

DIAGNOSIS PENYAKIT KANKER PAYUDARA MENGGUNAKAN METODE NAIVE BAYES BERBASIS DESKTOP

Achmad Ramadhan Safutra¹, Dwi Wahyu Prabowo¹

¹Jurusan Sistem Informasi, Fakultas Ilmu Komputer
Universitas Darwan Ali, Sampit

Email: apakatarama@gmail.com, dwi.wahyu9@gmail.com

Abstraksi - Kanker payudara (*Carcinoma mamae*) dalam bahasa Inggrisnya disebut *breast cancer* merupakan kanker pada jaringan payudara. Saat ini, kanker payudara merupakan penyebab kematian kedua akibat kanker pada wanita, setelah kanker leher rahim, dan merupakan kanker yang paling banyak ditemui diantara wanita.

Ada beberapa tahapan pemeriksaan lebih lanjut dalam mendeteksi penyakit kanker payudara apakah termasuk dalam kategori jinak atau ganas. Pemeriksaan Patologi Anatomi dilakukan di laboratorium dengan memeriksa contoh jaringan tumor yang diambil melalui biopsi. Dalam pemberian keputusan, dokter patologi anatomi harus menghitung apa yang telah dihasilkan oleh laboratorium dalam pemeriksaan seorang pasien. Perhitungan tersebut memerlukan waktu yang tidak singkat karena menyangkut vonis seorang pasien kanker payudara.

Tujuan dari penelitian ini adalah melakukan diagnosis penyakit kanker payudara berbasis komputer (aplikasi) yang dapat menghasilkan vonis apakah seorang pasien masuk dalam kategori jinak atau ganas. Aplikasi diagnosis penyakit kanker payudara ini menggunakan *dataset Wisconsin Diagnostic* sebagai data pembelajarannya. Metode *Naïve Bayes* diterapkan dalam aplikasi diagnosis penyakit kanker payudara karena teknik klasifikasi dari *Naïve Bayes* tersebut dapat menghasilkan hasil diagnosis yang akurat sebesar 98,6726%.

Kata Kunci— Penyakit Kanker Payudara, *Wisconsin Diagnostic*, *Naïve Bayes*

I. PENDAHULUAN

Kanker adalah penyakit yang kurang disadari oleh sebagian masyarakat, karena sebagian masyarakat yang terkena penyakit ini kurang mengetahui kanker itu sendiri dan kurangnya melakukan deteksi dini terhadap kanker. Akibatnya sebagian besar kanker ditemukan pada stadium lanjut dan sulit ditanggulangi sehingga memberikan beban yang besar bagi pasien kanker. Kanker buah dari perubahan sel yang mengalami pertumbuhan tidak normal dan tidak terkontrol [1].

Kanker Payudara (*Carcinoma mamae*) dalam Bahasa Inggrisnya disebut *breast cancer* merupakan kanker pada jaringan payudara [2]. Saat ini, kanker payudara merupakan penyebab kematian kedua akibat kanker pada wanita, setelah kanker leher rahim, dan merupakan kanker yang paling banyak ditemui diantara wanita. Berdasarkan data dari American Cancer Society, sekitar 1,3 juta wanita terdiagnosa menderita kanker payudara, dan tiap tahunnya diseluruh dunia kurang lebih 465.000 wanita meninggal oleh karena penyakit ini [1].

Ada beberapa tahapan pemeriksaan lebih lanjut dalam mendeteksi penyakit kanker payudara apakah termasuk dalam kategori jinak atau ganas. Pemeriksaan patologi anatomi dilakukan di laboratorium dengan memeriksa contoh jaringan tumor yang diambil melalui biopsi. Biopsi adalah pemeriksaan dengan mengambil sampel jaringan dari benjolan yang dicurigai kanker untuk kemudian diperiksa di bawah mikroskop apakah terdapat sel kanker atau tidak. Tujuan dari pemeriksaan Patologi Anatomi ini adalah untuk menentukan apakah jenis sel kanker masuk dalam kategori jinak atau ganas melalui keputusan yang diberikan oleh dokter patologi anatomi [3].

Dalam hal pemberian keputusan, dokter patologi anatomi harus menghitung apa yang sudah dihasilkan oleh laboratorium patologi anatominya. Perhitungan dokter memerlukan waktu yang tidak singkat karena menyangkut vonis seorang pasien kanker payudara. Oleh karena itu, hal ini memotivasi pengembangan suatu metode komputer (aplikasi) yang dapat mendiagnosis keberadaan penyakit kanker payudara. Aplikasi ini dapat menyediakan prosedur diagnosis penyakit kanker payudara terhadap pasien dengan waktu yang singkat dan memudahkan pekerjaan dokter.

Dewasa ini terdapat penelitian yang mengembangkan teknik komputasi cerdas untuk mendiagnosis penyakit kanker payudara, salah satu yang pernah diusulkan adalah dengan menggunakan metode *Naïve Bayes*. Metode *Naïve Bayes* [4] diusulkan untuk mendiagnosis penyakit kanker payudara. Penelitian terdahulu menggunakan *dataset Wisconsin Prognostic* sebagai data pembelajaran dan menghasilkan akurasi sebesar 74,24%. Namun jika dilihat dari akurasi tersebut, akurasi sebesar 74,24% belum cukup baik dalam hal performa.

Penelitian yang akan dikembangkan akan menggunakan *dataset Wisconsin Diagnostic* [5] sebagai data pembelajarannya. *Dataset Wisconsin Diagnostic* merupakan *dataset* populer dan secara luas digunakan oleh peneliti data mining untuk mendiagnosis penyakit kanker payudara. Penggunaan *dataset Wisconsin Diagnostic* ini diharapkan dapat meningkatkan akurasi diagnosis. Oleh karena itu, pada penelitian ini diusulkan pembuatan aplikasi berbasis desktop dengan menggunakan metode *Naïve Bayes* untuk mendiagnosis penyakit kanker payudara.

II. LANDASAN TEORI

A. Kanker Payudara

Kanker payudara adalah keganasan yang berasal dari sel kelenjar, saluran kelenjar dan jaringan penunjang payudara,tidak termasuk kulit payudara [2]. Saat ini, kanker payudara merupakan penyebab kematian kedua akibat kanker pada wanita, setelah kanker leher rahim, dan merupakan kanker yang paling banyak ditemui diantara wanita. Berdasarkan data dari *American Cancer Society*, sekitar 1,3 juta wanita terdiagnosis menderita kanker payudara,dan tiap tahunnya diseluruh dunia kurang lebih 465.000 wanita meninggal oleh karena penyakit ini [1].

B. Kecerdasan Buatan

Kecerdasan buatan berasal dari bahasa inggris “*Artificial Intelligence*” atau disingkat AI, yaitu *intelligence* adalah sifat yang berarti cerdas, sedangkan *artificial* artinya buatan. Kecerdasan buatan yang dimaksud di sini merujuk pada mesin yang mampu berpikir, menimbang tindakan yang akan diambil, dan mampu mengambil keputusan seperti yang dilakukan oleh manusia. Rich dan Knight mendefinisikan kecerdasan buatan sebagai sebuah studi tentang bagaimana membuat komputer melakukan hal-hal yang pada saat ini dapat dilakukan lebih baik oleh manusia [6].

C. Metode Konvensional untuk mendiagnosis penyakit kanker payudara

Ada beberapa metode konvensional yang biasa dilakukan untuk mendiagnosis penyakit kanker payudara, antara lain :

1. Biopsi

Biopsi merupakan metode diagnosis penyakit kanker payudara dengan melakukan pengambilan sampel jaringan dari benjolan yang dicurigai kanker untuk kemudian diperiksa dibawah mikroskop apakah terdapat sel kanker atau tidak. Kelemahannya biopsi bersifat *invasive*, mahal serta menimbulkan luka terhadap pasien [7].

2. Patologi Anatomi

Pemeriksaan Patologi Anatomi dilakukan di laboratorium dengan memeriksa contoh jaringan tumor yang diambil melalui teknik biopsi. Tujuan dari pemeriksaan Patologi Anatomi ini adalah untuk menentukan apakah jenis sel kanker masuk dalam kategori jinak atau ganas melalui keputusan yang diberikan oleh dokter patologi anatomi [8].

D. Metode Naïve Bayes

Naïve Bayes merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai Teorema Bayes. Teorema tersebut dikombinasikan dengan *Naïve* dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *Naïve Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya [9].

Persamaan (1) dari teorema *Bayes* (Bustami) adalah:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Keterangan:

X : Data dengan *class* yang belum diketahui

H : Hipotesis data X merupakan suatu *class* spesifik

P(H|X) : Probabilitas hipotesis H berdasar kondisi X (*posteriori probability*)

P(H) : Probabilitas hipotesis H (*prior probability*)

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

P(X) : Probabilitas X

Karena dalam klasifikasi nilai P(X) selalu konstan untuk semua kelas, maka:

$$P(H|X) = P(X|H) \cdot P(H) \quad (2)$$

E. Akurasi

Diagnosa terhadap kanker payudara memiliki empat buah kemungkinan yang dapat dicapai yaitu *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN). *True positive* dan *true negative* merupakan hasil diagnosis yang benar. *False positive* adalah kesalahan prediksi yang terjadi ketika seorang pasien mengalami kanker payudara tipe jinak tetapi diprediksi mengalami kanker payudara tipe ganas. *False negative* adalah kesalahan prediksi yang terjadi ketika seorang pasien mengalami kanker payudara tipe ganas tetapi diprediksi mengalami kanker payudara tipe jinak [10]. Tabel 1 menunjukkan *confusion matrix* yang memberikan perbandingan antara *true positive*, *true negative*, *false positive*, dan *false negative*.

TABEL 1
CONFUSION MATRIX DARI HASIL DIAGNOSIS

		Kelas hasil prediksi	
		Kanker Jinak	Kanker Ganas
Kelas Sebenarnya	Kanker Jinak	TP	FN
	Kanker Ganas	FP	TN

Persamaan (3) menunjukkan perhitungan akurasi dalam suatu aplikasi .

$$\text{Akurasi} = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \quad (3)$$

III. ANALISIS DAN PERANCANGAN SISTEM

A. Dataset Wisconsin Diagnostic

Penelitian ini menggunakan data dari *dataset Wisconsin Diagnostic* yang digunakan sebagai *data training* untuk mendiagnosis kanker payudara. Total atribut yang dimiliki oleh *dataset Wisconsin Diagnostic* adalah sebanyak 10 atribut [5]. Tabel berikut menjelaskan tentang setiap atribut dalam *dataset Wisconsin* .

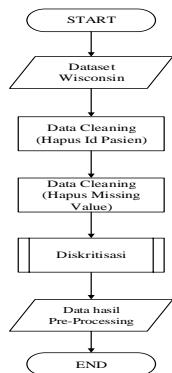
TABEL 2
RINGKASAN ATRIBUT DATASET WISCONSIN

Atribut	Nilai
Clump Thickness	1 – 10
Uniformity of Cell Size	1 – 10
Uniformity of Cell Shape	1 – 10
Marginal Adhesion	1 – 10
Single Epithelial Cell Size	1 – 10
Bare Nuclei	1 – 10
Bland Chromatin	1 – 10
Normal Nucleoli	1 – 10
Mitoses	1 – 10
Class	2 for benign, 4 for malignant

B. Metode

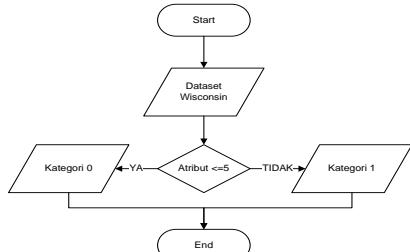
1) Data Preprocessing

Data preprocessing merupakan langkah awal dalam mendiagnosa kanker payudara. Pada penelitian ini data preprocessing terbagi menjadi data cleaning (menghapus id pasien dan missing value) dan diskritisasi. Gambar 1 memperlihatkan langkah-langkah pada tahap data preprocessing.



Gambar 1 Data Preprocessing

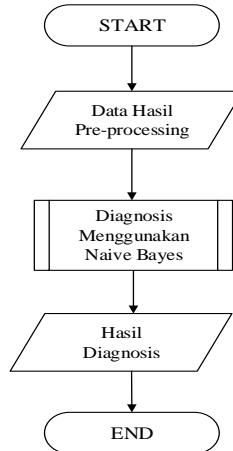
Langkah selanjutnya adalah diskritisasi data. Gambar 2 memperlihatkan langkah-langkah pada tahap diskritisasi data.



Gambar 1 Diskritisasi Data

2) Diagnosis Kanker Payudara

Data processing dilakukan setelah pembuatan data preprocessing untuk mendiagnosa kanker payudara. Diagnosis kanker payudara berbasis komputer dilakukan dengan cara mengklasifikasikan *Dataset Wisconsin* menggunakan metode Naive Bayes. Gambar 3 memperlihatkan langkah-langkah pada tahap diagnosis kanker payudara.



Gambar 2 Data Processing (Alir diagnosis kanker payudara)

Berdasarkan Gambar 3, alir dimulai dari input hasil data preprocessing, kemudian data tersebut diproses menggunakan metode *Naive Bayes*, dan setelah itu akan didapatkan suatu hasil dari diagnosis tersebut apakah kanker yang diberita termasuk tipe jinak atau ganas.

3) Contoh Kasus

Berikut adalah langkah perhitungan probabilitas diagnosis kanker payudara dari contoh kasus berdasarkan aturan yang sudah ada di tabel :

Kasus saudara Tika dengan hasil laboratorium sebagai berikut:

TABEL 3
CONTOH KASUS

Nama Atribut	Hasil Pemeriksaan
Clump Thickness	1-5
Uniformity of Cell Size	1-5
Uniformity of Cell Shape	6-10
Marginal Adhesion	1-5
Single Epithelial Cell Size	1-5
Bare Nuclei	1-5
Bland Chromatin	1-5
Normal Nucleoli	1-5
Mitoses	1-5

1. Memisahkan 2 *Class* menjadi C1 untuk “Kanker Positif = Jinak” dan C2 untuk “Kanker Positif = Ganas”

2. Mencari *Prior Probability*:

- a. Probabilitas Kanker Positif dengan hasil Jinak

$$\frac{\text{Kanker Positif dengan hasil Jinak (C1)}}{\text{Total Record}} = \frac{270}{456} = 0,592105263$$

b. Probabilitas Kanker Positif dengan hasil Ganas

$$\frac{\text{Kanker positif dengan hasil Ganas (C2)}}{\text{Total Record}} = \frac{186}{456} = 0,407894737$$

3. Mencari Likelihood dari setiap atribut

a. Clump thickness

Mencari Likelihood pada atribut *clump thickness* dilakukan dalam dua tahap, yakni:

1. Likelihood *clump thickness* Jinak (karena di kasus, hasil yang diperoleh adalah 0) dengan hasil class C1 (Kanker Positif Jinak).

$$\frac{\text{Likelihood clump thickness Jinak}}{\text{C1}} = \frac{257}{270} = 0,951852$$

2. Likelihood *clump thickness* Jinak (karena di kasus, hasil yang diperoleh adalah 0) dengan hasil class C2 (Kanker Positif Ganas).

$$\frac{\text{Likelihood clump thickness Ganas}}{\text{C2}} = \frac{52}{186} = 0,27957$$

b. Uniformity of Cell Size

1. C1

$$\frac{\text{Likelihood Uniformity of cell size Jinak}}{\text{C1}} = \frac{267}{270} = 0,988889$$

2. C2

$$\frac{\text{Likelihood Uniformity of cell size Ganas}}{\text{C2}} = \frac{86}{186} = 0,462366$$

c. Uniformity of Cell Shape

1. C1

$$\frac{\text{Likelihood Uniformity of cell shape Jinak}}{\text{C1}} = \frac{5}{270} = 0,018519$$

2. C2

$$\frac{\text{Likelihood Uniformity of cell shape Ganas}}{\text{C2}} = \frac{108}{186} = 0,580645$$

d. Marginal Adhesion

1. C1

$$\frac{\text{Likelihood Marginal Adhesion Jinak}}{\text{C1}} = \frac{268}{270} = 0,992593$$

2. C2

$$\frac{\text{Likelihood Marginal Adhesion Ganas}}{\text{C2}} = \frac{105}{186} = 0,564516$$

e. Single Epithelial Cell Size

1. C1

$$\frac{\text{Likelihood Single Epithelial Cell Size Jinak}}{\text{C1}} = \frac{265}{270} = 0,981481$$

2. C2

$$\frac{\text{Likelihood Single Epithelial Cell Size Ganas}}{\text{C2}} = \frac{104}{186} = 0,55914$$

f. Bare Nuclei

1. C1

$$\frac{\text{Likelihood Bare Nuclei Jinak}}{\text{C1}} = \frac{265}{270} = 0,981481$$

2. C2

$$\frac{\text{Likelihood Bare Nuclei Ganas}}{\text{C2}} = \frac{50}{186} = 0,268817$$

g. Bland Chromatin

1. C1

$$\frac{\text{Likelihood Bland Chromatin Jinak}}{\text{C1}} = \frac{264}{270} = 0,977778$$

2. C2

$$\frac{\text{Likelihood Bland Chromatin Ganas}}{\text{C2}} = \frac{98}{186} = 0,526882$$

h. Normal Nucleoli

1. C1

$$\frac{\text{Likelihood Normal Nucleoli Jinak}}{\text{C1}} = \frac{263}{270} = 0,974074$$

2. C2

$$\frac{\text{Likelihood Normal Nucleoli Ganas}}{\text{C2}} = \frac{86}{186} = 0,462366$$

i. Mitoses

1. C1

$$\frac{\text{Likelihood Mitoses Jinak}}{\text{C1}} = \frac{269}{270} = 0,996296$$

2. C2

$$\frac{\text{Likelihood Mitoses Ganas}}{\text{C2}} = \frac{160}{186} = 0,860215$$

4. Mencari Likelihood Probability:

a. Probabilitas Diagnosis dengan hasil Kanker Positif Jinak berdasarkan 9 Likelihood Probability dari setiap atribut.

$$0,951852 * 0,988889 * 0,018519 * 0,992593 * 0,981481 * 0,981481 * 0,977778 * 0,974074 * 0,996296 = 0,015815$$

b. Probabilitas Diagnosis dengan hasil Kanker Positif Ganas berdasarkan 9 Likelihood Probability dari setiap atribut.

$$0,27957 * 0,462366 * 0,580645 * 0,564516 * 0,55914 * 0,268817 * 0,526882 * 0,462366 * 0,860215 = 0,001335$$

5. Menghitung Posterior Probability:

a. Likelihood C1 * C1

$$0,015815 * 0,996296 = 0,009364$$

b. Likelihood C2 * C2

$$0,001335 * 0,860215 = 0,000544$$

Akhirnya diketahui hasil diagnosis dengan membandingkan hasil *Posterior probability* C1 = 0,009346 dan C2 = 0,000544, sehingga hasil diagnosis adalah C1, atau Kanker Positif='Jinak', karena C1 lebih besar dari C2.

4) Confusion Matrix

Confusion matrix digunakan untuk pengukuran efektifitas klasifikasi. Dengan menggunakan persamaan (2-3), dilakukan proses perhitungan akurasi untuk 226 data uji. Vonis diagnosis dengan kelas sebenarnya dimasukkan dalam tabel testing, sedangkan Vonis diagnosis dengan kelas hasil aplikasi dimasukkan dalam tabel hasil diagnosis.

Dari 226 data uji, diperoleh nilai TP = 51, TN = 172, FP = 1, FN = 2. Langkah berikutnya yaitu menggunakan persamaan (3), maka perhitungan nilai akurasi sebagai berikut:

$$\text{Akurasi} = \frac{51 + 172}{51 + 2 + 1 + 172} \times 100\% = \frac{224}{226} = 0,986726 \\ = 98,6726\%$$

Kesimpulan yang diperoleh yaitu, dari 226 data uji diperoleh akurasi kecocokan kelas sebenarnya terhadap kelas prediksi dengan *Naïve Bayes* sebesar 99%.

IV. IMPLEMENTASI

A. Implementasi Aplikasi Diagnosis

Implementasi aplikasi diagnosis adalah menerangkan tentang penjelasan seputar cara untuk menjalankan sebuah sistem yang bertujuan untuk membentuk suatu hubungan yang komunikatif antara pengguna dengan aplikasi.

1) Menu Login

Menu ini menampilkan *form login* yang berfungsi untuk mengontrol pengaksesan pengguna (*user*) pada saat menjalankan aplikasi. Proses pertama pada saat *form login* berjalan, pengguna (*user*) diminta untuk memasukkan kode aksesnya pada kotak isian "Username" dan kotak isian "Password". Adapun form login dapat dilihat sebagai berikut :

Gambar 4 Form login

2) Menu Master Data Pasien

Gambar 5 Form input data pasien

3) Menu Transaksi Diagnosis

Gambar 6 Form transaksi diagnosis

4) Menu Akurasi

Menu ini akan menampilkan *form akurasi* yang berfungsi untuk meningkatkan performa suatu aplikasi. *Form akurasi* tersebut dapat dilihat sebagai berikut :

Gambar 7 Form akurasi

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Kesimpulan yang dapat diambil dari perancangan sistem ini adalah:

1. a. Penerapan metode *Naïve Bayes* pada aplikasi diagnosis kanker payudara diawali dengan pemilihan *dataset Wisconsin Diagnostic* sebagai data *training* untuk mendiagnosis penyakit kanker payudara. *Dataset* ini harus terlebih dahulu dilakukan tahapan saringan seperti tahapan *preprocessing* sebelum digunakan dalam diagnosis.
- b. Setelah dilakukan tahapan *preprocessing*, maka penerapan metode *Naïve Bayes* dapat diterapkan pada aplikasi diagnosis. Adapun tahapan metode *Naïve Bayes* untuk mendapatkan hasil kelas jinak atau ganas yaitu dengan mencari *prior probability*, kemudian mencari *likelihood probability* setiap atribut dan menghitung *posterior probability*.
2. Perancangan dan pembuatan aplikasi dilakukan dengan cara menyediakan hasil perhitungan kemungkinan kanker jinak dan ganas dalam setiap diagnosis seorang pasien. Dalam perhitungan tersebut, nilai kemungkinan yang tertinggi dapat menjadi kesimpulan hasil diagnosis.
3. Pembuatan persentase kemungkinan penyakit kanker payudara dapat dilakukan dengan perhitungan akurasi. Perhitungan akurasi ini menggunakan metode *Confusion Matrix* yang digunakan untuk pengukuran efektifitas klasifikasi. Hasil dari penggunaan metode *Confusion Matrix* telah menunjukkan bahwa aplikasi diagnosis penyakit kanker payudara dapat menghasilkan persentase kemungkinan yang akurat yaitu sebesar 98,6726%.

B. Saran

Pada penelitian ini proses diskritisasi dilakukan dengan cara mengkonversi data setiap atribut pada *dataset Wisconsin*

Diagnostic menjadi kategori 0 dan kategori 1. Proses diskritisasi tersebut masih bersifat subjektif yang artinya hasil diskritisasi masih bergantung pada peneliti. Untuk memberikan hasil diskritisasi yang objektif, maka disarankan untuk penelitian selanjutnya menggunakan metode-metode diskritisasi yang sudah ada.

REFERENSI

- [1] I. Rasjidi, *Deteksi Dini dan Pencegahan Kanker Pada Wanita*, Jakarta: Sagung Seto, 2009.
- [2] A. Aswita and D. F. A. Putri, *Sistem Pakar Diagnosa Penyakit Kanker Payudara Menggunakan Certainty Factor*.
- [3] C. E. Speicher, *Pemilihan Uji Laboratorium yang Efektif*, S. B. Kresno, Ed., Jakarta: Penerbit Buku Kedokteran EGC, 1994.
- [4] D. Dumitru, "Prediction of Recurrent Events in Breast Cancer Using the Naive Bayesian Classification," vol. 36(2), 2009.
- [5] "Breast Cancer Wisconsin Dataset," [Online]. Available: <https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/>. [Accessed 4 Oktober 2015].
- [6] T. Sutojo, E. Mulyanto and V. Suhartono, *Kecerdasan Buatan*, Yogyakarta: Penerbit ANDI, 2011.
- [7] S. R. Wardhani, *Biopsi Dalam Bidang Dermatologi*, p. 14.
- [8] P. R. Indonesia, Ed. *Kanker Payudara*.
- [9] Bustami, *Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasi Data Nasabah Asuransi*.
- [10] A. Indriani, *Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier*, 2014.