

KLASIFIKASI KECENDERUNGAN PENYAKIT MATA DI JAWA BARAT DENGAN ASSOCIATION RULE DAN NAÏVE BAYES CLASSIFIER

Zillan Taufiq Budiman¹⁾, Wina Witanti²⁾, Dian Nursantika³⁾

^{1), 2), 3)} Informatika Universitas Jenderal Achmad Yani Cimahi

Jl. Terusan Jendral Sudirman, Cimahi, Jawa Barat, 40285, Indonesia

Email : zillan_taufiq@ymail.com¹⁾, witanti@gmail.com²⁾, dianursantika@gmail.com³⁾

Abstrak

Klasifikasi kecenderungan penyakit mata untuk menentukan tempat dalam kegiatan bantuan kesehatan penyakit mata merupakan hal penting dalam salah satu proses meningkatkan derajat kesehatan penyakit mata di masyarakat. Pemanfaatan data rekam medis pasien pada tahun sebelumnya dapat dilakukan untuk melakukan pengelompokan data penyakit dari setiap daerah, namun hal tersebut menjadi kendala ketika data yang diolah berukuran sangat besar dan masih tidak obyektifnya pemilihan tempat untuk kegiatan bantuan kesehatan. Pada penelitian ini digunakan sebuah Metode Naïve Bayes Classifier untuk mengelompokkan data rekam medis dari berbagai daerah serta dari berbagai umur dan metode perhitungan Association Rule untuk mencari kombinasi antar item dari data rekam medis. Proses Association Rule dimulai dari menentukan bobot masing-masing penyakit sampai dengan hasil nilai tertinggi. Proses pengelompokan dimulai dari menentukan jumlah class setiap penyakit mata yang akan dihitung dan menghasilkan nilai klasifikasi penyakit mata. Penelitian ini, menggunakan data rekam medis sebanyak 12373 data pada bulan Pebruari sampai April tahun 2016 dan 13 penyakit mata. Berdasarkan hasil pengujian yang dilakukan memiliki penilaian sebesar 51.31% untuk penyakit mata Acute Atopic.

Kata kunci: Klasifikasi, Naïve Bayes Classifier, Association Rule.

1. Pendahuluan

Klasifikasi kecenderungan penyakit mata merupakan suatu proses untuk mengetahui potensi penyakit mata yang paling banyak diderita pasien, sehingga memiliki arti yang cukup penting bagi rumah sakit. Karena dengan adanya klasifikasi penyakit mata, rumah sakit dapat mengetahui kecenderungan penyakit mata dari banyak pasien di berbagai daerah. Hal ini membantu pihak rumah sakit untuk menganalisis penyakit mata dalam menentukan tempat untuk kegiatan bantuan kesehatan dan juga dapat membantu meningkatkan derajat kesehatan mata di lingkungan masyarakat khususnya di Jawa Barat, akan tetapi dalam agenda penyaluran kesehatan yang diselenggarakan oleh

pihak rumah sakit masih tidak obyektifnya akan pemilihan tempat untuk kegiatan bantuan kesehatan penyakit mata. Hal ini dikarenakan kurangnya informasi untuk menentukan tempat dalam kegiatan bantuan kesehatan. Rumah sakit merupakan suatu bagian dari organisasi medis dan sosial yang mempunyai fungsi untuk memberikan pelayanan kesehatan lengkap kepada masyarakat, baik secara kuratif maupun preventif, pelayanannya dapat menjangkau keluarga maupun lingkungan rumah. Rumah sakit dapat dijadikan pusat untuk latihan tenaga kesehatan dan penelitian biologi.

Association rule merupakan teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item[1] dan Naïve Bayes Classifier merupakan pengklasifikasian probabilitas sederhana berdasarkan pada teorema bayes. Teorema bayes dikombinasikan dengan "Naïve" yang berarti setiap atribut atau variable bersifat bebas (independent). Naïve Bayes Classifier dapat dilatih dengan efisien dalam pembelajaran terawasi, proses Naïve Bayes Classifier mengasumsikan bahwa ada atau tidaknya suatu fitur pada suatu kelas tidak berhubungan dengan ada atau tidaknya fitur lain dikelas yang sama[2].

Tujuan penelitian ini yaitu membangun perangkat lunak untuk membantu rumah sakit dalam menentukan tingkat penyakit mata pada setiap kota dan kabupaten di Jawa Barat untuk dapat merekomendasikan atau memberi informasi tempat dalam kegiatan bantuan kesehatan dengan menggunakan Association Rule dan Naïve Bayes Classifier. Sehingga daerah yang utama untuk dilakukan bantuan kesehatan dapat ditentukan. Penelitian ini didasarkan pada teori-teori yang mendukung pemecahan masalah dan bertujuan untuk memberikan gambaran pada penelitian yang dilakukan dengan association rule dan naïve bayes classifier. Tahapan tersebut antara lain yaitu :

a. Pengumpulan Data

Tahap dan cara pengumpulan data yang dilakukan dalam penelitian ini dengan cara teknik wawancara secara tatap muka dan Tanya jawab langsung antara pengumpul data maupun peneliti terhadap narasumber atau sumber data dengan tujuan untuk memperoleh informasi

b. Metode yang Digunakan

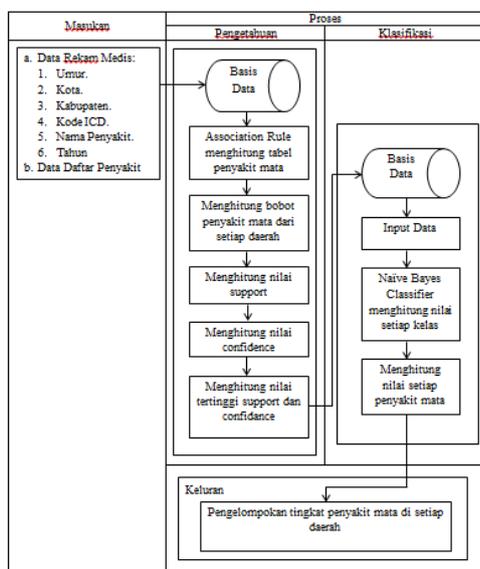
Metode yang digunakan pada penelitian ini yaitu *Association Rule* dilakukan untuk mencari kombinasi dari data rekam medis dan data penyakit mata dan Metode *Naive Bayes Classifier* digunakan untuk mengelompokkan penyakit mata dari berbagai kota dan kabupaten di Jawa Barat.

c. Metode Pengembangan Perangkat Lunak

Metode pengembangan perangkat lunak ini menggunakan Model Waterfall yang merupakan model untuk mengusulkan sebuah pendekatan kepada pengembangan software yang sistematis dan sekuensial yang meliputi aktivitas-aktivitas seperti analisis kebutuhan, desain, pengkodean, pengujian dan evaluasi, serta pemeliharaan.

d. Gambaran Umum Sistem

Gambaran umum sistem merupakan gambaran umum dari sistem yang akan dibangun untuk pembuatan perangkat lunak. Gambar metodologi penelitian dapat dilihat pada Gambar 1.



Gambar 1 Metodologi Penelitian

1. Masukan

Sistem yang akan dibangun pada penelitian ini terdiri dari masukan (*input*) data rekam medis pasien dan nama penyakit mata. Tabel masukan dapat dilihat pada Tabel 1.

Tabel 1 Deskripsi atribut pada variabel masukan

No	Atribut Data Rekam Medis	Nilai Atribut
1.	Umur	<1-12
2.	Umur	13-25
3.	Umur	26-40
4.	Umur	>40
5.	Kota	Nama kota
6.	Kabupaten	Nama kabupaten
7.	Nama penyakit	Nama penyakit
8.	Kode ICD	H10.1
9.	Kode ICD	H52.2

10.	Kode ICD	H01.0
11.	Kode ICD	H26.9
12.	Kode ICD	H53.2
13.	Kode ICD	H10.9
14.	Kode ICD	H00.0
15.	Kode ICD	H52.0
16.	Kode ICD	H51.1
17.	Kode ICD	H04.1
18.	Kode ICD	H57.8
19.	Kode ICD	H52.4
20.	Kode ICD	H40.2
21.	Tahun	Tahun

2. Proses

Pada tahap proses terdiri dari proses menghitung perbandingan jumlah item yang ada pada data daftar penyakit mata dan data daftar rekam medis seperti umur, kode ICD, kota, kabupaten, dan nama penyakit, selanjutnya menghitung nilai *support* dan menghitung nilai *confidence*, yang selanjutnya menghitung nilai tertinggi *support* dan *confidence*. Setelah itu dapat dilakukan proses menghitung nilai setiap kelas dan menghitung nilai setiap penyakit mata dari berbagai kota dan kabupaten dengan menggunakan *Naive Bayes Classifier* untuk mengelompokkan penyakit mata dari setiap daerah yang telah dihitung.

3. Keluaran

Keluaran atau output pada penelitian ini yaitu pengelompokan penyakit mata dengan tingkatan penyakit sesuai dengan kota dan kabupaten di Jawa Barat dalam menentukan tempat untuk kegiatan bantuan kesehatan.

2. Pembahasan

Association rule merupakan salah satu teknik di dalam data mining untuk menentukan hubungan antar item dalam suatu dataset (sekumpulan data) yang telah ditentukan. Teknik ini mencari kemungkinan kombinasi yang sering muncul (*frequent*) dari suatu *itemset* (sekumpulan item). terdapat dua langkah dalam algoritma ini, langkah pertama ialah melakukan perhitungan untuk menemukan *frequent itemset* dan langkah kedua mencari kaidah asosiasi (*rules*) dari sekumpulan *frequent itemset* [3]. mencari kombinasi item yang memenuhi syarat minimum dari nilai *support* dalam *database*. Nilai *support* sebuah item diperoleh dengan rumus berikut:

$$Support(A) = \frac{A}{\Sigma} \times 100\% \dots \dots \dots (1)$$

Keterangan : A = Merupakan item yang muncul bersamaan

$$\Sigma = \text{Total keseluruhan}$$

Sedangkan nilai *support* dari 2 item diperoleh dari rumus berikut:

$$Support (A \cap B) = \frac{A \text{ dan } B}{\Sigma} \times 100\% \dots\dots\dots(2)$$

Setelah semua pola frekuensi tinggi ditemukan, maka dicari aturan *assosiatif* yang memenuhi syarat minimum untuk *confidence* dengan menghitung aturan *assosiatif* $A \rightarrow B$. Nilai *confidence* $A \rightarrow B$ diperoleh dari rumus berikut:

$$Confidence = P(B|A) = \frac{Support A \rightarrow B}{A} \times 100\% \dots\dots\dots(3)$$

Proses asosiasi yang akan dilakukan yaitu dengan menghitung data rekam medis, data penyakit mata, dan kode ICD, yang bertujuan untuk mengasihkan tingkatan dari masing-masing penyakit mata per-daerah berdasarkan data yang telah ditetapkan. Pada proses asosiasi ini akan digunakan Metode *Association Rule*, dalam metode tersebut terdapat tahapan asosiasi yaitu menghitung tabel penyakit mata, menghitung bobot penyakit mata dari setiap daerah, perhitungan nilai *support*, perhitungan nilai *confidence* sampai menghitung nilai tertinggi *support* dan *confidence*. Langkah pertama akan dibuat tabel tabular dengan menghitung bobot penyakit mata dari setiap umur, kota dan kabupaten, dalam tabel tabular yang akan dihitung yaitu penyakit mata *Acute Atopic Conjunctivitis*. Tabel tabular data dapat dilihat pada Tabel 2.

Tabel 2. Tabular data

Pasien	Kode ICD	<1-12	Bdg	Kab. Idr	Cmh	Kab. Bdg
1	H10.1	1	1	0	1	0
2	H10.1	1	1	0	1	0
3	H10.1	1	1	0	1	0
4	H10.1	1	1	0	1	0
5	H10.1	1	1	0	0	1
6	H10.1	1	1	0	0	1
7	H10.1	1	1	0	0	1
8	H10.1	1	1	0	0	1
9	H10.1	1	1	0	0	1
18	H10.1	1	1	0	0	0
19	H10.1	1	1	0	0	0
20	H10.1	1	0	0	1	0
21	H10.1	1	0	0	1	0
22	H10.1	1	0	0	1	0
23	H10.1	1	0	0	1	0
24	H10.1	1	0	0	1	0
25	H10.1	1	0	0	1	0
26	H10.1	0	0	0	0	0
27	H10.1	0	0	0	0	0
28	H10.1	0	0	0	0	0
29	H10.1	0	0	0	0	0
30	H10.1	0	0	0	0	0
Σ		25	19	0	10	9

Setelah menghitung bobot maka langkah selanjutnya ialah menghitung nilai *support* dan

confidence. Pada langkah untuk menghitung nilai *support* ialah hitung Σ dari unsur – unsur item yang telah didapat dari langkah sebelumnya dengan dibagi jumlah seluruh data pasien lalu dikalikan 100%. Sedangkan *confidence* langkah untuk menghitungnya ialah Σ jumlah item yang terdapat dari unsur item lalu dibagi dengan jumlah *antecedent* dan dikalikan dengan 100%.. tabel nilai *support* dan *confidence* dapat dilihat pada Tabel 3.

Tabel 3. Nilai *support* dan *confidence*

No	If antecedent then consequent	Support	Confidence
1.	if H10.1 and <1-12 then Bdg	19/30 = 63,3%	19/30 = 63,3%
2.	if H10.1 and Bdg then <1-12	19/30 = 63,3%	19/30 = 63,3%
3.	if <1-12 and Bdg then H10.1	19/30 = 63,3%	19/25 = 76%
4.	if H10.1 and <1-12 then Cmh	10/30 = 33,3%	10/30 = 33,3%
5.	if H10.1 and Cmh then <1-12	10/30 = 33,3%	10/30 = 33,3%
6.	if <1-12 and Cmh then H10.1	10/30 = 33,3%	10/25 = 40%
7.	if H10.1 and <1-12 then kab. Bdg	9/30 = 30%	9/30 = 30%
8.	if H10.1 and kab. Bdg then <1-12	9/30 = 30%	9/30 = 30%
9.	if <1-12 and kab. Bdg then H10.1	9/30 = 30%	9/25 = 36%

Setelah langkah sebelumnya telah selesai maka langkah selanjutnya mengkalikan nilai *support* dan nilai *confidence* sehingga di dapat hasil nilai tertinggi. Tabel nilai tertinggi dapat dilihat pada Tabel 4.

Tabel 4. Hasil nilai tertinggi

No	If antecedent then consequent	Perkalian paling besar
1.	if <1-12 and Bdg then H10.1	0,48108
2.	if <1-12 and Cmh then H10.1	0,1332
3.	if <1-12 and kab. Bdg then H10.1	0,108

Naïve Bayes Classifier merupakan sebuah metode klasifikasi yang berakar pada teorema bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistic yang dikemukakan oleh ilmuwan inggris Thomas Bayes. Yaitu mencari peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dari *naïve bayes classifier* ialah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi atau kejadian [4].

$$P(A|B) = (P(B|A) * P(A))/P(B).....(4)$$

Peluang dari kejadian A sebagai B ditentukan dari peluang B saat A, peluang A, dan peluang B. pada pengaplikasiannya nanti rumus ini berubah menjadi:

$$P(C_i|D) = (P(D|C_i) * P(C_i))/P(D).....(5)$$

Naïve Bayes Classifier atau dapat disebut sebagai *multinomial*. *Naïve Bayes* merupakan model penyederhanaan dari metoda *bayes* yang cocok dalam pengklasifikasian. Persamaan *naïve bayes* sebagai berikut:

$$V_{map} = \arg \max P(V_j | a_1, a_2, \dots, a_n).....(6)$$

Menurut persamaan (6), maka persamaan (4) dapat ditulis sebagai berikut:

$$V_{map} = \frac{\arg \max_{v_j \in V} P(a_1, a_2, \dots, a_n | P(v_j))}{P(a_1, a_2, \dots, a_n)}.....(7)$$

Dimana : V_{map} = probabilitas tertinggi, a_1, a_2, \dots, a_n = atribut dan $P(a_1, a_2, \dots, a_n)$ konstan, sehingga dapat dihilangkan menjadi:

$$V_{map} = \arg \max_{v_j \in V} P(a_1, a_2, \dots, a_n | v_j) P(v_j).....(8)$$

Karena $p(a_1, a_2, \dots, a_n | v_j)$ sulit untuk dihitung, maka akan diasumsikan bahwa setiap kata pada data atau dokumen tidak mempunyai keterkaitan.

Langkah selanjutnya ialah menghitung *Naïve Bayes Classifier*. Pada tahapan ini yaitu proses untuk pengelompokkan setiap penyakit mata dari berbagai kota dan kabupaten yang dilakukan dengan menggunakan perhitungan *Naïve Bayes Classifier*. Langkah pertama yaitu membuat keterangan penyakit mata, kota serta kabupaten untuk mempermudah langkah dalam menentukan nilai n_c untuk setiap class, berikut perhitungan pertama dalam keterangan penyakit mata.

Keterangan penyakit mata *Acute Atopic Conjunctivitis*, Kota Bandung:

1. H10.1
2. Bdg
3. <1-12
4. 13-25
5. 26-40
6. >40

Langkah selanjutnya menentukan nilai n_c untuk setiap class yang sesuai dengan keterangan yang telah dibuat. Berikut langkah menentukan nilai n_c untuk setiap class. Tabel menentukan nilai n_c dapat dilihat pada Tabel 5.

Tabel 5. Menentukan nilai n_c

	H10.1	Bdg	<1-12	13-25	25-40	>40	Ket
H10.1 dan Bdg	1	1	1	1	1	1	Yes
H10.1 dan Cmh	1	0	1	1	1	1	No

H10.1 dan Kab. Idr	1	0	1	1	1	1	No
H10.1 dan Kab. Bdg	1	0	1	1	1	1	No

Kemudian proses selanjutnya menghitung nilai $P(A_i | V_j)$ dan menghitung nilai $P(V_j)$ dalam langkah ini proses yang akan ditentukan berdasarkan langkah pertama yang telah diraih prosesnya. Berikut langkah kedua dalam menghitung nilai $P(A_i | V_j)$ dan menghitung $P(V_j)$. Tabel menghitung nilai $P(A_i | V_j)$ dan $P(V_j)$ dapat dilihat pada Tabel 6.

Tabel 6. Menghitung nilai $P(A_i | V_j)$ dan $P(V_j)$

	a	b	c	d	e	f
H10.1 dan Bdg	0.10 25	0.102 5	0.10 25	0.10 25	0.10 25	0.10 25
H10.1 dan Cmh	0.10 25	0.071 25	0.10 25	0.10 25	0.10 25	0.10 25
H10.1 dan Kab. Idr	0.10 25	0.071 25	0.10 25	0.10 25	0.10 25	0.10 25
H10.1 dan Kab. Bdg	0.10 25	0.071 25	0.10 25	0.10 25	0.10 25	0.10 25

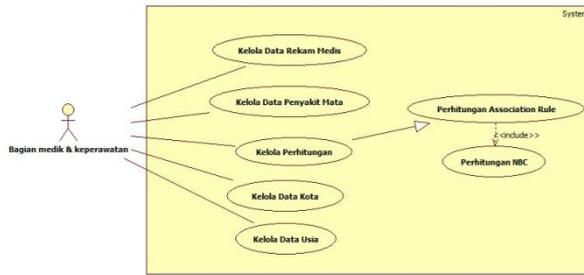
Kemudian setelah melakukan langkah kedua, langkah selanjutnya menghitung nilai $P(A_i | V_j) * P(V_j)$ untuk setiap v . langkah ini berfungsi untuk menentukan nilai tertinggi dari masing-masing class yang tujuannya untuk mengelompokkan setiap penyakit mata dari berbagai kota dan kabupaten. Tabel nilai tertinggi dapat dilihat pada Tabel 7.

Tabel 7. Nilai tertinggi

Klasifikasi	Hasil
H10.1 dan Bdg	0.0000000881
H10.1 dan Cmh	0.0000000613
H10.1 dan Kab. Idr	0.0000000613
H10.1 dan Kab. Bdg	0.0000000613

2.1. Perancangan Sistem

Perancangan sistem ini dibuat menggunakan *software star UML*. Pertama membuat *use case* yang terdiri dari empat *use case* utama dan satu actor. Gambar *use case diagram* dapat dilihat pada Gambar 2.



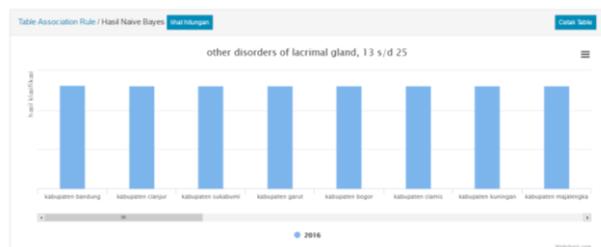
Gambar 2. Use case diagram

2.2. Hasil

Hasil dari penelitian ini telah diujikan dengan menggunakan data rekam medis serta menggunakan metode *Association Rule* dan metode *naïve bayes classifier*. Hasil dari penelitian ini dapat dilihat pada Gambar 3 hasil perhitungan *association rule* dan Gambar 4 grafik perhitungan *naïve bayes classifier*.

No	Kota	Usia	Penyakit Mata	Hasil
1	kota Bandung	40&1100	glaukoma	0.020690527811005
2	kota Bandung	40&1100	other disorders of lacrimal gland	0.0113221107291905
3	kota Bandung	40&1100	prebiyosia	0.0392404081268976
4	kabupaten Bandung	40&1100	glaukoma	0.0088124401913876
5	kota Bandung	13&1025	astigmatism	0.0079623904362025
6	kota Bandung	0&1112	astigmatism	0.0049802057509624
7	kabupaten Bandung	40&1100	prebiyosia	0.00475912402050096
8	kabupaten Bandung	40&1100	other disorders of lacrimal gland	0.0045023544502394
9	kota Bandung	13&1025	myopia	0.003190220430107763
10	kota Bandung	40&1100	stigmatis	0.0030094063293208

Gambar 3. Hasil perhitungan *association rule*



Gambar 4. Grafik perhitungan *naïve bayes classifier*

3. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan dapat disimpulkan bahwa :

1. *Association Rule* dapat digunakan untuk melakukan asosiasi atau mencari kombinasi item untuk kegiatan kesehatan.
2. *Naïve Bayes Classifier* dapat digunakan untuk mengelompokkan penyakit mata dan kota maupun kabupaten di Jawa Barat.
3. Perancangan dibuat berdasarkan gambaran umum sistem.
4. Sistem ini dapat membantu rumah sakit untuk pemilihan tempat dalam kegiatan bantuan kesehatan.

Daftar Pustaka

- [1] Edy Kurniawan, "Analisa Tujuan Berobat Pasien Berdasarkan Jenis Penyakit (ICD) Menggunakan Teknik Data Mining Association Rule Dengan Algoritma Clustering," *Multilek Indonesia*, vol. 7, p. 1, Juni 2013.
- [2] Wahyudi Setiawan, "Sistem Pakar Diagnosa Penyakit Mata Menggunakan Naïve Bayes Classifier," *Seminar Nasional Sains dan Teknologi*, 2014.
- [3] Adhi Bayu Tama, "Penetapan Strategi Penjualan Menggunakan Association Rule Dalam Konteks CRM," *Jurnal Generic*, 2010.
- [4] Nathaniel Chandra Gede Indrawan, I Nyoman Sukajaya Denny, "Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naïve Bayes Classifier Dengan Fitur N-Gram," *Jurnal Ilmiah Teknologi dan Informasi ASIA (JITIKA)*, 2016.

Biodata Penulis

Zillan Taufiq Budiman, Mahasiswa Jurusan Informatika Fakultas MIPA, Universitas Jenderal Achmad Yani. Angkatan 2011. Saat ini sedang menyelesaikan skripsi strata satu (S1).

Wina Witanti, memperoleh sarjana teknik (S.T), Jurusan Teknik Informatika, STT YBS Internasional Bandung, lulus tahun 1998, Program Pasca Sarjana Magister Teknik (M.T), Institut Teknologi Bandung lulus tahun 2006. Saat ini menjadi dosen tetap di Universitas Jenderal Achmad Yani Cimahi.

Dian Nursantika, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Ilmu Komputer Universitas Gajah Mada Yogyakarta, lulus tahun 2008. Memperoleh gelar Magister Komputer (M.Cs) Program Pasca Sarjana Magister Ilmu Komputer Universitas Gajah Mada Yogyakarta, lulus tahun 2010. Saat ini menjadi Dosen di Universitas Jenderal Achmad Yani Cimahi.

