



SKRIPSI

ANALISIS RISIKO PERESEPAN OBAT MENGUNAKAN INTEGRASI ALGORITMA C5.0 DAN ASSOCIATION RULE MINING: STUDI KASUS PUSKESMAS KLAMPIS

NURUL KAMALIA ZAHRA
NPM 22083010021

DOSEN PEMBIMBING

Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
Andri Fauzan Adziima, M.Si

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2026**

LEMBAR PENGESAHAN

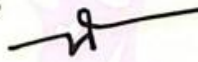
ANALISIS RISIKO PERESEPAN OBAT MENGGUNAKAN INTEGRASI ALGORITMA C5.0 DAN ASSOCIATION RULE MINING: STUDI KASUS PUSKESMAS KLAMPIS

Oleh:
NURUL KAMALIA ZAHRA
NPM. 22083010021

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur pada tanggal 09 Juni 2026:

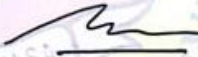
Menyetujui,

Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
NIP. 19830310 202121 1 006


.....

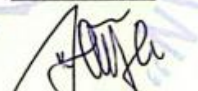
(Pembimbing I)

Andri fauzan Adziima, M. Si.
NIP. 19950512 202406 1 001


.....

(Pembimbing II)

Amri Muhaimin, S.Stat., M.Stat., M.S.
NIP. 19950723 202406 1 002


.....

(Ketua Penguji)

Shindi Shella May Wara, M. Stat.
NIP. 19960518 202406 2 003


.....

(Penguji I)

Mengetahui,

Dekan Fakultas Ilmu Komputer

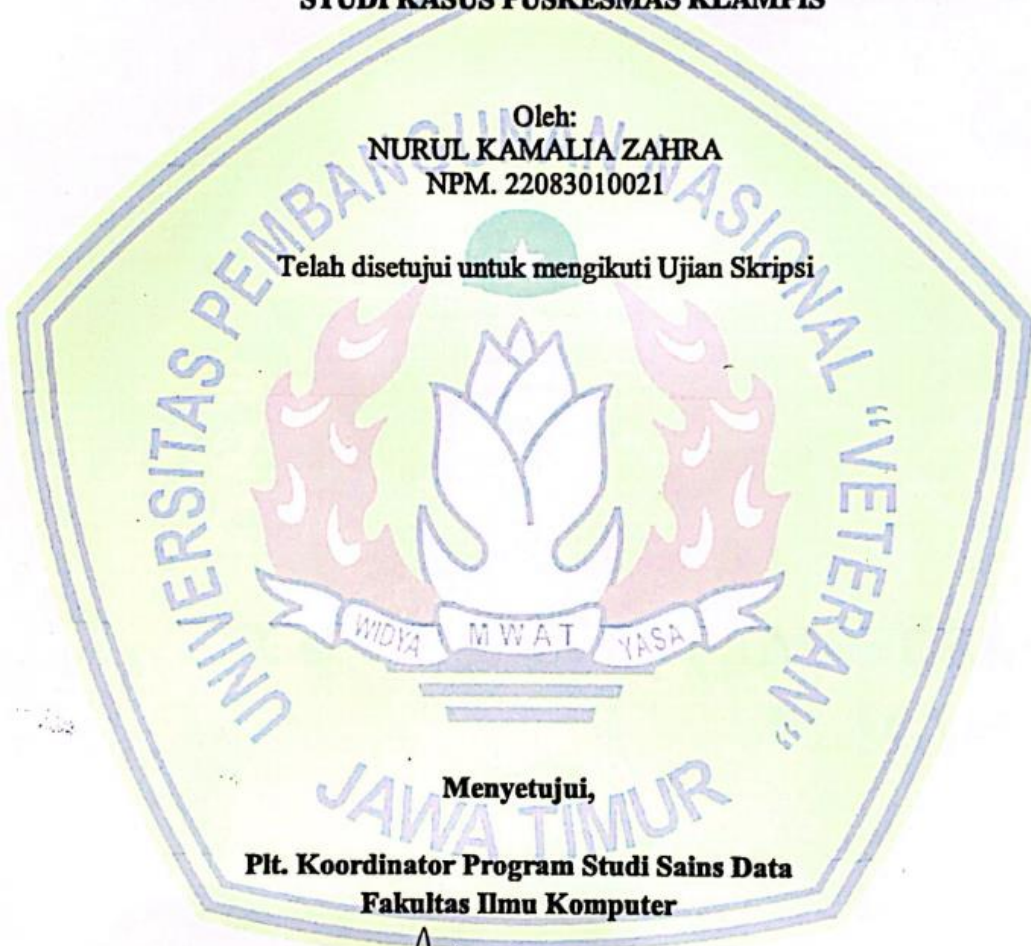

Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

ANALISIS RISIKO PERESEPAN OBAT MENGGUNAKAN INTEGRASI ALGORITMA C5.0 DAN ASSOCIATION RULE MINING: STUDI KASUS PUSKESMAS KLAMPIS

Oleh:
NURUL KAMALIA ZAHRA
NPM. 22083010021

Telah disetujui untuk mengikuti Ujian Skripsi



Menyetujui,

**Plt. Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer**

Dr. I Gede Susrama Mas Divasa, ST., MT
NIP. 19700619 2021211 009

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Nurul Kamalia Zahra
NPM : 22083010021
Program : Sarjana (S1)
Program Studi : Sains Data
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila di kemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 09 Juni 2026
Yang Membuat Pernyataan,



NURUL KAMALIA ZAHRA
NPM. 22083010021

ABSTRAK

Nama Mahasiswa / NPM : Nurul Kamalia Zahra / 22083010021
Judul Skripsi : Analisis Risiko Peresepan Obat Menggunakan Integrasi Algoritma C5.0 dan *Association Rule Mining*: Studi Kasus Puskesmas Klampis
Dosen Pembimbing : 1. Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
2. Andri Fauzan Adziima, M.Si.

Peresepan obat merupakan bagian penting dalam pelayanan kesehatan yang berpotensi menimbulkan berbagai risiko, seperti interaksi obat, efek samping, dan kesalahan penggunaan obat. Tingkat risiko tersebut dapat meningkat seiring bertambahnya jumlah obat yang diresepkan serta karakteristik obat yang digunakan. Oleh sebab itu, diperlukan metode yang tidak hanya mampu mengelompokkan tingkat risiko peresepan, tetapi juga mengidentifikasi pola hubungan antarobat yang terdapat dalam resep. Meskipun demikian, sebagian besar penelitian sebelumnya masih menganalisis klasifikasi risiko dan pola keterkaitan obat secara terpisah. Penelitian ini mengintegrasikan algoritma C5.0 dan *Association Rule Mining* (ARM) untuk menganalisis tingkat risiko peresepan obat sekaligus mengidentifikasi pola kombinasi obat pada setiap kategori risiko. Algoritma C5.0 digunakan untuk mengklasifikasikan data peresepan ke dalam tiga kelas risiko, yaitu sangat berisiko, cukup berisiko, dan sedikit berisiko. Untuk menyederhanakan model tanpa menurunkan kinerjanya secara signifikan, diterapkan metode *Reduced Error Pruning* (REP). Selanjutnya, ARM dengan algoritma Apriori digunakan untuk menemukan pola keterkaitan antarobat pada masing-masing kelas risiko. Hasil analisis kemudian disajikan melalui *Graphical User Interface* (GUI) yang interaktif dan mudah dipahami. Hasil penelitian menunjukkan bahwa model C5.0 setelah penerapan *Reduced Error Pruning* memperoleh nilai *accuracy* sebesar 94,32% dan *Kappa* sebesar 0,9135. Faktor utama yang memengaruhi tingkat risiko peresepan adalah golongan obat, jumlah obat, dan reaksi alergi. Selain itu, setiap kelas risiko memiliki karakteristik penggunaan obat dan pola keterkaitan antarobat yang berbeda. Integrasi C5.0 dan ARM terbukti efektif dalam mendukung identifikasi risiko peresepan obat serta menyediakan informasi yang bermanfaat bagi skrining resep, pemantauan terapi, dan pengambilan keputusan di fasilitas pelayanan kesehatan.

Kata Kunci: C5.0, *Reduced Error Pruning*, *Association Rule Mining*, Risiko Peresepan Obat.

ABSTRACT

Student Name / NPM : Nurul Kamalia Zahra / 22083010021
Undergraduate thesis title : *Prescription Risk Analysis Using the Integration of the C5.0 Algorithm and Association Rule Mining: A Case Study at Klampis Community Health Center*
Advisors : 1. Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
2. Andri Fauzan Adziima, M.Si.

Drug prescribing is an essential process in healthcare services that may lead to various risks, including drug interactions, adverse drug reactions, and inappropriate medication use. The level of risk tends to increase as the number of prescribed medications grows and according to the characteristics of the drugs involved. Therefore, a method is needed not only to classify prescribing risk levels but also to identify patterns of relationships among drugs within prescriptions. However, previous studies have generally focused on either risk classification or drug association analysis separately. This study integrates the C5.0 algorithm and Association Rule Mining (ARM) to analyze prescription risk levels and identify drug combination patterns within each risk category. The C5.0 algorithm was employed to classify prescription data into three risk categories: high risk, moderate risk, and low risk. To simplify the model without significantly reducing its performance, the Reduced Error Pruning (REP) method was applied. Furthermore, ARM using the Apriori algorithm was utilized to discover drug association patterns within each classified risk category. The analytical results were then presented through an interactive and user-friendly Graphical User Interface (GUI). The results indicate that the C5.0 model with Reduced Error Pruning achieved an accuracy of 94.32% and a Kappa value of 0.9135. The most influential factors affecting prescription risk levels were drug category, number of prescribed drugs, and allergic reactions. In addition, each risk category exhibited distinct characteristics in terms of drug utilization and drug association patterns. The integration of C5.0 and ARM proved effective in supporting prescription risk identification and providing valuable information for prescription screening, therapy monitoring, and decision-making in healthcare facilities.

Keywords: *C5.0, Reduced Error Pruning, Association Rule Mining, Drug Prescribing Risk.*

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Analisis Risiko Peresepan Obat Menggunakan Integrasi Algoritma C5.0 dan Association Rule Mining: Studi Kasus Puskesmas Klampis”** dapat terselesaikan dengan baik.

Penyusunan skripsi ini merupakan perjalanan yang penuh tantangan, pembelajaran, dan pengalaman berharga bagi penulis. Berkat bantuan, bimbingan, dukungan, doa, serta motivasi dari berbagai pihak, skripsi ini akhirnya dapat diselesaikan dengan baik. Oleh karena itu, dengan rasa syukur dan hormat, penulis menyampaikan ucapan terima kasih kepada:

1. Tuhan Yang Maha Esa, atas segala rahmat, karunia, kesehatan, kekuatan, serta kemudahan yang diberikan kepada penulis sehingga skripsi ini dapat diselesaikan dengan baik.
2. Diri sendiri, yang telah berusaha sebaik mungkin, bangkit di setiap kesulitan, dan tetap melangkah hingga seluruh proses ini dapat dilalui dengan baik.
3. Almarhum ayah tercinta, yang mungkin tak lagi dapat menyaksikan perjalanan ini, namun kasih sayang dan kenangannya akan selalu hidup dalam hati serta menjadi kekuatan bagi penulis. Penulis menyampaikan terima kasih yang tulus kepada ibu tercinta atas doa yang tak pernah putus, kasih sayang yang tulus, serta dukungan dan pengorbanan yang selalu menjadi kekuatan bagi penulis dalam menyelesaikan skripsi ini.
4. Ibu Dr. Ir. Novirina Hendrasarie, M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Bapak Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T. selaku dosen pembimbing utama dan Bapak Andri Fauzan Adziima, M.Si. selaku dosen pembimbing pendamping yang telah meluangkan waktu, memberikan arahan, masukan, saran, serta dukungan kepada penulis selama proses penyusunan skripsi ini.
6. Bapak/Ibu Dosen Program Studi Sains Data, yang telah membagikan ilmu, wawasan, dan pengalaman berharga kepada penulis selama masa perkuliahan.

7. Keluarga terkasih, khususnya Ibu, yang senantiasa memberikan doa, kasih sayang, dukungan, pengorbanan, serta semangat yang tiada henti kepada penulis dalam setiap proses penyelesaian skripsi ini.
8. Rizky Wahyu Ariyanto yang selalu menemani penulis, baik dalam masa-masa penuh semangat maupun saat menghadapi berbagai kesulitan selama penyusunan skripsi ini. Terima kasih atas cinta, doa, kesabaran, dukungan, dan keyakinan yang diberikan tanpa henti, sehingga menjadi penguat bagi penulis untuk terus melangkah hingga mencapai tahap ini.
9. Fainus Sholihin dan Sumiati, yang selalu memberikan motivasi, support, serta menjadi tempat berbagi cerita bagi penulis selama menjalani perkuliahan hingga penyelesaian skripsi ini.
10. Kepala Apoteker Puskesmas Klampis, yang telah berkenan membantu, memberikan arahan, serta memfasilitasi penulis dalam proses pengambilan data untuk keperluan penelitian ini.
11. Naila Mugnifa Qalbi, sebagai sahabat sekaligus teman satu Program Studi Sains Data yang telah menemani penulis sejak masa calon mahasiswa baru hingga semester akhir, serta selalu memberikan dukungan dan semangat.
12. Teman-teman, yang telah menjadi tempat berbagi cerita, semangat, dan dukungan selama perjalanan perkuliahan hingga skripsi ini terselesaikan.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, 09 Juni 2026

Penulis

DAFTAR ISI

HALAMAN SAMPUL.....	i
LEMBAR PENGESAHAN.....	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI.....	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xv
DAFTAR NOTASI	xvi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian	6
1.5. Manfaat Penelitian.....	7
BAB II TINJAUAN PUSTAKA.....	9
2.1. Penelitian Terdahulu.....	9
2.2. Kerangka Teori	17
2.2.1. Data Mining.....	17
2.1.2. Association Rule Mining	20

2.1.3.	<i>Decision Tree</i>	25
2.1.4.	Algoritma C5.0	26
2.1.5.	Teknik <i>Reduced Error Prunning</i>	29
2.1.6.	Faktor Resiko Obat	30
2.1.7.	GUI	32
BAB III METODOLOGI PENELITIAN.....		33
3.1.	Variabel Penelitian dan Sumber Data.....	33
3.2.	Langkah Analisis	35
3.2.1.	Pengumpulan data.....	35
3.2.2.	<i>Preprocessing Data</i>	37
3.2.3.	Variabel Prediktor dan Variabel target	38
3.2.4.	Pembagian Data (<i>training,validation, testing</i>)	39
3.2.5.	Pelatihan Model C5.0.....	39
3.2.6.	<i>Reduced Error Prunning</i>	40
3.2.7.	Evaluasi Model C5.0.....	41
3.2.8.	Integrasi Hasil C50 dengan Data Peresepan	41
3.2.9.	<i>Support, Confidence</i> dan <i>Lift</i>	42
3.2.10.	<i>Association Rule Mining</i>	42
3.3.	Desain Sistem.....	42
BAB IV HASIL DAN PEMBAHASAN.....		45
4.1.	Persiapan Dataset	45
4.1.1.	Pemilihan atribut.....	48
4.1.2.	<i>Data Cleaning</i>	52
4.1.3.	<i>Tokenisasi</i>	54
4.1.4.	Normalisasi Nama Obat	55
4.2.	Klasifikasi Resiko Peresepan Obat.....	58
4.2.1.	Transformasi data (<i>flattening</i>)	58
4.2.2.	Integrasi Data dengan Master Obat	60
4.2.3.	Agregasi Transaksi	63

4.2.4.	<i>Selection of Predictor Variables</i>	65
4.2.4.	<i>Selection Target Variables</i>	73
4.2.4.	Modeling C50.....	76
4.2.4.	<i>Reduce Error Prunning</i>	90
4.2.4.	Evaluasi model	96
4.2.5.	Hasil Pemodelan Klasifikasi Peresepan Obat	98
4.3.	Integrasi Hasil C50 dengan Asosiasi.....	101
4.3.1.	Penggabungan Hasil C50 dan Data Peresepan Obat	101
4.3.2.	<i>Support</i> Obat dan kombinasi Obat	102
4.3.3.	<i>Confidence</i> dan <i>Lift</i> Tiap Kelas Risiko	105
4.3.4.	<i>Association Rules Mining</i>	108
4.3.5.	Hasil Integrasi C5.0 dan Asosiasi.....	118
4.4.	Tampilan Antarmuka(GUI)	122
BAB V PENUTUP		127
5.1.	Kesimpulan	127
5.2.	Saran Pengembangan	129
DAFTAR PUSTAKA		131
LAMPIRAN		137

DAFTAR GAMBAR

Gambar 2.1. Tahapan <i>Data Mining</i>	17
Gambar 2.2. Pembentukan Kandidat <i>Itemset</i>	21
Gambar 2.3. Kerangka Algoritma C5.0.....	27
Gambar 2.4. Pemilihan Atribut	27
Gambar 2.5. Pemilihan <i>Root</i> Utama.....	29
Gambar 2.6. Pemilihan Cabang	29
Gambar 3.1. Alur Penelitian	36
Gambar 3.2. Katagori Risiko Obat.....	40
Gambar 3.3. Desain GUI C5.0.....	43
Gambar 3.4. Desain GUI ARM	44
Gambar 4.1. Distribusi Jumlah Obat	51
Gambar 4.2. <i>Tree</i> C5.0	85
Gambar 4.3. <i>Tree</i> C5.0 setelah <i>Prunning</i>	94
Gambar 4.4. <i>Dashboard</i> Klasifikasi.....	122
Gambar 4.5. <i>Dashboard</i> Asosiasi	123
Gambar 4.6. Fitur <i>Input</i> Nama Obat.....	123
Gambar 4.7. <i>Output</i> Klasifikasi	124
Gambar 4.8. Informasi Karakteristik Obat	124
Gambar 4.9. Fitur Asosiasi	125
Gambar 4.10. Hasil Asosiasi.....	126
Gambar 4.11. Hasil Asosiasi Pada Kelas Target	126

DAFTAR TABEL

Tabel 2.1. Penelitian Terdahulu	9
Tabel 3.1. Variabel Penelitian dan Sumber Data	33
Tabel 3.2. Variabel Data Informasi Obat	34
Tabel 4.1. Data Penelitian.....	45
Tabel 4.2. Data Informasi Obat.....	46
Tabel 4.3. Dataset Transaksi.....	49
Tabel 4.4. Statistik Deskriptif	50
Tabel 4.5. Cek Data Transaksi yang Hilang	52
Tabel 4.6. Cek Data Informasi Obat yang Hilang.....	52
Tabel 4.7. Tokenisasi Data	54
Tabel 4.8. Data Informasi Obat.....	55
Tabel 4.9. Transformasi Data	59
Tabel 4.10. Penggabungan Data	61
Tabel 4.11. Cek Nilai yang Kosong	62
Tabel 4.12. Agregasi Transaksi	64
Tabel 4.13. Pemetaan Jumlah Obat.....	67
Tabel 4.14. Pemetaan Golongan Obat.....	69
Tabel 4.15. Pemetaan Faktor Resiko.....	73
Tabel 4.16. Variabel Target	75
Tabel 4.17. <i>Entropy</i> Total.....	80
Tabel 4.18. Atribut Golongan Obat.....	81
Tabel 4.19. Kompleksitas dan Tingkat Kesalahan Model C5.0	87
Tabel 4.20. <i>Attribute Usage</i> Sebelum <i>Pruning</i>	88
Tabel 4.21. <i>ConfusionMatrix</i>	89
Tabel 4.22. Akurasi dan <i>kappa</i>	89
Tabel 4.23. Evaluasi Model	90
Tabel 4.24. Parameter <i>Pruning</i>	92
Tabel 4.25. Kompleksitas dan Tingkat Kesalahan Setelah <i>Pruning</i>	95
Tabel 4.26. <i>Attribute Usage</i> Setelah <i>Pruning</i>	95

Tabel 4.27. <i>Confusion Matrix Pruning</i>	96
Tabel 4.28. Akurasi dan Kappa <i>Pruning</i>	97
Tabel 4.29. Evaluasi Model <i>Pruning</i>	97
Tabel 4.30. <i>Support</i> 1 Nama Obat.....	103
Tabel 4.31. <i>Support</i> Kombinasi Nama Obat.....	104
Tabel 4.32. <i>Confidence</i> dan <i>Lift</i> Pada Setiap Peresepan	106
Tabel 4.33. Rata-rata <i>Lift</i> Pada Kelas Target.....	107
Tabel 4.34. <i>Support</i> Kelas Sangat Beresiko	111
Tabel 4.35. Asosiasi Kelas Sangat Beresiko	111
Tabel 4.36. <i>Support</i> Kelas Cukup Beresiko.....	113
Tabel 4.37. Asosiasi Kelas Cukup Beresiko.....	114
Tabel 4.38. <i>Support</i> Kelas Sedikit Beresiko	116
Tabel 4.39. Asosiasi Kelas Sedikit Beresiko	116

DAFTAR LAMPIRAN

Lampiran 1. Dataset	137
Lampiran 2. <i>Code script</i> Penelitian	138
Lampiran 3. GUI.....	139
Lampiran 4. LOA Jurnal	140
Lampiran 5. Surat Ijin Penelitian Puskesmas	141

DAFTAR NOTASI

$P(A \cap B)$	:	Peluang Terjadinya Transaksi A dan B
$Support\ A$	:	Nilai dukungan (<i>support</i>), menunjukkan seberapa sering <i>itemset</i> X muncul dalam seluruh transaksi
$Support\ (A,B)$	:	Nilai dukungan (<i>support</i>) 2 <i>itemset</i>
$Confidence(A,B)$	:	Nilai kepercayaan (<i>confidence</i>), menunjukkan seberapa besar kemungkinan item A muncul jika terdapat item B
$Lift\ Ratio$	:	Menghitung tingkat kekuatan asosiasi antara <i>antecedent</i> (A) dan <i>consequent</i> (B)
$Entropy$	:	Mengukur keacakan dalam suatu himpunan data.
S	:	Himpunan kasus
n	:	Jumlah kelas
pi	:	Proporsi dari S_i terhadap S
$Gain\ (S,A)$	:	Menghitung seberapa besar penurunan ketidakpastian pada dataset S ketika dipisahkan menggunakan atribut A .
$ S_i $	:	Jumlah kasus pada kelas ke- i
$ S $	:	Jumlah Kasus dalam S sebelum
$Gain\ (S_i)$	:	Nilai <i>Gain</i> dari satu variabel ke variabel ke- i
$\sum_{i=1}^n Entropy\ (S_i)$	:	Jumlah Nilai <i>entropy</i> suatu variabel ke- i kategori ke j
$\frac{Gain\ (A)}{\sum_{i=1}^n Entropy\ (S_i)}$	:	Menentukan atribut terbaik yang akan dijadikan <i>node</i> pemisah
$SplitInfo\ (S,A)$	:	Mengukur jumlah informasi yang dihasilkan saat sebuah dataset dibagi (di- <i>split</i>) berdasarkan nilai dari suatu atribut tertentu
$\frac{Gain\ (S,A)}{SplitInfo\ (S,A)}$	:	Rasio antara nilai <i>Information Gain</i> dengan nilai <i>Split Information</i> pada suatu atribut.