

Analisis Algoritma Klasifikasi C 4.5 Untuk Memprediksi Keberhasilan Immunotherapy Pada Penyakit Kulit

Ady Hermawan¹; Ardi Ramadhan Sukma²; Riqardi Halfis³

¹Magister Ilmu Komputer/STMIK Nusa Mandiri Jakarta
ady.hermawan31@gmail.com

²Magister Ilmu Komputer/STMIK Nusa Mandiri Jakarta
ardisukma40@gmail.com

³Magister Ilmu Komputer/STMIK Nusa Mandiri Jakarta
arizhboys@gmail.com

Abstract - Maintaining skin health is one thing that is also needed. Not only health from inside, health from the outside must also be considered. There are so many skin problems that arise in the human body. Wart disease is characterized by small bumps on the surface of the skin which are generally caused by the Human Papiloma Virus (HPV) virus. One technique for treating wart disease is immunotherapy, this method is a treatment by increasing the immune system to deal with wart disease. Clinical predictions are growing very rapidly by adopting computer science and information technology in managing health and drug data, this clinical prediction can be produced from processing using data mining methods. Data mining is a popular method used to explore patterns or knowledge from large data stacks. C 4.5 algorithm which is one of the decision tree induction algorithms is also a method of data mining algorithms used to classify. This study aims to predict the success rate of immunotherapy treatment methods on wart disease with algorithm C 4.5 using RapidMiner. From the study it was known that the accuracy rate for processing immunotherapy data on wart disease to predict its success using the C 4.5 algorithm of 74.07%.

Keywords: data mining, C 4.5, immunotherapy

PENDAHULUAN

Menjaga kesehatan kulit merupakan salah satu hal yang juga diperlukan. Tidak hanya kesehatan dari dalam saja, kesehatan dari luar pun juga harus diperhatikan. Banyak sekali masalah-masalah kulit yang timbul pada tubuh manusia. Diantaranya adalah penyakit kutil. Kutil dapat tumbuh pada permukaan kulit manapun. Kondisi tersebut merupakan hal yang umum. Namun jika dibiarkan tanpa perawatan tentunya dapat memperburuk kondisi kutil seperti nyeri, bengkak, merak atau bahkan bernanah.

Penyakit kutil ditandai dengan benjolan kecil pada permukaan kulit yang pada umumnya disebabkan oleh virus *Human Papiloma Virus* (HPV). Virus tersebut menginfeksi lapisan atas kulit. Seseorang dapat dengan mudah terinfeksi apabila orang tersebut mempunyai luka terbuka serta menyentuh penderita kutil. Namun tidak semua orang yang terkena kontak langsung dengan pengidap virus HPV akan menimbulkan penyakit kutil. Hal tersebut bergantung pada tingkat imunitas individu yang bersangkutan.

Salah satu teknik pengobatan penyakit kutil adalah imunoterapi, metode ini merupakan

pengobatan dengan meningkatkan sistem kekebalan tubuh untuk mengatasi penyakit kutil. Imunoterapi intraleksi adalah pengobatan efektif untuk kutil. Metode ini memiliki respons terapeutik yang lebih baik, membutuhkan lebih sedikit sesi, dan mampu mengobati kutil yang sudah lama (Supriyatna, 2018).

Ada tren baru terhadap penggunaan imunoterapi dalam pengobatan kutil, karena sistem kekebalan tubuh tampaknya memainkan peran penting dalam pengendalian infeksi kutil (Shah, 2016).

Prediksi klinis berkembang sangat pesat dengan mengadopsi ilmu computer dan teknologi informasi dalam pengelolaan data kesehatan dan obat, prediksi klinik ini dapat dihasilkan dari pengolahan dengan menggunakan metode data mining. Data mining merupakan metode yang digunakan untuk menggali ilmu pengetahuan dari kumpulan data. Kumpulan atau tumpukan data yang hanya tersimpan diolah dengan data mining untuk menghasilkan ilmu pengetahuan yang dapat dimanfaatkan untuk kebutuhan prediksi klinis (Amalia, 2018).

Penelitian terkait digunakan sebagai acuan dan referensi yang dapat memudahkan peneliti dalam menentukan langkah sistematis dari berbagai teori

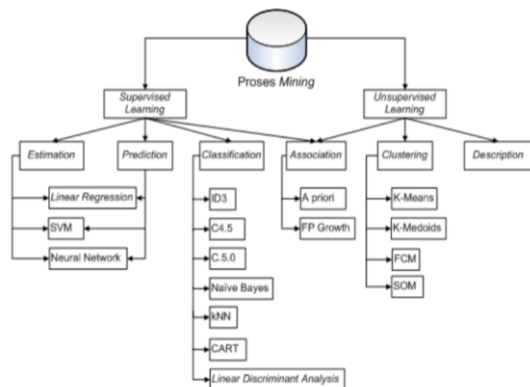
konseptual yang akan dikaji. Penelitian terkait berfungsi untuk menghindari duplikasi dan penanggulangan penelitian atau kesalahan yang sama seperti yang dibuat oleh peneliti sebelumnya (Masyhuri, 2008). Berikut penelitian yang juga pernah membahas tentang prediksi keberhasilan *immunotherapy* pada penyakit kutil:

1. Penelitian Adi Supriatna dan Wida Prima Mustika

Penelitian yang berjudul “Komparasi Algoritma Naive bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kutil” ini menganalisa keberhasilan imunoterapi pada penyakit kutil dengan membandingkan algoritma Naive bayes dan SVM. Dimana dari hasil pengujian menggunakan algoritma Naives bayes dapat memprediksi *class* untuk semua *instance* secara tepat dengan nilai akurasi 1, sedangkan metode *Support Vector Machine* (SVM) masih terdapat prediksi yang tidak tepat dengan tingkat akurasi sebesar 0.8.

Data Mining

Data mining adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis (Widayu, 2017). Salah satu teknik yang dibuat dalam data mining adalah bagaimana menelusuri data yang ada untuk membangun sebuah model, kemudian menggunakan model tersebut agar dapat mengenali pola data yang lain yang tidak berada dalam basis data yang tersimpan. Kebutuhan untuk prediksi juga dapat memanfaatkan teknik ini (Haryati, 2015). Dalam *data mining* inilah tumpukan data masa lalu dianggap sebagai tambang yang dapat diolah untuk menghasilkan sebuah pengetahuan yang sangat berharga.



Gambar 1. Pengelompokan Teknik Data Mining

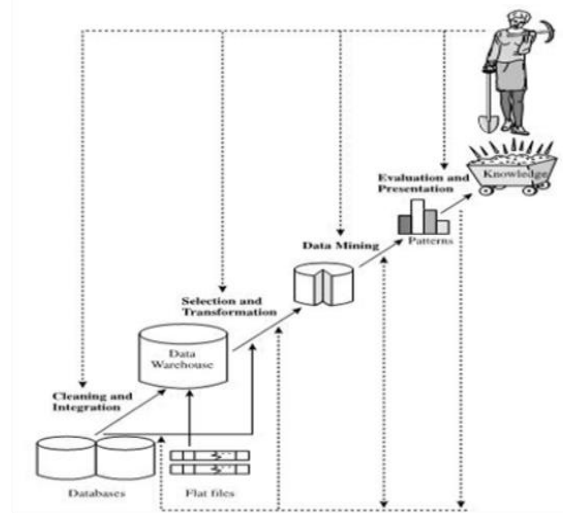
Pada dasarnya, data mining mempunyai empat fungsi dasar, yaitu:

1. Fungsi Prediksi (*prediction*). Proses untuk menemukan pola dari data dengan menggunakan beberapa variabel untuk memprediksikan variabel lain yang tidak diketahui jenis atau nilainya.
2. Fungsi Deskripsi (*description*). Proses untuk menemukan suatu karakteristik penting dari data dalam suatu basis data.
3. Fungsi Klasifikasi (*classification*). Klasifikasi merupakan suatu proses untuk menemukan model atau fungsi untuk menggambarkan *class* atau konsep dari suatu data. Proses yang digunakan untuk mendeskripsikan data yang penting serta dapat meramalkan kecenderungan data pada masa depan.
4. Fungsi Asosiasi (*association*). Proses ini digunakan untuk menemukan suatu hubungan yang terdapat pada nilai atribut dari sekumpulan data.

Algoritma C 4.5

Algoritma yang dapat digunakan untuk memprediksi atau mengklarifikasi suatu kejadian dengan pembentukan pohon keputusan antara lain C4.5, yang merupakan salah satu algoritma induksi pohon keputusan yang dikembangkan oleh J. Ross Quinlan. Secara umum alur proses algoritma C4.5 untuk membangun pohon keputusan dalam data mining adalah:

1. Pilih atribut sebagai simpul akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.



Gambar 2. Tahapan Dalam Data Mining

Pemilihan atribut sebagai simpul, baik simpul akar(*root*) atau simpul internal didasarkan pada nilai *Gain* tertinggi dari atribut-atribut yang ada. Penghitungannilai *Gain* digunakan rumus seperti dalam Persamaan 1 (Yunus, 2014). Perhitungan metode algoritma C4.5 menggunakan konsep

information gain atau entropy reduction untuk memilih pembagian yang optimal. Tahapan dalam membuat sebuah pohon keputusan dengan algoritma C4.5 yaitu (Amalia, 2017):

1. Mempersiapkan data training, dapat diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu
2. Menentukan akar dari pohon dengan menghitung nilai gain yang tertinggi dari masing-masing atribut atau berdasarkan nilai indeks entropi terendah. Sebelumnya dihitung terlebih dahulu nilai indeks entropi, dengan persamaan(1).

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots (1)$$

- S : Himpunan kasus
- A : Atribut
- n : Jumlah partisi atribut A
- |Si| : Jumlah kasus pada partisi ke-i
- |S| : Jumlah kasus dalam S

untuk menghitung nilai Entropy dapat dilihat pada Persamaan 2.

$$Entropy(S) = \sum_{i=1}^n - pi * \log_2 pi \dots \dots \dots (2)$$

- n : Jumlah partisi S
- pi : Proporsi dari Si terhadap S

Terdapat empat langkah dalam proses pembuatan pohon keputusan pada algoritma C4.5, yaitu:

1. Memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada.
2. Membuat cabang untuk masing-masing nilai, artinya membuat cabang sesuai dengan jumlah nilai variabel gain tertinggi.
3. Membagi setiap kasus dalam cabang, berdasarkan perhitungan nilai gain tertinggi dan perhitungan dilakukan setelah perhitungan nilai gain tertinggi awal dan kemudian dilakukan proses perhitungan gain tertinggi kembali tanpa meyertakan nilai variabel gain awal.
4. Mengulangi proses dalam setiap cabang sehingga semua kasus dalam cabang

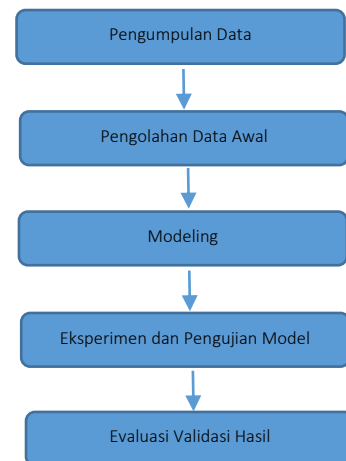
memiliki kelas yang sama, mengulangi semua proses perhitungan gain tertinggi untuk masing-masing cabang kasus sampai tidak bisa lagi dilakukan proses perhitungan.

Algoritma C4.5 secara rekursif mengunjungi setiap simpul keputusan, memilih pembagian yang optimal, sampai tidak bisa dibagi lagi. Dari ketiga peneliti yang telah dilakukan tersebut, klasifikasi dengan Algoritma C4.5 digunakan oleh para peneliti sebagai solusi untuk mengambil keputusan yang diharapkan mampu membantu dalam pengambilan keputusan dengan lebih mudah dan cepat.

Dalam penerapan dan penggunaan algoritma C4.5, dapat digunakan untuk melakukan prediksi dan klasifikasi terhadap calon pegawai yang berpotensi untuk masuk ke dalam perusahaan dengan cara membuat pohon keputusan berdasarkan data-data yang sudah ada dan melakukan prediksi terhadap calon pegawai baru yang ingin masuk ke perusahaan. Selain itu algoritma ini digunakan untuk klasifikasi predikat keberhasilan mahasiswa disebuah universitas. Variabel yang memiliki prioritas utama terhadap predikat keberhasilan mahasiswa adalah mahasiswa yang memilih sesi perkuliahan pada pagi hari.

METODOLOGI PENELITIAN

Metode Penelitian yang digunakan dalam pengujian data keberhasilan *immunotherapy* pada penyakit kutil dengan menggunakan algoritma C 4.5, menggunakan rancangan penelitian sebagai berikut:



1. **Pengumpulan Data**
Pada penelitian ini penulis mendapatkan data dari *repository archive.ics.uci.edu* yang sampai saat penelitian ini dilakukan, data yang berhasil terkumpul adalah sebanyak 91 *record* yang terdiri dari delapan atribut dan satu atribut label.
2. **Pengolahan Data Awal**

Dalam pengolahan data awal penulis melakukan beberapa tahapan:

a. Data Validation

Pada tahap ini penulis melakukan identifikasi dan validasi data yang ada dengan cara mencari tahu apakah ada data yang tidak konsisten atau data yang tidak lengkap (*missing value*). Dalam penelitian ini sebanyak 91 data telah lulus validasi. Sehingga tidak ada data yang perlu dihapus.

b. Data integration and Transformation

Tahap ini digunakan untuk meningkatkan akurasi dan efisiensi algoritma. Dalam penelitian ini beberapa data diubah menjadi data kategorikal yang mampu mewakili data seperti atribut hasil *treatment immunotherapy*.

c. Data size reduction and discretization

Tahap ini digunakan untuk memperoleh dataset dengan jumlah atribut dan record yang lebih sedikit tetapi bersifat informatif. Kemudian data yang digunakan adalah data yang bisa mewakili dataset.

3. Pohon Keputusan (Decision Tree)

Pohon keputusan adalah hasil dari proses perhitungan *entropy* dan *information gain*, setelah perhitungan berulang-ulang sampai semua atribut pohon memiliki kelas dan tidak bisa lagi dilakukan proses perhitungan.

4. Rule Model

Aturan-aturan/ Rule model adalah uraian penjelasan yang merepresentasikan sebuah pohon keputusan.

5. Validasi dan Pengujian

Validasi dan pengujian adalah Pengujian yang dilakukan untuk mengetahui semua fungsi bekerja dengan baik atau tidak. Validasi dilakukan dengan *Ten-fold Cross Validation*. *Ten-fold Cross Validation* adalah validasi yang dilakukan dengan cara membagi suatu set data menjadi sepuluh segmen yang berukuran sama besar dengan cara melakukan pengacakan data. Validasi dan pengujian dilakukan untuk mengetahui tingkat akurasi, presisi, dan recall dari hasil prediksi klasifikasi. Akurasi adalah persentase dari catatan yang diklasifikasikan dengan benar dalam pengujian dataset. Presisi adalah persentase data yang diklasifikasikan sebagai model baik yang sebenarnya juga baik. Recall adalah pengukuran tingkat pengenalan positif sebenarnya.

HASIL DAN PEMBAHASAN

Dalam penelitian ini metode yang digunakan adalah metode algoritma C4.5, dan tool yang digunakan untuk pengolahan data yaitu RapidMiner.

Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data hasil pengobatan penyakit kutil menggunakan *immunotherapy* yang ditunjukkan pada Tabel 1.

Tabel 1. Data Training Immunotherapy

No	Gender	Usia	Lama	Jumlah Kutil	Jenis	Area	Besar (mm)	Hasil
1	Laki-Laki	22	1.25	3	3	47	3	Berhasil
2	Perempuan	19	2.25	2	1	60	7	Berhasil
3	Perempuan	26	10.5	6	1	50	9	Gagal
4	Laki-Laki	25	5.75	2	1	300	7	Berhasil
5	Perempuan	17	11.25	4	3	70	7	Berhasil
6	Laki-Laki	27	5	2	1	20	5	Berhasil
7	Perempuan	24	4.75	10	3	30	45	Berhasil
8	Laki-Laki	15	11	6	1	30	25	Gagal
9	Perempuan	34	11.5	12	1	25	50	Gagal
10	Perempuan	20	7.75	18	3	45	2	Berhasil
11	Perempuan	38	2.5	1	3	43	50	Berhasil
12	Laki-Laki	23	3	2	3	87	70	Berhasil
91	Perempuan	23	6.75	6	1	19	2	Berhasil

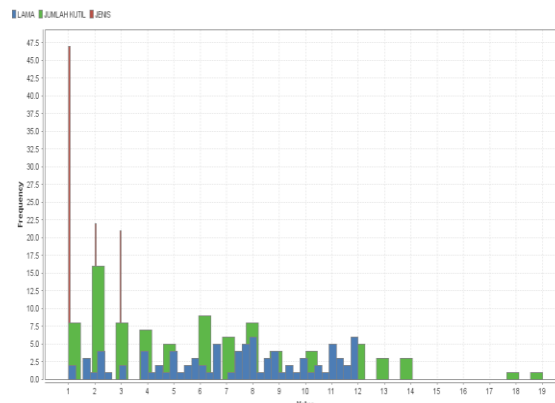
Validasi dan Pengujian

Pengujian dilakukan dengan validasi silang. Salah satu jenis validasi silang adalah *cross validation*. Berikut hasil dari *cross validation*:

Tabel 3. Hasil Cross Validation

Accuracy: 74.07 %			
	True Berhasil	True Gagal	Class Precision
Pred. Berhasil	20	6	76.92%
Pred. Gagal	1	0	0.00%

Berdasarkan hasil pengujian dengan menggunakan *Cross-Validation* maka di dapatkan nilai akurasi 74.07% dengan class precesion Berhasil 76.92% dan class precision Gagal 0.00%. Selain menghasilkan nilai akurasi dari pengolahan RapidMiner juga menghasilkan kurva ROC seperti pada Gambar 3.

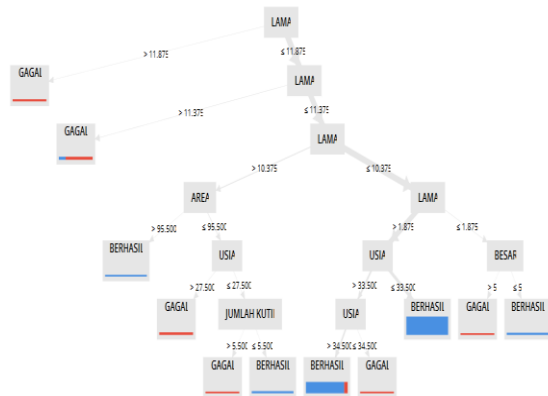


Gambar 3. Kurva ROC Algoritma C4.5.

Sumber: hasil pengolahan data Rapidminer

Pohon Keputusan (Decision Tree)

Dari hasil pengujian data keberhasilan *immunotherapy* pada penyakit kutil dengan algoritma C 4.5 menggunakan Rapidminer didapat pohon keputusan (decision tree) seperti yang terlihat pada gambar 4.



Gambar 4. Model Pohon Keputusan Algoritma C4.5
Sumber: hasil pengolahan data Rapidminer

Rule Model

Dari pohon keputusan (decision tree) yang terbentuk pada Gambar 4 diatas didapat aturan-aturan/*rule*. Ada beberapa aturan yang terbentuk, dapat dilihat sebagai berikut:

1. *If* Lama > 11,875 *then* Hasil = Gagal
2. *If* Lama > 11,375 *then* Hasil = Gagal
3. *If* Lama > 10,375 *and* Area > 95,500 *then* Hasil = Berhasil
4. *If* Lama > 10,375 *and* Area > 95,500, Usia > 27,500 *then* Hasil = Gagal
5. *If* Lama > 10,375 *and* Area > 95,500, Usia > 27,500, Jumlah Kutil > 5,500 *then* Hasil = Gagal
6. *If* Lama > 10,375 *and* Area > 95,500, Usia > 27,500, Jumlah Kutil ≤ 5,500 *then* Hasil = Berhasil
7. *If* Lama ≤ 1,875 *and* Besar > 5 *then* Hasil = Gagal
8. *If* Lama ≤ 1,875 *and* Besar ≤ 5 *then* Hasil = Berhasil
9. *If* Lama > 1,875 *and* Usia ≤ 33,500, *then* Hasil = Berhasil
10. *If* Lama > 1,875 *and* Usia ≤ 34,500, *then* Hasil = Gagal
11. *If* Lama > 1,875 *and* Usia > 34,500, *then* Hasil = Berhasil

KESIMPULAN

Berdasarkan hasil pengujian data keberhasilan *immunotherapy* pada penyakit kutil dengan menggunakan algoritma C 4.5, maka dapat disimpulkan bahwa:

1. Tingkat akurasi algoritma tersebut mencapai 74.07%. Dimana artinya pengujian data keberhasilan *immunotherapy* dengan menggunakan algoritma C 4.5 tidak lebih baik dibandingkan Naïve bayes seperti penelitian yang telah dilakukan sebelumnya dengan data yang sama. Untuk penelitian berikutnya diperlukan penelitian lebih lanjut mengenai atribut yang mempengaruhi pengolahan data serta menggunakan lebih banyak data guna meningkatkan kualitas hasil penelitian.
2. Terbentuk beberapa aturan atau rule model yang dapat digunakan sebagai acuan dalam membuat aplikasi GUI.

REFERENSI

- Amalia, H & Evicienna. (2017). Komparasi Metode Data Mining Untuk Penentuan Proses Persalinan Ibu Melahirkan. *Jurnal Sistem Informasi (Journal of Information Systems)*. 2/13 (2017), 103-109. Retrieved from <http://jsi.cs.ui.ac.id/index.php/jsi/article/view/545>
- Yunus, M., Dahlan, H S., Santoso, P B. (2014). SPK Pemilihan Calon Pendoron Darah Potensial dengan Algoritma C4.5 dan Fuzzy Tahani. *Jurnal EECCIS Vol. 8, No. 1*. Retrieved from <http://jurnaleeccis.ub.ac.id/index.php/eccis/article/view/235>
- Haryati, S., Sudarsono, A., Suryana, E. (2015). Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu). *Jurnal Media Infotama Vol. 11 No. 2*. Retrieved from <https://jurnal.unived.ac.id/index.php/jmi/article/view/260>
- Supriyatna, A., Mustika, W P. (2018). Komparasi Algoritma Naive bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kutil. *Jurnal Sains Komputer Informatika (J-SAKTI) Volume (2) No.2*. Retrieved from https://www.researchgate.net/publication/327870768_Komparasi_Algoritma_Naive_bayes

[dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kutil](#)

Shah, A N., Patel, D., Ravishankar, V. (2016). Measles, mumps and rubella vaccine as an intralesional immunotherapy in treatment of warts. International Journal of Research in Medical Sciences Shah AN et al. Int J Res Med Sci. 2016 Feb;4(2):472-476. Retrieved from <https://www.msjonline.org/index.php/ijrms/article/view/522>

Masyhuri, M. Z. (2008). *Metodologi Penelitian Pendekatan Praktis dan Aplikatif*. Bandung: PT. Refika Aditama.

PROFIL PENULIS

Ady Hermawan, lahir di Kebumen, 31

Desember 1994, meraih gelar Sarjana Komputer (S.Kom) di STMIK Nusa Mandiri ada 2018, saat ini aktif sebagai *programmer* di sebuah perusahaan swasta yang bergerak di bidang Konsultan TI dan sedang menempuh jenjang pendidikan Magister Ilmu Komputer di STMIK Nusa Mandiri.

Ardi Ramadhan Sukma, S.Kom, lulus dari Program Sarjana(S1) Teknik Informatika di Universitas Pembangunan Nasional “Veteran” Jakarta. Sekarang sedang melanjutkan Program Pasca Sarjana(S2) Magister Ilmu Komputer di STMIK Nusa Mandiri.

Riqadri Halfis, S.Kom, lulus dari Program Sarjana(S1) Sistem Informasi di STMIK Jayabaya. Sekarang sedang melanjutkan Program Pasca Sarjana(S2) Magister Ilmu Komputer di STMIK Nusa Mandiri.