. Budiadma, "5 Perbedaan Akurasi dan Presisi dalam Pengukuran," *Usaha 321 Ragam Informasi*, 13 Juni 2019. [Online]. Available: <https://usaha321.net/perbedaan-akurasi-dan-presisi.html>. [Accessed 15 Juli 2019].
- [25] M. Bijaksana, A. Romadhony, B. Utomo, "Klasifikasi Jawaban dengan Menggunakan Multiple Features Extraction pada Community Question Answering Answer Classification using Multiple Features Extraction in Community Question Answering," *e-Proceeding of Engineering*, vol. 3, pp. 1-11, 2016.
- [26] C. Manliguez, "Generalized Confusion Matrix for Multiple Classes," *University of the Philippines*, vol. 1, pp. 1-3, 2016.
- [27] Supargiyono, E. H. Murhandrawati, F. Mau, "Koefisien Kappa sebagai Indes Kesepekatan Hasil Diagnosis Mikroskopis Malaria di Kabupaten Belu Nusa Tenggara Timur," *Buletin Penelitian Kesehatan*, vol. 43, pp. 1-8, 2015.
- [28] D. Napitupulu, "Studi Validitas Dan Reabilitas Faktor Sukses Implementasi E-Government Berdasarkan Pendekatan Kappa," *Jurnal Sistem Informasi*, vol. 1, pp. 1-5, 2014.