

PENERAPAN ALGORITMA DESICION TREE C4.5 DALAM KLASIFIKASI REKOMENDASI PENDONOR DARAH PADA UNIT TRANSFUSI DARAH JEMBER

¹Esa Dinar Wulansari (1510651006),

²Ilham Saifudin, S.Pd., M.Si.

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

Jl. Karimata No. 49 Jember Kode Pos 68121

Email: ¹esadinarw@gmail.com, ²ilham.saifudin@unmuhjember.ac.id.

ABSTRAK

Donor darah merupakan proses penyaluran darah atau produk berbasis darah dari satu orang ke sistem peredaran darah orang lain, dengan melakukan donor darah maka sel-sel darah di dalam tubuh menjadi lebih cepat terganti dengan yang baru. Penelitian ini bertujuan mengetahui berapa tingkat akurasi dan presisi hasil klasifikasi rekomendasi pendonor darah menggunakan algoritma *c4.5* berdasarkan atribut umur, jumlah donor, ketepatan waktu, dimana ketepatan dilihat dari riwayat pendonor sebelumnya. Setelah dilakukan pengujian data sebanyak 27 kali dengan k-fold 2,3,5,6,7,10 menggunakan teknik validasi *cross validation* diperoleh nilai akurasi dan presisi yang berbeda-beda pada setiap pengujian, bahwa pengujian 2fold iterasi ke 1, 7fold iterasi ke 3 dimana mempunyai nilai akurasi 97% dan presisi 100% maka dapat disimpulkan bahwa metode klasifikasi algoritma *c4.5* dapat digunakan dalam rekomendasi pendonor darah.

Kata kunci: Pendonor Darah, Algoritma *C4.5*, *Confusion Matrix*, *Cross Validation*.

APPLICATION OF DESICION TREE C4.5 ALGORITHM IN THE CLASSIFICATION OF BLOOD DONOR RECOMMENDATIONS ON THE JEMBER BLOOD TRANSFUSION UNIT

¹Esa Dinar Wulansari (1510651006),

²Ilham Saifudin, S.Pd., M.Si.

Informatics Engineering Study Program, Faculty of Engineering, Muhammadiyah University, Jember

Jl. Karimata No. 49 Jember Postal Code 68121

Email: ¹esadinarw@gmail.com, ²ilham.saifudin@unmuhjember.ac.id.

ABSTRACT

Blood donation is the process of channeling blood or blood-based products from one person to another person's circulatory system, by making a blood donor the blood cells in the body become more quickly replaced with new ones. This study aims to determine the level of accuracy and precision of the classification of blood donor recommendations using the *c4.5* algorithm based on the attributes of age, number of donors, timeliness, where accuracy is seen from the history of previous donors. After testing the data 27 times with k-fold 2,3,5,6,7,10 using cross validation validation techniques obtained accuracy and precision values are different in each test, that testing 2fold iteration to 1, 7fold iteration to 3 which has an accuracy value of 97% and 100% precision, it can be concluded that the classification method of *c4.5* algorithm can be used in blood donor recommendations.

Keywords: *Blood Donor, C4.5 Algorithm, Confusion Matrix, Cross Validation.*

PENDAHULUAN

Donor darah proses penyaluran darah atau produk berbasis darah dari satu orang ke sistem peredaran darah orang lain. Donor darah berhubungan dengan kondisi medis seperti kehilangan darah dalam jumlah besar yang disebabkan oleh trauma, operasi, syok, dan tidak berfungsinya organ pembentuk sel darah merah. Banyak sekali orang yang tidak mengetahui manfaat donor darah bagi kesehatan. Dengan melakukan donor darah, maka sel-sel darah di dalam tubuh menjadi lebih cepat terganti dengan yang baru.

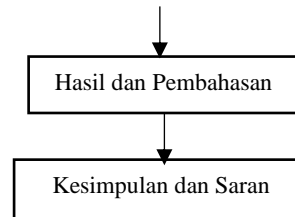
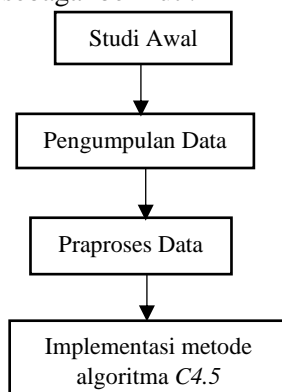
Unit Transfusi Darah merupakan bagian dari PMI yang melayani kegiatan donor darah, persediaan stok darah di unit transfusi darah terkadang tidak tetap. Stok darah yang banyak pun belum bisa menjamin kebutuhan atau permintaan darah dari rumah sakit, karena pasien yang membutuhkan darah terkadang juga lebih banyak begitupun sebaliknya yang membutuhkan darah di rumah sakit sedikit namun persediaan darah di PMI banyak mengakibatkan kantong darah yang tersedia menjadi kadaluarsa. Hal tersebut berpengaruh pada ketepatan jadwal pendonor, data pendonor darah akan di olah untuk meprediksi dan mengetahui pendonor darah yang dapat direkomendasikan di unit transfusi darah dengan menggunakan data mining.

Berdasarkan pada penelitian sebelumnya penulis tertarik menggunakan metode klasifikasi *desicion tree* yaitu algoritma C4.5. Dengan parameter meliputi umur, jumlah donor, ketepatan waktu. Dari hasil perhitungan data akan menghasilkan keputusan dan di uji akurasinya.

METODOLOGI PENELITIAN

a. Kerangka Penelitian

Penjelasan mengenai tahapan penelitian tentang penerapan algoritma C4.5 dalam klasifikasi pendonor darah potensial pada Unit Transfusi Darah Jember. Metodologi penelitian yang dilakukan ini melalui beberapa tahapan yaitu studi awal, pengumpulan data, praproses data, implementasi metode algoritma C4.5, confusion matrix, hasil dan pembahasan, kesimpulan dan saran. Berikut ini adalah diagram metodologi penelitian yang berisi kerangka penelitian yang akan dilakukan sebagai berikut :



Gambar 2 Diagram Metodologi Penelitian

b. Implementasi Algoritma C4.5

Proses implementasi pada data menggunakan algoritma C4.5 awalnya dilakukan proses transformasi atau mengubah data ke bentuk yang sesuai dengan beberapa nilai atribut yang sebelumnya bernilai numerik menjadi karakter yang sesuai dengan data agar dapat di proses dalam perhitungan algoritma C4.5. Berikut langkah-langkah dalam proses transformasi data pendonor pada Unit Transfusi Darah Jember :

1. Menghitung nilai entropy. Berikut rumus *entropy* pada Persamaan 1:

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

Keterangan:

S : himpunan kasus.

n : jumlah partisi S

P_i : proporsi S_i terhadap S

2. Kemudian hitung nilai *gain* menggunakan rumus seperti dalam Persamaan 2:

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

Keterangan :

S : himpunan kasus

A : atribut

$|S_i|$: jumlah kasus pada partisi ke- i

$|S|$: jumlah kasus dalam S

$Entropy(S_i)$: *entropy* untuk sampel-sampel yang memiliki nilai i .

3. Ulangi perhitungan entropy dan gain pada atribut-atribut yang belum terpartisi
4. Proses partisi pohon keputusan akan berhenti saat
 - 1) Semua record dalam simpul N mendapat kelas yang sama.
 - 2) Tidak ada atribut di dalam *record* yang dipartisi lagi.
 - 3) Tidak ada *record* di cabang yang kosong.
5. Pohon Keputusan.
6. *Cross Validation*
7. *Confusion Matrix*

HASIL DAN PEMBAHASAN

a. Implementasi Sistem

Tahap implementasi sistem merupakan suatu tahapan yang menerapkan semua desain sistem yang telah dibuat kedalam bentuk interface sehingga suatu software dapat digunakan dengan mudah. Maka dilakukan implementasi sistem “Penerapan Algoritma *Decision Tree C4.5* Dalam Klasifikasi Pendoron Darah Potensial Pada Unit Transfusi Darah Jember”. Sistem ini dibangun menggunakan bahasa pemrograman PHP dan Database MySQL.

b. Pengujian Data

Pada penelitian ini data yang sudah dikumpulkan akan dilakukan preprocessing data dimana tahap ini memilah atribut yang akan dipakai untuk pengujian, atribut yang dipakai antara lain umur, jumlah donor, ketepatan waktu. Data yang akan digunakan sebagai pengujian pada tugas akhir ini sebanyak 210 data pendonor. Sebelumnya data akan dibagi menjadi 5 kategori pengujian data yaitu pembagian pengujian dengan data latih dan data uji yang berbeda-beda

1) Hasil Uji Skenario 1 (2fold)

Berdasarkan pada skenario 1 dengan pengujian k=2 yaitu membagi data uji pada nomor 1-105 dan sisa datanya menjadi data latih. Untuk menghasilkan nilai presentase akurasi dan presisi pada data uji, sebelumnya akan dilakukan perhitungan data latih pengujian k=2. Berikut merupakan hasil perhitungan *entropy* dan *information gain* pada pengujian k=2 dengan menghitung data nomor 1-105 :

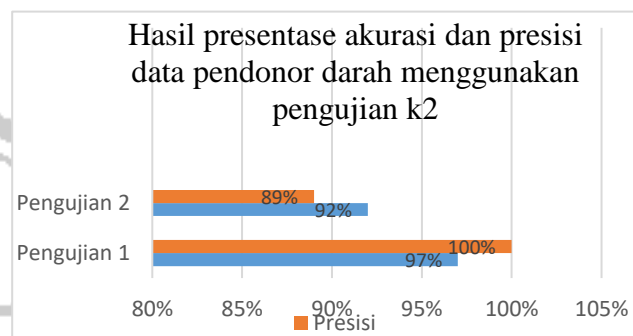
ATRIBUT		JUMLAH KASUS	Ya	Tidak	Entropy	Information Gain
		(S)	(Si)	(Si)		Σ
Total		105	68	37	0,93618	
Umur					0,93089	0,00529
	Dewasa	68	42	26	0,95969	
	Tua	37	26	11	0,87796	
Jumlah Donor					0,17076	0,76542
	Sedikit	34	0	34	0	
	Banyak	71	68	3	0,25253	
Ketepatan					0,35138	0,58480
	Tepat Waktu	76	68	8	0,48546	
	Terlambat	29	0	29	0	

Tabel 1 Perhitungan C4.5

Setelah diperoleh hasil nilai *entropy* dan *information gain* dari tabel perhitungan pengujian k=2 pada iterasi 1 diatas disimpulkan bahwa atribut jumlah donor memiliki nilai gain tertinggi. Selanjutnya membuat cabang untuk tiap-tiap nilai dan ulangi kembali proses setiap cabang sampai mendapatkan nilai. Setelah nilai *information gain* di dapat akan membentuk pohon keputusan seperti penyelesaian

perhitungan algoritma *c4.5* pada bab 3 dan hasil perhitungan proses pengujian k=2 iterasi 2 sama seperti k=1. Sama halnya untuk pengujian beberapa k seperti k=3, k=5, k=16, k=7, dan k=10 untuk proses perhitungan sama seperti pada uji skenario k=2 atau 2fold

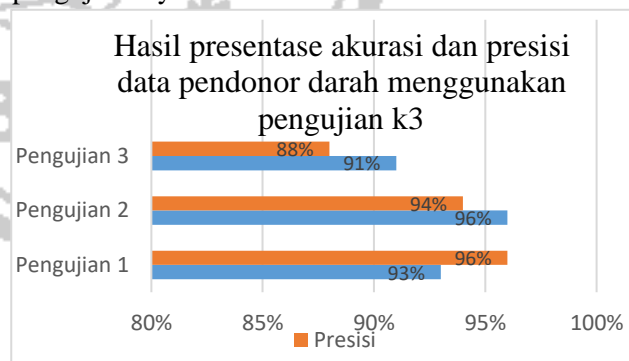
Diperoleh hasil presentasinya dimana pengujian 1 nilai akurasi 97% presisi 100%, pengujian 2 nilai akurasi 92% presisi 89%. Maka diambil tingkat akurasi yang tertinggi dari uji skenario 1, pada pengujian 1 yaitu 97%.



Gambar 1 Diagram Batang Hasil 2fold

2) Hasil Uji Skenario 2 (3fold)

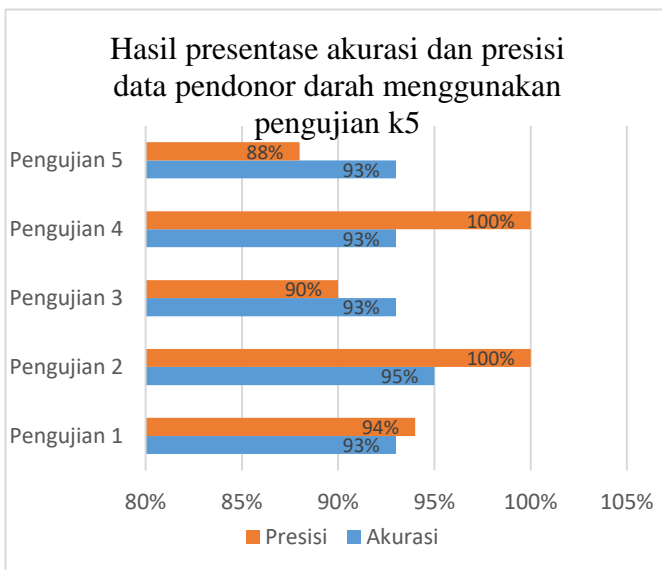
Hasil uji skenario 2 yang telah dilakukan dengan 3fold, menggunakan 3 pengujian dengan data pendonor sebanyak 210 data diperoleh hasil akurasi dan presisi yang berbeda pada masing-masing pengujian, berikut hasil presentasinya dimana pengujian 1 nilai akurasi 93% presisi 96%, pengujian 2 nilai akurasi 96% presisi 94%, pengujian 3 nilai akurasi 91% presisi 88%. Maka diambil tingkat akurasi yang tertinggi dari uji skenario 2, pada pengujian 2 yaitu 96%.



Gambar 2 Diagram Batang Hasil 3fold

3) Hasil Uji Skenario 3 (5fold)

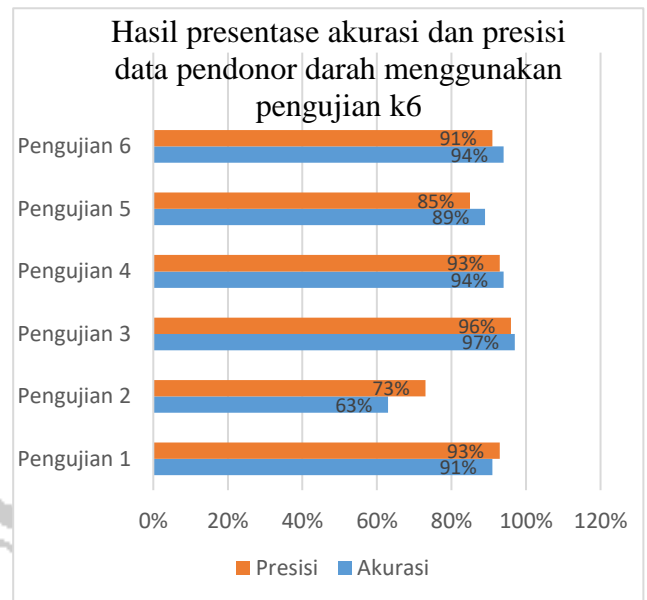
Hasil uji skenario 3 yang telah dilakukan dengan 5fold, menggunakan 5 pengujian dengan data pendonor sebanyak 210 data diperoleh hasil akurasi dan presisi yang berbeda pada masing-masing pengujian, berikut hasil presentasinya dimana pengujian 1 nilai akurasi 93% presisi 94%, pengujian 2 nilai akurasi 95% presisi 100%, pengujian 3 nilai akurasi 93% presisi 90%, pengujian 4 nilai akurasi 93% presisi 100%, pengujian 5 nilai akurasi 93% presisi 88%. Maka diambil tingkat akurasi yang tertinggi dari uji skenario 3, pada pengujian 2 yaitu 95%.



Gambar 3 Diagram Batang Hasil 5fold

4) Hasil Uji Skenario 4 (6fold)

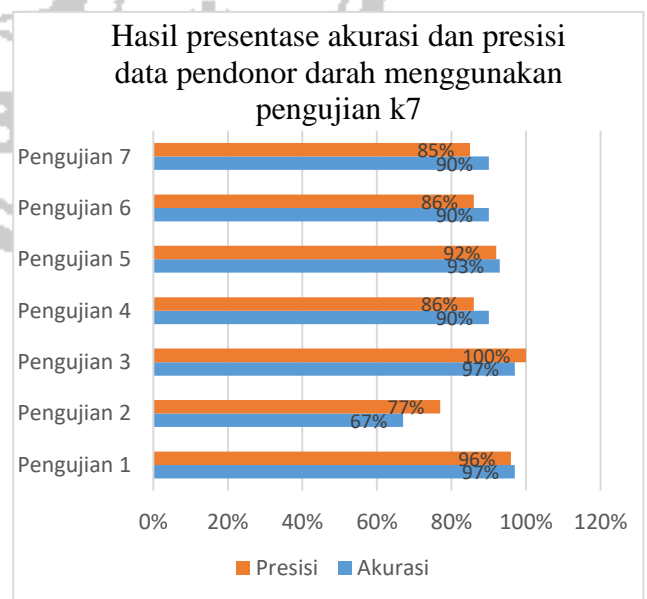
Hasil uji skenario 4 yang telah dilakukan dengan 6fold, menggunakan 6 pengujian dengan data pendonor sebanyak 210 data diperoleh hasil akurasi dan presisi yang berbeda pada masing-masing pengujian, berikut hasil presentasinya dimana pengujian 1 nilai akurasi 91% presisi 93%, pengujian 2 nilai akurasi 63% presisi 73%, pengujian 3 nilai akurasi 97% presisi 96%, pengujian 4 nilai akurasi 94% presisi 93%, pengujian 5 nilai akurasi 89% presisi 85%, pengujian 6 nilai akurasi 94% presisi 91%. Maka diambil tingkat akurasi yang tertinggi dari uji skenario 4, pada pengujian 3 yaitu 97%.



Gambar 4 Diagram Batang Hasil 6fold

5) Hasil Uji Skenario 6 (7fold)

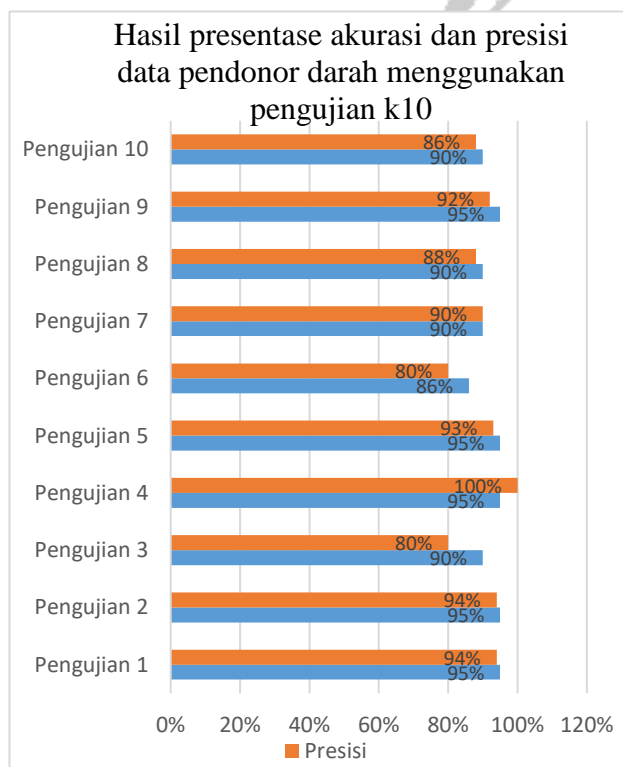
Hasil uji skenario 5 yang telah dilakukan dengan 7fold, menggunakan 7 pengujian dengan data pendonor sebanyak 210 data diperoleh hasil akurasi dan presisi yang berbeda pada masing-masing pengujian, berikut hasil presentasinya dimana pengujian 1 nilai akurasi 97% presisi 96%, pengujian 2 nilai akurasi 67% presisi 77%, pengujian 3 nilai akurasi 97% presisi 100%, pengujian 4 nilai akurasi 90% presisi 86%, pengujian 5 nilai akurasi 93% presisi 92%, pengujian 6 nilai akurasi 90% presisi 86%, pengujian 7 nilai akurasi 90% presisi 85%. Maka diambil tingkat akurasi yang tertinggi dari uji skenario 5, pada pengujian 3 yaitu 97%.



Gambar 5 Diagram Batang Hasil 7fold

6) Hasil Uji Skenario 7 (10fold)

Hasil uji skenario 6 yang telah dilakukan dengan 10fold, menggunakan 10 pengujian dengan data pendonor sebanyak 210 data diperoleh hasil akurasi dan presisi yang berbeda pada masing-masing pengujian, berikut hasil presentasinya dimana pengujian 1 nilai akurasi 95% presisi 94%, pengujian 2 nilai akurasi 95% presisi 94%, pengujian 3 nilai akurasi 90% presisi 80%, pengujian 4 nilai akurasi 95% presisi 100%, pengujian 5 nilai akurasi 95% presisi 93%, pengujian 6 nilai akurasi 86% presisi 80%, pengujian 7 nilai akurasi 90% presisi 90%, pengujian 8 nilai akurasi 90% presisi 88%, pengujian 9 nilai akurasi 95% presisi 92%, pengujian 10 nilai akurasi 90% presisi 86%. Maka diambil tingkat akurasi yang tertinggi dari uji skenario 5, pada pengujian 3 dan 4 yaitu akurasi 95% presisi 100%.



Gambar 6 Diagram Batang Hasil 10fold

c. Pembahasan

Pengujian data yang sudah dilakukan dengan menghitung jumlah keseluruhan data sebanyak 27kali menggunakan teknik *cross validation* yang digunakan untuk mengevaluasi kinerja model atau algoritma dimana data dipisahkan menjadi dua subset yaitu latih dan uji. Melihat hasil jumlah k rata-rata memperoleh nilai yang berbeda pada tiap pengujian nilai k, namun juga terdapat mayoritas hasil yang sama sehingga diambil nilai akurasi dan presisi yang tertinggi, adapun setiap k-fold terdapat nilai yang tertinggi, maka hasil dari uji skenario sebanyak 6 kali dimana jumlah

keseluruhan data dihitung 27 kali menghasilkan nilai akurasi dan presisi yang tertinggi pada k-fold 2 iterasi 1 dan k-fold 7 iterasi 3 yaitu nilai akurasi 97% dan presisi 100%.

KESIMPULAN DAN SARAN

Berdasarkan uraian dari hasil dan pembahasan pada bab sebelumnya, maka didapatkan beberapa kesimpulan dan saran sebagai berikut:

a. Kesimpulan

Berdasarkan penelitian yang sudah dijelaskan pada hasil dan pembahasan sebelumnya yaitu rekomendasi pendonor darah di Unit Transfusi Darah Jember dengan melakukan beberapa pengujian data menggunakan atribut umur, jumlah donor, dan ketepatan, dimana ketepatan dapat dilihat dari riwayat pendonor sebelumnya. Setelah dilakukan uji skenario 1 sampai 6, dimana jumlah keseluruhan data dihitung sebanyak 27 kali dengan k-fold 2,3,5,6,7,10 menggunakan algoritma *c4.5*. Dalam klasifikasi rekomendasi pendonor darah diperoleh nilai akurasi dan presisi yang akan diambil nilai tertinggi pada pengujian yang sudah dilakukan. Diperoleh beberapa nilai akurasi dan presisi tertinggi yang sama antara lain k-fold 2fold iterasi ke 1, 7fold iterasi ke 3 dimana mempunyai nilai akurasi 97% dan presisi 100% maka dapat disimpulkan bahwa metode klasifikasi algoritma *c4.5* dapat digunakan dalam rekomendasi pendonor darah.

b. Saran

1. Dalam penelitian selanjutnya dapat mengembangkan dan menambah data rekomendasi pendonor yang lebih spesifik agar lebih meningkatkan keakuratan kinerja dari metode yang akan digunakan.
2. Penulis dapat mengembangkan dan menggunakan algoritma klasifikasi yang lain untuk membandingkan dalam mencari algoritma terbaik.
3. Dalam mengambil sebuah referensi untuk atribut terbaru sesuai perkembangan dunia kesehatan terutama tentang donor darah.

DAFTAR PUSTAKA

- Arifin, M. F. (2017). Penerapan Algoritma Klasifikasi C4.5 dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada.
- Astria, F. H. (2016). Sistem Informasi Perpustakaan Online Berbasis Web Astria. *E-journal Teknik Elektro dan Komputer* .
- Dhimas, B. R. (2017). Klasifikasi Calon Pendorong Darah Menggunakan Metode Naive Bayes Classifier (Studi Kasus : Calon Pendorong Darah di Kota Semarang). *Jurnal Gaussian, Volume 6, Nomor 2*, 193-200.
- Eka, P. C. (2018). Metode Decision Tree Algoritma C.45 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji. *Jurnal Riset Sistem Informasi Dan Teknik Informatika (Jurasiik)*.
- Elisa, E. (2017). Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti. *Join*, Volume 2 No. 1 |.
- Fridayanthie, E. W. (2015). Analisa Data Mining Untuk Prediksi Penyakit Hepatitis Dengan Menggunakan Metode Naive Bayes Dan Support Vector Machine. *Jurnal Khatulistiwa Informatika*, Vol. 3, No. 1.
- Haryati, S. S. (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.
- Ika, M. (2015). Klasifikasi Data Mining Dalam Menentukan Pemberian Kredit Bagi Nasabah Koperasi. *Jurnal Ilmiah Teknosains* , Vol. 1 No. 1 .
- Kennedi, T. &. (2013). Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan. *Informasi dan Teknologi Ilmiah (INTI)*, Volume : I, Nomor : 1.
- Palit, R. V. (2015). Rancangan Sistem Informasi Keuangan Gereja Berbasis Web Di Jemaat GMIM Bukit Moria Malalayang. *E-Journal Teknik Elektro dan Komputer*, 4, 7.
- Teguh, P. A. (2017). Partisipasi Masyarakat Dalam Kegiatan Donor Darah Di Palangka Raya. Volume 3 Nomor.
- Udi, B. H. (2018). Tinjauan Kegiatan Donor Darah Terhadap Kesehatan Di PMI Karanganyar, Jawa Tengah Tahun 2018. *Infokes*, Vol 8 No 1.