

Evaluasi Performa Metode Rough Set dan Algoritma Apriori dalam Mengidentifikasi Pola Penyakit Demam Tifoid

Beny Irawan^{1*}, Erwin Daniel Sitanggang²

^{1*}Institut Kesehatan Medistra Lubuk Pakam

Jl. Sudirman No.38, Petapahan, Kec. Lubuk Pakam, Kabupaten Deli Serdang, Sumatera Utara - 20512

²Universitas Mandiri Bina Prestasi

Jl. Letjend. Djamin Ginting No.285-287, Padang Bulan, Medan Baru, Kota Medan, Sumatera Utara, Indonesia - 20155

^{1*}benyirawan@medistra.ac.id, ²rwins.sitanggang@gmail.com

DOI: 10.58918/cmws6d52

Abstrak

Informasi merupakan kebutuhan vital dalam kehidupan sehari-hari dan memiliki peranan penting dalam proses pengambilan keputusan. Namun, memperoleh informasi yang tepat dan akurat sering kali menjadi tantangan tersendiri. Penelitian ini bertujuan untuk membantu pengambil keputusan dengan membandingkan kinerja metode Rough Set dan algoritma Apriori dalam proses analisis data. Metode Rough Set dimanfaatkan untuk menghasilkan aturan yang ringkas melalui proses reduksi, sedangkan algoritma Apriori digunakan untuk menemukan kombinasi itemset yang sering muncul dalam basis data berdasarkan nilai minimum support (minsup), dengan tahapan utama berupa proses join dan prune. Tujuan utama dari penelitian ini adalah meningkatkan akurasi kedua metode dalam mengidentifikasi penyakit demam tifoid. Penelitian ini menggunakan tujuh variabel input utama, dan berdasarkan pengujian terhadap dataset demam tifoid, metode Rough Set menghasilkan aturan dengan panjang empat, sedangkan algoritma Apriori menghasilkan aturan dengan panjang tiga. Hasil akhir menunjukkan bahwa kedua metode memberikan rata-rata akurasi sebesar 87,4%.

Kata Kunci: Metode Rough Set, Algoritma Apriori, Demam, Demam Tifoid.

1. Pendahuluan

Informasi merupakan elemen yang sangat penting dalam pengambilan keputusan di berbagai bidang, termasuk bidang kesehatan. Seiring dengan meningkatnya jumlah data yang tersedia, tantangan utama bukan lagi pada pengumpulan data, melainkan bagaimana menggali dan mengolah data tersebut agar menghasilkan informasi yang akurat, relevan, dan bermanfaat. Dalam konteks ini, teknik data mining menjadi sangat penting karena mampu mengekstraksi pola dan pengetahuan tersembunyi dari sejumlah besar data secara otomatis dan efisien.

Data mining adalah suatu proses seleksi, eksplorasi, dan pemodelan terhadap data dalam jumlah besar untuk menemukan pola atau tren yang tidak tampak secara eksplisit (Turban et al., 2005). Proses ini menggunakan berbagai metode seperti algoritma Apriori, Rough Set, Fuzzy Logic, dan metode lainnya untuk melakukan analisis dan prediksi berbasis data. Data mining juga telah banyak digunakan dalam bidang kesehatan, seperti untuk prediksi penyakit, pengelompokan data pasien, serta penentuan pola hubungan antar gejala dan penyakit.

Metode Rough Set, yang pertama kali dikembangkan oleh Zdzisław Pawlak pada tahun 1980-an, merupakan salah satu pendekatan matematika yang digunakan untuk menangani ketidakpastian, ketidaktepatan, dan ambiguitas dalam data. Rough Set tidak memerlukan informasi tambahan seperti probabilitas atau tingkat keanggotaan sebagaimana yang dibutuhkan dalam teori probabilitas dan fuzzy. Melalui konsep seperti lower approximation, boundary region, dan reduct, metode ini dapat digunakan untuk menemukan aturan keputusan dari data secara efisien (Hakim & Rusli, 2013; Li et al., 2007).

Sementara itu, algoritma Apriori merupakan teknik association rule mining yang digunakan untuk menemukan asosiasi antar item dalam suatu basis data, berdasarkan nilai support dan confidence (Jiawei & Kamber, 2006). Algoritma ini bekerja melalui dua tahap utama, yaitu join (penggabungan itemset) dan prune (pemangkasan berdasarkan minimum support), untuk membentuk aturan yang merepresentasikan hubungan antar atribut dalam dataset.

Beberapa penelitian sebelumnya telah menunjukkan bahwa metode Rough Set dan algoritma Apriori mampu memberikan hasil yang signifikan dalam analisis data medis. Misalnya, Budiono et al.

(2014) menggunakan algoritma Apriori untuk menemukan pola penyakit radang sendi, sedangkan penelitian oleh Adeyemo et al. (2015) menunjukkan bahwa algoritma Multilayer Perceptron (MLP) memiliki akurasi tinggi untuk diagnosis demam tifoid, meskipun C4.5 lebih cepat dalam proses pelatihan.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis kinerja metode Rough Set dan algoritma Apriori dalam mengidentifikasi pola penyakit demam tifoid. Dengan membandingkan kedua metode ini, diharapkan dapat diperoleh hasil analisis yang optimal, baik dari segi akurasi, efisiensi, maupun kualitas aturan yang dihasilkan, sehingga dapat mendukung proses diagnosis dan pengambilan keputusan di bidang kesehatan.

2. Tinjauan Pustaka

2.1. Data Mining

Data mining adalah proses penambangan pengetahuan dari sejumlah besar data untuk menemukan pola dan informasi yang bermakna (Han & Kamber, 2006). Proses ini merupakan bagian dari *Knowledge Discovery in Databases* (KDD) dan melibatkan ekstraksi, pengenalan, serta penyajian informasi dari basis data yang besar untuk mendukung pengambilan keputusan (Connolly & Begg, 2005; Berry & Linoff, 2004). Penggunaan *data mining* menjadi penting karena pertumbuhan data yang pesat dan kebutuhan untuk menginterpretasikan data secara efektif. Disiplin ilmu yang terlibat dalam *data mining* antara lain statistika, basis data, pembelajaran mesin, dan visualisasi data.

2.1.1. Data Cleaning

Data cleaning adalah tahap awal dalam proses KDD yang bertujuan untuk menangani data tidak lengkap, redundan, atau inkonsisten (Kurniawati, 2015). Salah satu teknik yang digunakan adalah *remove incomplete data*, yaitu menghapus baris data yang tidak memiliki informasi lengkap. Misalnya, jika dalam tabel data pasien terdapat entri yang kosong, maka data tersebut dihapus agar hanya data yang lengkap yang digunakan dalam analisis.

2.1.2. Data Transformation

Data transformation adalah proses mengubah data mentah ke dalam format yang sesuai untuk analisis, salah satunya dengan menggunakan algoritma fungsi interval. Transformasi ini mengonversi data kategorik menjadi bentuk numerik biner agar dapat digunakan dalam metode analisis data mining.

2.2. Teori Rough Set

Teori Rough Set dikembangkan oleh Z. Pawlak pada tahun 1982 dan merupakan metode matematis untuk menganalisis klasifikasi data dalam bentuk tabel, khususnya data diskrit (Thangavel et al., 2006). Tujuannya adalah memperoleh decision rules yang sederhana dan relevan dari data yang mengandung ketidakpastian, ketidaklengkapan, dan ketidakkonsistenan. Rough Set tidak memerlukan informasi tambahan atau distribusi probabilistik dalam proses analisis, namun kurang efektif dalam menangani atribut kontinu.

Pendekatan ini sangat berguna dalam knowledge discovery in database (KDD) dan telah diterapkan dalam berbagai bidang seperti kedokteran, farmasi, perbankan, teknik, dan pengolahan citra. Beberapa keunggulannya adalah kemampuannya dalam mereduksi data, mengevaluasi signifikansi atribut, menghasilkan aturan keputusan, dan menyediakan interpretasi langsung yang mudah dipahami. Beberapa konsep dasar dalam teori Rough Set mencakup Information System (IS) dan Decision System (DS), yaitu struktur data yang merepresentasikan objek beserta atribut kondisi dan atribut keputusan. Indiscernibility Relation digunakan untuk menggambarkan relasi antar objek yang tidak dapat dibedakan berdasarkan atribut kondisi yang sama. Konsep Equivalence Class mengelompokkan objek-objek yang memiliki nilai atribut identik. Selanjutnya, Discernibility Matrix dan Modulo D dimanfaatkan untuk mengidentifikasi atribut yang dapat membedakan antar objek. Reduct merupakan himpunan minimal dari atribut-atribut yang tetap mempertahankan kemampuan sistem dalam mengklasifikasikan data. Dari data yang telah direduksi tersebut, dilakukan proses Generating Rules untuk membentuk aturan-aturan keputusan yang merepresentasikan pola dari data yang dianalisis.

2.3. Algoritma Apriori

Algoritma Apriori dimanfaatkan untuk mengidentifikasi himpunan item yang sering muncul dalam basis data melalui proses berulang. Setiap tahap iterasi mencakup dua proses utama, yaitu pembentukan kandidat dan perhitungan nilai support. Proses dimulai dari himpunan 1-item, kemudian diperluas menjadi k-itemset dengan menggabungkan itemset dari iterasi sebelumnya yang telah memenuhi ambang batas minimum support.

Ciri khas algoritma ini adalah:

1. *Pruning*: Kandidat itemset dibuang jika subset-nya tidak termasuk dalam *frequent itemset* sebelumnya.

2. *Pemindaian ulang database*: Diperlukan pada setiap iterasi untuk menghitung *support*.

Ukuran utama dalam algoritma Apriori:

1. *Support*: Proporsi transaksi yang memuat *antecedent* dan *consequent*.
2. *Confidence*: Proporsi transaksi yang memuat *consequent* dari seluruh transaksi yang memuat *antecedent*.

$$\text{Support} = \frac{|A \cap B|}{\text{Total Transaksi}} \times 100\% \quad \text{Confidence} = \frac{|A \cap B|}{|A|} \times 100\%$$

2.3.1. Classification-Based Association

Classification-based association menggabungkan *association rule mining* dan klasifikasi. Dua pendekatan utama yaitu:

1. *ARCS*: Mengelompokkan data terlebih dahulu, lalu menghasilkan aturan berbentuk *Aquant1 ∧ Aquant2 → Acat*.
2. *Associative Classification*: Menghasilkan aturan *condset → y* dengan basis *frequent itemset*. Aturan dipilih berdasarkan *support* dan *confidence* tertinggi.

3. Metodologi Penelitian

3.1. Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksploratif dan deskriptif. Tujuan utama penelitian adalah untuk mengeksplorasi pola-pola data serta mengidentifikasi aturan-aturan pengambilan keputusan dalam diagnosis penyakit Tifoid menggunakan metode data mining berbasis *Rough Set Theory*.

3.2. Jenis dan Sumber Data

Data yang digunakan merupakan data sekunder yang diperoleh dari sistem rekam medis RS Grand Medistra Lubuk Pakam. Data mencakup variabel-variabel sebagai berikut:

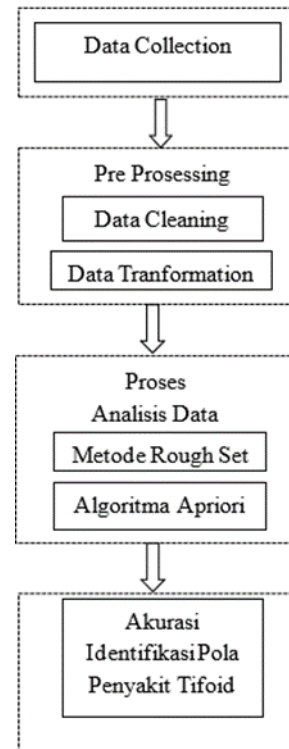
1. Identitas pasien (inisial, umur, jenis kelamin);
2. Hasil pemeriksaan laboratorium (nilai Widal, leukosit, hematokrit, dan lainnya);
3. Gejala klinis (demam, nyeri perut, diare, dan lain-lain);
4. Status diagnosis akhir (positif/negatif Tifoid).

3.3. Teknik Pengumpulan Data

Pengumpulan data dilakukan melalui studi dokumentasi dengan mencatat informasi dari sistem

informasi rekam medis elektronik rumah sakit berdasarkan kriteria yang telah ditetapkan.

3.4. Teknik Analisis Data



Gbr. 1. Diagram Alir Proses Analisis Data dengan Metode Rough Set

Analisis data dilakukan dengan pendekatan Rough Set, melalui tahapan sebagai berikut:

1. *Pra-pemrosesan Data (Preprocessing)*
 - o Seleksi atribut berdasarkan relevansi terhadap diagnosis Tifoid;
 - o Pembersihan data (*data cleaning*) untuk menangani data yang hilang atau tidak konsisten;
 - o Diskretisasi data, yaitu mengubah nilai atribut numerik ke dalam kategori diskrit sesuai kebutuhan metode *Rough Set*.
2. *Pembuatan Decision Table*

Data yang telah diproses disusun dalam bentuk *decision table*, di mana setiap baris merepresentasikan satu kasus pasien, kolom merepresentasikan atribut kondisi, dan kolom terakhir adalah atribut keputusan (diagnosis Tifoid).
3. *Penerapan Rough Set Theory*
 - o Menentukan *indiscernibility relation* berdasarkan nilai-nilai atribut;
 - o Menentukan *lower approximation* dan *upper approximation* terhadap kelas keputusan;
 - o Menghasilkan *reduct*, yaitu subset atribut minimal yang mempertahankan kemampuan klasifikasi;

- o Menyusun *decision rules* yang merepresentasikan pola pengambilan keputusan berdasarkan data.

4. Validasi Aturan Keputusan

Aturan-aturan yang dihasilkan dievaluasi menggunakan pengukuran akurasi, sensitivitas, dan spesifisitas. Validasi dilakukan melalui metode *cross-validation* atau dengan membagi data menjadi data latih dan data uji, jika memungkinkan.

4. Pembahasan dan Hasil Penelitian

4.1. Pra Pengolahan Data

Proses pra-pengolahan data dilakukan untuk memperoleh *rule* (aturan) dari kasus demam tifoid dengan menggunakan metode *Rough Set*, serta untuk mengevaluasi akurasi dari *rule* yang dihasilkan. Pengujian dilakukan dengan bantuan perangkat lunak *Rosetta*.

Data yang digunakan disusun dalam bentuk *Decision System* dan disimpan menggunakan perangkat lunak *Microsoft Excel*, seperti ditunjukkan pada Gbr. 2.

	A	B	C	D	E	F	G	H
1	Demam	Sakit Kepala	Pusing	Sakit Perut	Muntah	Batuk	Nafsu Makan Menurun	Diagnosa
2	tidak	ya	tidak	ya	ya	ya		negatif
3	tidak	ya	tidak	ya	ya	ya		negatif
4	tidak	ya	tidak	ya	ya	ya		negatif
5	tidak	ya	tidak	ya	ya	ya		negatif
6	tidak	ya	tidak	ya	ya	ya		negatif
7	ya	tidak	tidak	tidak	tidak	tidak		negatif
8	ya	tidak	tidak	tidak	tidak	tidak		negatif
9	ya	tidak	tidak	ya	tidak	tidak		positif
10	ya	tidak	tidak	ya	tidak	tidak		positif
11	ya	tidak	tidak	ya	tidak	tidak		positif
12	ya	tidak	tidak	ya	tidak	tidak		positif
13	ya	tidak	tidak	ya	tidak	tidak		positif
14	ya	tidak	tidak	ya	tidak	tidak		positif
15	ya	tidak	tidak	ya	tidak	tidak		positif
16	ya	tidak	tidak	ya	tidak	tidak		positif
17	ya	tidak	tidak	ya	tidak	tidak		positif
18	ya	tidak	tidak	ya	tidak	tidak		positif
19	ya	tidak	tidak	ya	tidak	tidak		positif
20	ya	tidak	tidak	ya	tidak	tidak		positif
21	ya	tidak	tidak	ya	tidak	tidak		positif
22	ya	tidak	tidak	ya	tidak	tidak		positif
23	ya	tidak	tidak	ya	tidak	tidak		positif
24	ya	tidak	tidak	ya	tidak	tidak		positif
25	ya	tidak	tidak	ya	tidak	tidak		positif
26	ya	tidak	tidak	ya	tidak	tidak		positif
27	ya	tidak	tidak	ya	tidak	tidak		positif
28	ya	tidak	tidak	ya	tidak	tidak		positif
29	ya	tidak	tidak	tidak	ya	tidak		negatif
30	ya	tidak	tidak	tidak	ya	tidak		negatif
31	ya	tidak	tidak	tidak	ya	tidak		negatif
32	ya	tidak	tidak	tidak	ya	tidak		negatif
33	ya	tidak	tidak	tidak	ya	tidak		negatif
34	ya	tidak	tidak	tidak	ya	tidak		negatif
35	ya	tidak	tidak	tidak	ya	tidak		negatif
36	ya	tidak	tidak	tidak	ya	tidak		negatif
37	ya	tidak	tidak	tidak	ya	tidak		negatif

Gbr. 2. Data Dalam Bentuk Excel

4.2. Hasil Implementasi Menggunakan Metode *Rough Set*

Untuk menguji keakuratan hasil pengolahan data secara manual pada bab sebelumnya, digunakan perangkat lunak *Rosetta* yang mendukung metode *Rough Set*. Proses ini diawali dengan mengimpor data dalam format Excel ke dalam *Rosetta*, seperti ditunjukkan pada Gbr. 3.

Gbr. 3. Tampilan Hasil Import Data Rosetta

4.3. Reduct

Reduct merupakan proses seleksi atribut minimal (atribut yang paling berpengaruh) dari sekumpulan atribut kondisi, menggunakan pendekatan *prime implicant* dari fungsi Boolean. Hasil reduct dari perangkat lunak *Rosetta* ditampilkan pada Gbr. 4.

	Reduct	Support	Length
1	{Demam, Sakit Kepala, Pusing, Sakit Perut}	1	4
2	{Demam, Pusing, Sakit Perut, Muntah}	1	4
3	{Demam, Sakit Kepala, Sakit Perut, Nafsu Makan Menurun}	1	4
4	{Demam, Sakit Perut, Muntah, Nafsu Makan Menurun}	1	4
5	{Demam, Sakit Perut, Batuk, Nafsu Makan Menurun}	1	4
6	{Pusing, Sakit Perut, Batuk}	1	3
7	{Sakit Kepala, Pusing, Sakit Perut, Nafsu Makan Menurun}	1	4
8	{Pusing, Sakit Perut, Muntah, Nafsu Makan Menurun}	1	4

Gbr. 4. Hasil Reduct

Dari hasil yang diperoleh, terdapat delapan *reduct* yang menjadi dasar pembentukan *rule*. Setiap reduct menunjukkan kombinasi atribut yang disederhanakan dan relevan untuk pengambilan keputusan. Kolom *support* menunjukkan banyaknya kondisi pada reduct yang sesuai dengan data awal, sementara *length* menunjukkan panjang kombinasi atribut pada reduct tersebut.

4.4. Generating Rules

Proses *generating rules* adalah langkah dalam metode *Rough Set* untuk menghasilkan aturan (rules) atau pengetahuan dari *equivalence class* dan hasil reduct. Hasil proses ini ditampilkan pada Gbr. 5.

Gbr. 5. Hasil Generating Rules

Dari pengolahan data menggunakan Rosetta, diperoleh sebanyak 57 rule. Namun, hanya delapan rule terbaik yang dipilih berdasarkan akurasi prediksi tertinggi sebesar 87,4%. Rincian rule dapat dilihat pada Tabel 1.

Tabel 1

Akurasi Rule Rough Set

No	Rule	Akurasi
1	Jika Demam (ya) AND Sakit Kepala (ya) AND Pusing (ya) AND Sakit Perut (ya) → Diagnosa (positif)	87,4%
2	Jika Demam (ya) AND Pusing (ya) AND Sakit Perut (ya) AND Muntah (ya) → Diagnosa (positif)	87,4%
3	Jika Demam (ya) AND Sakit Kepala (ya) AND Sakit Perut (ya) AND Nafsu Makan Menurun (ya) → Diagnosa (positif)	87,4%
4	Jika Demam (ya) AND Sakit Perut (ya) AND Muntah (ya) AND Nafsu Makan Menurun (ya) → Diagnosa (positif)	87,4%
5	Jika Demam (ya) AND Sakit Perut (ya) AND Batuk (ya) AND Nafsu Makan Menurun (ya) → Diagnosa (positif)	87,4%
6	Jika Pusing (ya) AND Sakit Perut (ya) AND Batuk (ya) → Diagnosa (positif)	87,4%
7	Jika Sakit Kepala (ya) AND Pusing (ya) AND Sakit Perut (ya) AND Nafsu Makan Menurun (ya) → Diagnosa (positif)	87,4%
8	Jika Pusing (ya) AND Sakit Perut (ya) AND Muntah (ya) AND Nafsu Makan Menurun (ya) → Diagnosa (positif)	87,4%
Rata-rata		87,4%

4.5. Analisis Penerapan Algoritma Apriori

Proses analisis dengan algoritma Apriori dilakukan dengan membuat kombinasi kandidat itemset berdasarkan aturan tertentu, kemudian diuji apakah kombinasi tersebut memenuhi syarat minimum *support*. Itemset yang memenuhi kriteria akan digunakan untuk membentuk *rule* dengan *confidence*

yang tinggi. Proses ini dilakukan menggunakan perangkat lunak Weka versi 3.6.9, setelah data diimpor dalam format .csv tanpa atribut keputusan seperti pada Gbr. 6.

Gbr. 6. Data dalam Format Excel

Gbr. 7. Hasil Rules Apriori

Hasil pembentukan rule menunjukkan gejala-gejala yang paling sering muncul bersamaan pada penderita demam tifoid, dengan nilai *confidence* yang tinggi. Beberapa di antaranya adalah:

1. Jika pasien mengalami Batuk (ya) → kemungkinan Nafsu Makan Menurun (ya) (*confidence* = 1).
2. Jika pasien mengalami Sakit Perut (ya) → kemungkinan Nafsu Makan Menurun (ya) (*confidence* = 1).
3. Jika pasien mengalami Batuk (ya) → kemungkinan Sakit Perut (ya) (*confidence* = 1).
4. Jika pasien mengalami Batuk (ya) dan Nafsu Makan Menurun (ya) → kemungkinan Sakit Perut (ya) (*confidence* = 1).
5. Jika pasien mengalami Muntah (ya) → kemungkinan Demam (ya) (*confidence* = 0,99).

Tabel 2

Akurasi Rule Apriori

Rule	Support	Confidence	Akurasi
Batuk (ya) dan Nafsu Makan (ya) → Diagnosa Positif	90%	100%	87,4%
Sakit Perut (ya) dan Nafsu Makan Menurun (ya) → Diagnosa Positif	90%	100%	87,4%
Batuk (ya) dan Sakit Perut (ya) → Diagnosa Positif	90%	100%	87,4%
Batuk (ya), Sakit Perut (ya), Nafsu Makan Menurun (ya) → Diagnosa Positif	90%	99%	87,4%
Muntah (ya) dan Demam (ya) → Diagnosa Positif	90%	99%	87,4%
Rata-Rata	-	-	87,4%

Berdasarkan hasil analisis Apriori, lima rule yang dihasilkan masing-masing juga menunjukkan akurasi sebesar 87,4%, menunjukkan konsistensi hasil dengan metode Rough Set.

4.6. Pembahasan

Berdasarkan hasil implementasi kedua metode, baik *Rough Set* maupun *Apriori* sama-sama menghasilkan akurasi yang tinggi dan konsisten sebesar 87,4%. Hal ini menunjukkan bahwa kedua metode mampu mengidentifikasi pola gejala demam tifoid dengan baik. *Rough Set* unggul dalam penyederhanaan atribut dan eksplorasi rule secara logis, sedangkan Apriori efektif dalam mengungkap asosiasi antar gejala dengan *support* dan *confidence* yang kuat. Temuan ini mendukung pemanfaatan data mining dalam sistem pendukung keputusan medis untuk meningkatkan kecepatan dan ketepatan diagnosis demam tifoid.

5. Kesimpulan dan Saran

Berdasarkan analisa kinerja metode rough set dan algoritma apriori pada kasus penyakit demam tifoid di RS Grand Medistra Lubuk Pakam menggunakan 500

dataset, maka dapat diambil kesimpulan bahwa hasil akurasi kinerja metode rough set dalam identifikasi penyakit demam tifoid sebesar 87,4%, sedangkan untuk kinerja algoritma apriori hasil akurasi yang didapat sebesar 87,4%. Juga aturan prediksi yang dihasilkan oleh metode rough set terdiri dari 4 gejala sedangkan aturan yang dihasilkan algoritma apriori terdiri dari 3 gejala.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada semua pihak yang telah membantu dalam penelitian ini.

Referensi

- [1] O. O. Adeyemo, T. O. Adeyeye, and O. Ogunbiyi, "Comparative study of ID3/C4.5 decision tree and multilayer perceptron algorithms for the prediction of typhoid fever," *Afr. J. Comput. ICTs*, vol. 8, no. 1, pp. 103–112, 2015.
- [2] M. J. A. Berry and G. S. Linoff, *Data Mining Techniques for Marketing, Sales, and Customer Relationship Management*, 2nd ed. Indianapolis, IN: Wiley Publishing, Inc., 2004.
- [3] T. Connolly and C. Begg, *Database Systems: A Practical Approach to Design, Implementation, and Management*. Boston, MA: Addison Wesley, 2004.
- [4] M. L. Hakim and M. Rusli, "Data mining menggunakan metode Rough Set untuk menentukan bakat mahasiswa," *Prosesor*, pp. 5–11, 2013.
- [5] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Waltham, MA: Elsevier Inc., 2012.
- [6] S. Kurniawati, "Penerapan metode Rough Set pada tingkat kepuasan konsumen terhadap kualitas pelayanan hotel," *Inf. dan Tek. Ilmiah (INTI)*, vol. 4, no. 2, pp. 138–142, 2015.
- [7] Analisa Sistem Pakar Penyakit Menular Pada Anak-Anak Dengan Metode Forward Chaining. (2023). *LOFIAN: Jurnal Teknologi Informasi Dan Komunikasi*, 2(2), 20-25. <https://doi.org/10.58918/lofian.v2i2.207>
- [8] K. Thangavel, Q. Shen, and A. Pethalakshmi, "Application of clustering for feature selection based on Rough Set theory approach," *Int. J. Artif. Intell. Mach. Learn. (AIML)*, vol. 6, no. 1, pp. 19–27, 2006.
- [9] E. Turban, J. E. Aronson, and T. P. Liang, *Decision Support Systems and Intelligent Systems*, 7th ed. Upper Saddle River, NJ: Prentice Hall, 2005.
- [10] X. Yin and J. Han, "CPAR: Classification based on predictive association rules," in *Proc. SIAM Int. Conf. Data Mining (SDM'03)*, San Francisco, 2003, pp. 331–335.
- [11] P. Berka, "Knowledge Discovery in Databases and Data Mining," in **Encyclopedia of Information Science and Technology**, IGI Global, 2024. [Online]. Available: <https://www.igi-global.com/chapter/knowledge-discovery-in-databases-and-data-mining/112586>