



KLASIFIKASI DIAGNOSA PERADANGAN KANDUNG KEMIH MENGUNAKAN METODE ALGORITMA NAÏVE BAYES

Muhammad Muharrom¹

¹Prodi Teknologi Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika,
muhammad.muu@bsi.ac.id

Jl. Krmat Raya No.98, Senen. Jakarta Pusat 10450 Telp. (021)23231170 Fax (021) 21236158

Keywords:

Disease, Naïve Bayes,
diagnosis

Inflammation of the bladder is one of the diseases caused by pain when urinating, the pain comes from the urinary tract which is between the parts that need a diagnosis such as patient temperature {35C-42C}, occurrence of nausea {yes, no} back pain { yes, no}, Pushing urine (need to urinate constantly) {yes, no}, urinary pain {yes, no}, Burning of the urethra, itching, swelling of the urethral outlet {yes, no}. To understand the problem better, it is necessary to research in order to prevent the worsening of the disease by classifying patient data to accurate disease outcomes suffered by patients and to be able to predict from existing data. The dataset is obtained from public datasets which can be used for research and the method used is Naïve Bayes, one of the classification algorithms that is quite reliable and is assisted by rapid mining tools to analyze its accuracy. By using the naïve Bayes method in the calculation, the compatibility between manual calculations using rapidminer tools and the resulting accuracy is 90.8% and the ROC calculation results in an AUC value of 0.956 which means excellent classification.

Kata Kunci:

Penyakit, Naïve bayes, diagnosis

Abstrak

Peradangan pada kandung kemih adalah salah satu penyakit yang ditimbulkan dari rasa nyeri saat buang air kecil, rasa nyeri itu berasal dari saluran kemih yang diantara bagian yang perlu diagnosis seperti Suhu pasien {35C-42C}, Terjadinya mual {ya, tidak} Nyeri Pinggang {ya, tidak}, Mendorong urin (kebutuhan untuk buang air kecil terus-menerus) {ya, tidak}, Sakit berkemih {ya, tidak}, Pembakaran uretra, gatal, pembengkakan saluran keluar uretra {ya, tidak} . Untuk memahami masalah dengan lebih baik, Maka perlunya penelitian agar dapat mencegah semakin parahnya penyakit yang diderita dengan cara mengklasifikasikan data pasien untuk mengakurasi hasil penyakit yang di derita oleh pasien dan dapat memprediksi dari data yang ada. Dataset didapatkan dari dataset publik yang mana dapat digunakan untuk penelitian dan metode yang digunakan adalah *naïve bayes* salah satu algoritma klasifikasi yang cukup handal dan dibantu dengan *tools rapid miner* untuk menganalisis keakuratannya. Dengan menggunakan metode *naïve bayes* dalam perhitungan dihasilkan kesesuaian antara perhitungan manual dengan menggunakan *tools rapidminer* dan dihasilkannya keakuratan adalah 90,8% dan perhitungan ROC menghasilkan nilai AUC adalah 0.956 yang berarti *excellent classification*.

Pendahuluan

Pada dasarnya kesehatan adalah salah satu faktor utama yang harus dijaga pada manusia. Apabila tubuh dalam keadaan kurang sehat

maka mudah terserang penyakit. Salah satunya sering terjadi yaitu penyakit akibat infeksi. Penyakit ini biasanya disebabkan oleh jamur, bakteri, jamur maupun virus. Risiko seseorang untuk terkena penyakit infeksi bisa meningkat

apabila ia memiliki kondisi medis tertentu seperti memiliki data tahan tubuh yang sedang melesah. Infeksi saluran kemih adalah salah satu penyakit yang merupakan reaksi inflamasi sel-sel *uroepitelium* yang melapisi saluran kemih yang dapat mengenai laki-laki maupun perempuan semua umur yang dapat beresiko terhadap kesehatan tubuh manusia.

Menurut Sjahrurrachaman "banyaknya infeksi kandung kemih disebabkan oleh bakteri yang mengendap dan hanya sebagian kecil yang disebabkan oleh jamur atau virus. Sehingga pengobatan yang utama pada infeksi kandung kemih adalah antibakteri"[1]. Pemeriksaan terhadap urin salah satu pemegang peranan penting dalam memonitoring penyakit infeksi saluran kemih. Jumlah hasil leukosit yang banyak di dalam urin menandakan adanya suatu infeksi atau radang pada saluran kemih yang didapatkan. Leukosit di dalam urin dapat dilihat dengan beberapa cara, salah satunya adalah sedimen urin *leukosit*. Sedimen urin leukosit dapat memberi informasi penting bagi klinis dalam membantu menegakkan diagnosis dan memantau perjalanan penyakit di saluran kemih.

Pada penelitian ini dilakukan untuk mengetahui perhitungan data diagnosa dugaan peradangan kandung kemih serta membandingkan dengan *tools* aplikasi rapid miner menggunakan algoritma *Naïve Bayes*. Menurut Prasetyo "*Naïve Bayes* merupakan salah satu teknik prediksi klasifikasi yang berbasis probabilistik sederhana yang berdasarkan pada penerapan *teorema Bayes* (aturan Bayes) dengan asumsi independensi (ketidaktergantungan) yang kuat (naif) berdasarkan atribut yang sudah ditentukan. Dengan kata lain, dalam *Naïve Bayes* model yang digunakan adalah "model fitur independen"[2].

Penggunaan data mining klasifikasi dengan algoritma *Naïve Bayes* sebagai pilihan untuk prediksi diagnosa peradangan kandung kemih yang dapat menjadi alternatif pilihan yang tepat. Ruang lingkup pada penelitian ini dengan menggunakan data pasien yang didapatkan dari *University of California Irvine Machine Learning Repository*, kemudian mengevaluasi hasil algoritma tersebut untuk

mengetahui metode klasifikasi dan tingkat keakuratan dengan *tools* rapid miner.

Landasan Teori

A. Data Mining

Data mining adalah proses yang menggunakan statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat. Data Mining didefinisikan sebagai proses penemuan pola dalam data. Berdasarkan tugasnya, data mining dikelompokkan menjadi deskripsi, estimasi, prediksi, klasifikasi, *clustering* dan asosiasi. Proses dalam tahap data mining terdiri dari tiga langkah Utama, yaitu *data Preparation* Pada langkah ini, data dipilih, dibersihkan, dan dilakukan *preprocessed* mengikuti pedoman dan *knowledge* dari ahli domain yang menangkap dan mengintegrasikan data internal dan eksternal ke dalam tinjauan organisasi secara menyeluruh. Penggunaan algoritma data mining dilakukan untuk menggali data yang terintegrasi untuk memudahkan identifikasi informasi bernilai [3]. Namun semakin besar data yang diolah maka semakin besar pula waktu prosesnya [4].

B. Algoritma

Kata algoritma (*algorithm*) kata tersebut berasal dari kata *algorism* yang diambil dari nama seorang penulis buku Arab yang terkenal, yaitu Abu Ja'far Muhammad ibnu Musa Al-Khuwarizmi (al-Khuwarizmi dibaca orang Barat menjadi *algorism*)[5].

Adapun pengertian algoritma adalah kumpulan suatu instruksi atau suatu perintah yang dibuat secara jelas, detail dan sistematis berdasarkan urutan yang logis yang dapat diterima oleh akal untuk penyelesaian suatu masalah sesuai dengan tingkat yang sudah ditentukan. 5 komponen utama dalam algoritma yaitu *finiteness*, *definiteness*, *input*, *output* dan *effectiveness*. Sehingga dalam merancang sebuah algoritma ada tiga komponen yang harus ada yaitu:

1. Komponen masukan (*input*)

Komponen ini biasanya terdiri dari pemilihan variabel, jenis variabel, tipe

variabel, konstanta dan parameter (dalam fungsi).

2. Komponen keluaran (*output*)

Komponen ini merupakan tujuan dari perancangan algoritma dan program. Permasalahan yang diselesaikan dalam algoritma dan program harus ditampilkan dalam komponen keluaran. Karakteristik keluaran yang baik adalah keluaran yang benar menjawab permasalahan dan tampilan (*interface*) yang ramah.

3. Komponen proses (*processing*)

Komponen ini merupakan bagian utama dan terpenting dalam merancang sebuah algoritma. Dalam bagian ini terdapat logika masalah, logika algoritma (sintaksis dan semantik), rumusan, metode (rekursi, perbandingan, penggabungan, pengurangan dan lain-lain).

Adapun jika dilihat dari segi kondisinya, kompoenen algoritma terdiri dari dua kondisi yaitu:

1. *Pre condition* adalah kondisi suatu program ketika proses algoritma siap untuk dijalankan (sebelum dilaksanakan algoritma). Dengan kata lain *pre condition* merupakan kondisi awal dimana proses algoritma akan dijalankan. *Pre condition* dinyatakan dengan mendefinisikan input dari sebuah algoritma berdasarkan pengetahuan dan data yang ada.

2. *Post condition*

Post condition merupakan kondisi setelah suatu algoritma selesai dijalankan. *Post condition* dinyatakan dengan mendefinisikan hasil (*output*) dari suatu algoritma untuk mencapai suatu hasil yang telah ditentukan sebelumnya.

C. *Naïve Bayes*

Menurut Prasetyo "*Bayes* merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema Bayes. Klasifikasi Naive Bayes praktis diterapkan karena merupakan salah satu probabilitas sederhana yang penerapannya didasarkan pada teorema bayes dengan asumsi independensi[6]".

Tahapan dari proses algoritma *Naive Bayes* adalah:

1). Menghitung jumlah kelas / label.

2). Menghitung Jumlah Kasus Per Kelas

3). Kalikan Semua Variable Kelas

4). Bandingkan Hasil Per Kelas

D. Confusion Matrix

Merupakan suatu cara untuk memberikan rincian secara klasifikasi, kelas yang diprediksi akan ditampilkan di bagian atas *matrix* dan kelas yang diobservasi ditampilkan di bagian kiri [7].

Evaluasi model *confussion matrix* menggunakan tabel seperti *matrix* dibawah ini:

Tabel 1. Model *Confusion Matrix*

CLASSIFICATION	PREDICTED CLASS	
	Class= Yes	Class= No
Observed Class		
Class= Yes	True Positive- (TP)	False Negative- (FN)
Class= No	False Positive- (FP)	True Negative- (TN)

Akurasi dapat dihitung dengan menggunakan rumus berikut:

$$Accuracy = (TP+TN)/(TP+TN+FP+FN)$$

TP : Jumlah kasus positif yang diklasifikasikan sebagai positif

FP : Jumlah kasus negatif yang diklasifikasikan sebagai positif

TN : Jumlah kasus negatif yang diklasifikasikan sebagai negatif

FN : Jumlah kasus positif yang diklasifikasikan sebagai negatif.

E. Kurva ROC

Kurva ini banyak digunakan untuk menilai hasil prediksi, kurva ROC adalah teknik untuk memvisualisasikan, mengatur, dan memilih pengklasifikasian berdasarkan kinerja mereka. Kurva ROC adalah peralatan dua dimensi yang digunakan untuk menilai kinerja klasifikasi yang menggunakan dua class keputusan, masing- masing objek dipetakan ke salah satu elemen dari himpunan pasangan, positif atau negatif. Pada kurva ROC, TP rate diplot pada sumbu Y dan FP rate diplot pada sumbu X.

Untuk klasifikasi data mining, nilai AUC dapat dibagi menjadi beberapa kelompok [7].

- a) 0.90-1.00= *Excellent Classification*
- b) 0.80-0.90= *Good Classification*
- c) 0.70-0.80= *Fair Classification*
- d) 0.60-0.70= *Poor Classification*
- e) 0.50-0.60= *Failur*

F. Rapid Miner

Rapid Miner adalah *platform* perangkat lunak data ilmu pengetahuan yang dikembangkan oleh perusahaan dengan nama yang sama, yang menyediakan lingkungan terpadu untuk pembelajaran mesin (*machine learning*). Pembelajaran mendalam (*deep learning*), penambangan teks (*teks mining*) dan analisis prediktif (*predictive analytic*). Aplikasi ini digunakan untuk aplikasi bisnis dan komersial serta untuk penelitian, pendidikan, pelatihan, pembuatan *prototype* dengan cepat, dan pengembangan aplikasi serta mendukung semua langkah proses pembelajaran mesin termasuk persiapan data, visualisasi hasil validasi dan pengoptimalan.[8]

Metode Pengumpulan Data

Berikut adalah Teknik pengumpulan data yang digunakan dalam proses penelitian ini:

Adapun perancangan pada penelitian telah disusun sehingga alur dari proses penelitian ini sesuai dengan yang telah ditentukan. penelitian ini menggunakan kuantitatif yang digunakan untuk meneliti pada populasi atau sample tertentu, pengumpulan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif/statistik dengan tujuan untuk menguji hipotesis yang telah ditentukan.

Penelitian ini dilakukan dengan menjalankan beberapa langkah proses penelitian, yaitu:

1) Pengumpulan Data

Data pasien diperoleh dari *University of California Irvine Machine Learning Data Repository*, *Dataset* ini banyak digunakan oleh peneliti untuk menguji algoritma klasifikasi. *Dataset* ini berisikan 120 *record* dengan 6 atribut. *Dataset* ini merupakan *dataset* yang

dapat digunakan untuk klasifikasi penyakit Peradangan Kandung Kemih [9].

2) Pengolahan Awal Data

Pada tahap ini data yang masih berupa nilai numerik dan juga *continue* ditransformasikan kedalam bentuk kategorikal dan dibuat skala atau interval sehingga menghasilkan range yang lebih kecil sebagai bahan pembelajaran algoritma *Naive Bayes* akan lebih mudah di klasifikasikan dengan menggunakan *Rapid Miner* sehingga memudahkan dalam memproses klasifikasi pada *dataset* yang telah tersedia sebelumnya.

3) Penerapan Algoritma *Naive Bayes* dan *Tools Rapid Miner*

Pada tahap ini memilih dan menerapkan teknik yang tepat untuk mendapatkan hasil yang optimal, pada penelitian ini algoritma yang digunakan adalah algoritma *Naive Bayes* dilakukan dengan data uji. tahap ini dilakukan dengan bantuan *tools Rapid Miner .70*.

4) Analisa Hasil perhitungan

Pada penelitian ini dilakukan analisa terhadap hasil yang diperoleh lewat perhitungan manual dan menggunakan *tools rapid miner* untuk mengetahui kesesuaian antara perhitungan manual dengan menggunakan *tools rapid miner*. Pengukuran dilakukan menggunakan metode *Confusion Matrix* dan Kurva ROC beserta *Area Under the Curve*-nya.

Hasil dan Pembahasan

Pada tahap ini dilakukan eksperimen dan pengujian metode yang digunakan yaitu menghitung dengan menggunakan algoritma yang diusulkan yaitu Algoritma *Naive Bayes*. Adapun atribut yang digunakan adalah: Suhu pasien {35C-42C}, Terjadinya mual {ya, tidak} Nyeri Pinggang {ya, tidak}, Mendorong urin (kebutuhan untuk buang air kecil terus-menerus) {ya, tidak}, Sakit berkemih {ya, tidak} Pembakaran uretra, gatal, pembengkakan saluran keluar uretra {ya, tidak}, Keputusan: Peradangan kandung kemih {ya, tidak}. Adapun dibuat satu data yang digunakan untuk pengujian yang nanti akan disesuaikan hasilnya antara perhitungan manual dan menggunakan *tools Rapidminer* sebagai berikut:

Data testing: '40' 'Suhu pasien, 'No' Terjadinya mual, 'No' sakit pinggang, 'Yes' Mendorong urin (kebutuhan buang air kecil terus menerus), 'Yes' Nyeri berkemih, 'Yes' Terbakarnya uretra, gatal, bengkaknya saluran keluar uretra, Keputusan Peradangan kandung kemih disesuaikan dari prediksi dari data yang ada sebanyak 120 data.

Langkah Pertama: Menghitung jumlah kelas/label: Keputusan: Peradangan kandung kemih: "No" = 61/120 "Yes" = 59/120. Langkah Kedua: Menghitung Jumlah Kasus Per Kelas.

Suhu Pasien (35C-42C): "No" = 5/61 "Yes" = 3/59 Terjadinya mual: "No" = 51/61 "Yes" = 40/59 Nyeri Pinggang: "No" = 10/61 "Yes" = 40/59. Mendorong urin (kebutuhan terus-menerus untuk buang air kecil): "No" = 21/61 "Yes" = 59/59. Rasa sakit berkemih: "No" = 10/61

"Yes" = 49/59. Pembakaran uretra, gatal, bengkaknya saluran keluar uretra: "No" = 21/61 "Yes" = 29/59. Langkah Ketiga: Kalikan Semua Variable Kelas Kelas No:

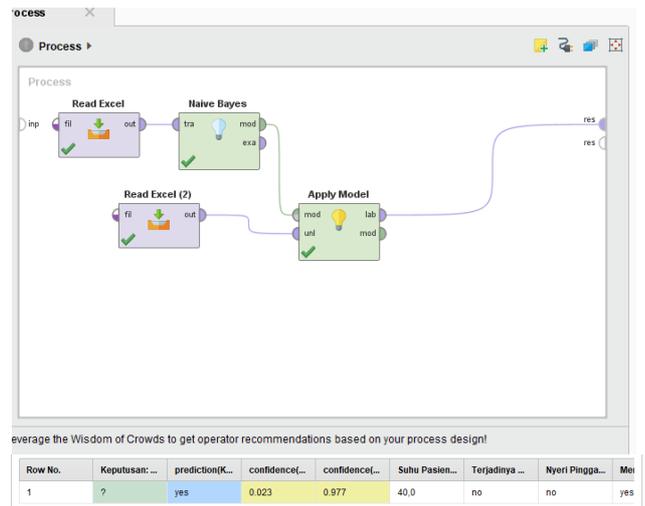
$$= \frac{5}{61} \times \frac{51}{61} \times \frac{10}{61} \times \frac{21}{61} \times \frac{10}{61} \times \frac{21}{61} = 0,0819672131147541 \times 0,836065573770492 \times 0,163934426229508 \times 0,344262295081967 \times 0,163934426229508 \times 0,344262295081967 = 0,0000000000$$

Kelas Yes:

$$= \frac{3}{59} \times \frac{40}{59} \times \frac{40}{59} \times \frac{59}{59} \times \frac{49}{59} \times \frac{29}{59} = 0,0508474576271186 \times 0,677966101694915 \times 0,677966101694915 \times 1 \times 0,830508474576271 \times 0,491525423728814 = 0,0095405905$$

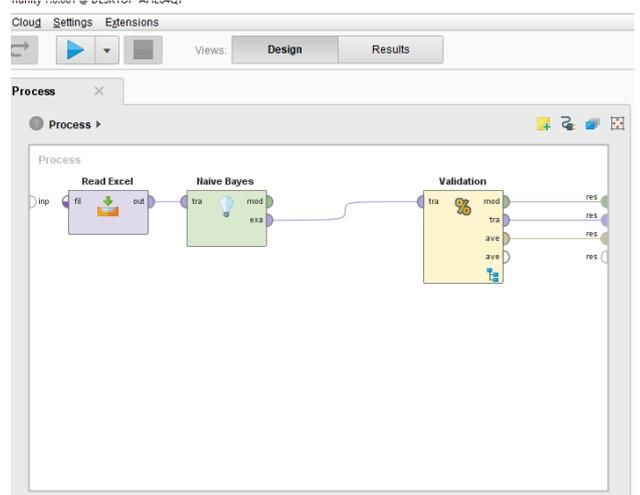
Langkah terakhir adalah membandingkan mana nilai hasil perhitungan yang lebih besar, karena kelas "YES" memiliki nilai lebih besar maka hasil prediksi dari algoritma naïve bayes terhadap data tersebut adalah "Keputusan: Peradangan kandung kemih = "YES".

Adapun menggunakan tools *rapidminer* sehingga dapat diketahui secara otomatis prediksi yang dilakukan sebagai berikut:



Gambar 1. Model klasifikasi naïve bayes Hasil yang didapat prediksi adalah yes, sesuai dengan perhitungan manual.

Adapun untuk model yang digunakan pada aplikasi *rapid miner* adalah



Gambar 2. Model Pengujian Validasi Naïve Bayes

Criterion	accuracy	precision	recall
AUC (optimistic)	accuracy: 90.83% +/- 5.83% (mikro: 90.83%)		
AUC			
AUC (pessimistic)			
	true no	true yes	class preci
pred_no	60	10	85.71%
pred_yes	1	49	98.00%
class recall	98.36%	83.05%	

Gambar 3. Percobaan model confusion matrix

Berdasarkan gambar 3 menunjukkan bahwa, diketahui dari 120 data pasien, hasil data testing pada tabel confusion matrix diatas ada 60 pasien diprediksi negatif maka hasilnya

sesuai dengan prediksi yaitu negatif, 10 pasien diprediksi negatif tetapi hasilnya adalah positif, sedangkan ada 1 pasien yang diprediksi positif tetapi hasilnya adalah negatif dan 49 diprediksi positif maka hasilnya sesuai dengan prediksi yaitu positif tingkat akurasi dengan menggunakan algoritma Naive Bayes adalah 90.83%, dan dapat dihitung untuk mencari nilai *accuracy*, yaitu:

Keterangan:

TP = 49

FP = 1

TN= 60

FN = 10

$$\begin{aligned} \text{Akurasi} &= (TP+TN) / (TP+TN+FP+FN) \\ &= (49+60) / (49+60+1+10) \\ &= 0.9083 \text{ (90.83\%)} \end{aligned}$$

Sedangkan hasil perhitungan divisualisasikan dengan kurva ROC Naive Bayes. Kurva ROC pada gambar 4 mengekspresikan *confusion matrix*. Garis horizontal adalah *false positives* dan garis vertikal *true positives*.



Gambar 4. Kurva ROC

Dari Gambar 4 terdapat grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.956 dimana diagnosa hasilnya *Excellent Classification*.

Kesimpulan dan Saran

Berdasarkan perumusan masalah dan serangkaian penelitian yang telah penulis lakukan. Maka dapat disimpulkan beberapa hal sebagai berikut :

1. Dilakukan proses perhitungan manual menggunakan algoritma naïve bayes dengan data testing menghasilkan prediksi yang sesuai dengan menggunakan *tools rapid miner*.
2. Jika dilakukan proses tingkat akurasi pada *rapidminer* dari *dataset* yang ada menghasilkan nilai keakuratan 90,8% yang

berarti cukup tinggi dilakukan dengan metode *naïve bayes*.

3. Hasil perhitungan ROC menggunakan algoritma naïve bayes menghasilkan nilai AUC adalah 0.956 yang berarti *excellent classification*.

Berikut ini beberapa saran yang dapat dipergunakan untuk pengembangan penelitian ini adalah.

1. Perlu adanya percobaan pada algoritma yang lain seperti *k-NN* atau *Decision tree* guna untuk dapat mengetahui hasil kekuatan dari metode klasifikasi yang lain serta tingkat keakuratannya.
2. Menggunakan data yang *real time*, atau yang terbaru, sehingga sangat akurat dalam menentukan hasil dan sesuai perkembangan saat ini.

Referensi

- [1] Endriani, R., Andriani, F., & Alfina, D. (2010). Pola resistensi bakteri penyebab infeksi saluran kemih (ISK) terhadap antibakteri di pekanbaru. *Jurnal Natur Indonesia*, 12(02), 130-135.
- [2] Aji, P. B. (2016). *Sistem Diagnosa Dini Penyakit Paru Dengan Metode Naïve Bayes*. STMIK Sinar Nusantara Surakarta. eprints.sinus.ac.id/443/
- [3] M. M. Baharuddin, T. Hasanuddin, and H. Azis, "Analisis Performa Metode K-Nearest Neighbor untuk Identifikasi Jenis Kaca," *Ilk. J. Ilm.*, vol. 11, no. 28, pp. 269-274, 2019.
- [4] A. Ilham, "Komparasi Algoritma Klasifikasi Dengan Pendekatan Level Data Untuk Menangani Data Kelas Tidak Seimbang," *J. Ilm. Ilmu Komput.*, vol. 3, no. 1, pp. 9-14, 2017.
- [5] Munir, Rinaldi. (2012). *Matematika Diskrit*. Bandung: Informatika.
- [6] Dewi, I. C., Soebroto, A. A., & Furqon, M. T. (2015). Sistem Pakar Diagnosa Penyakit Sapi Potong Dengan Metode Naive Bayes. *Journal of Environmental Engineering and Sustainable Technology*, 2(2), 72-78.
- [7] Gorunescu, Florin. (2011). *Data Mining Concepts, Models and Techniques*.

- [8] Ramdan, Nur Ariesanto., Abdul K, Agyztia Preman. (2020). Modul Belajar Data Mining & RapidMiner. Klaten: Lakeisha.
- [9] Dataset diakses 15 Oktober 2020. <https://archive.ics.uci.edu/ml/datasets/Acute+Inflammations>