

ANALISIS ASSOCIATION RULE UNTUK IDENTIFIKASI POLA GEJALA PENYAKIT HIPERTENSI MENGGUNAKAN ALGORITMA APRIORI (STUDI KASUS: KLINIK RAFINA MEDICAL CENTER)

Salsabilla C Nurzanah¹, Syariful Alam², Teguh I Hermanto³

^{1,2,3}Program Studi Teknik Informatika Sekolah Tinggi Teknologi Wastukencana Purwakarta
Email: ¹salsabillachoeurnisa91@wastukencana.ac.id, ²syarifulalam@wastukencana.ac.id,
³teguhiman@wastukencana.ac.id

(Naskah masuk: 19 Juli 2022, diterima untuk diterbitkan: 27 Juli 2022)

Abstrak

Hipertensi didefinisikan sebagai penyakit kronis yang dapat menyebabkan tekanan darah meningkat melebihi 140/90 mmHg. Tidak adanya gejala spesifik dari penyakit hipertensi dapat menyebabkan penyakit yang serius atau komplikasi pada penderita hipertensi, oleh karena itu hipertensi dapat dilakukan sebagai “*silent killer*”. Klinik Rafina Medical Center merupakan suatu fasilitas kesehatan tingkat pertama di Kabupaten Purwakarta, Jawa Barat. Penyakit hipertensi dari tahun 2020 sampai awal tahun 2022 di Klinik Rafina Medical Center selalu termasuk ke dalam 10 diagnosa terbanyak, artinya banyak pasien yang terdiagnosa hipertensi di klinik tersebut, oleh karena itu dilakukan identifikasi menggunakan metode *data mining* yang bertujuan untuk menghasilkan informasi mengenai pola gejala hipertensi. Metode analisis data pada penelitian ini yaitu metode *Knowledge Discovery in Database* (KDD). Metode *data mining* yang digunakan yaitu metode *association rule*, dimana metode ini dapat mencari dan mengidentifikasi pola asosiasi antar atribut dalam suatu dataset. Algoritma pada penelitian ini adalah algoritma apriori. *Tools* dalam penelitian ini menggunakan *software* Orange. Penelitian ini menggunakan minimum *support* 3%, dan minimum *confidence* 70% sehingga menghasilkan 2 aturan asosiasi. Aturan asosiasi dengan nilai *confidence* dan *lift ratio* tertinggi yaitu *High Glucose*, Nyeri Tengku → Nyeri Kepala artinya jika penderita hipertensi mempunyai riwayat *High Glucose*, dan merasakan gejala Nyeri Tengku, maka penderita hipertensi 98,9% akan merasakan gejala Nyeri Kepala dengan nilai *lift ratio* sebesar 2,41 yang artinya bahwa aturan asosiasi tersebut valid. Hasil tersebut diharapkan dapat memberikan pengetahuan baru mengenai pola gejala hipertensi sehingga dapat digunakan dalam penyuluhan dan penyediaan obat bagi penderita hipertensi khususnya di Klinik Rafina Medical Center.

Kata kunci: *Hipertensi, Data Mining, Association Rule, Apriori, KDD*

ASSOCIATION RULE ANALYSIS FOR IDENTIFICATION OF HYPERTENSION SYMPTOMS PATTERNS USING APRIORI ALGORITHM (CASE STUDY: RAFINA MEDICAL CENTER CLINIC)

Abstract

Hypertension defined as a chronic disease that can cause blood pressure to rise above 140/90 mmHg. The absence of specific symptoms of hypertension can cause serious illness or complications in people with hypertension, therefore hypertension is called the “silent killer”. Rafina Medical Center Clinic is a first-arte health facility in Purwakarta, West Java. Hypertension from 2020 to early 2022 at the Rafina Medical Center Clinic is always included in the 10 most diagnoses, meaning that many patients are diagnosed wuth hypertension at the clinic, therefore identification was carried out using data mining method which aimed to generate information about the pattern of hypertension symptoms. The data analysis method in this research is the Knowledge Discovery in Database (KDD) method. The data mining method used is the association rule method, where this method can search for and identify association patterns between attributes in a dataset. The algorithm in this research is the apriori algorithm. The tools used in this research is Orange software. This research uses a minimum support of 3%, and a minimum confidence of 70% so as to produce 2 association rules. The association rules that have the highest confidence value and lift ratio are High Glucose, Neck Pain → Headache, meaning that if a hypertensive patient has a history of High Glucose, and feels symptoms of Neck Pain, then 98,9% hypertension sufferers will feel the symptoms of Headache with a lift ratio value of 2,41 which means that the association rules are valid. These results are expected to provide a new knowledge about the pattern of hypertension symptoms so that they can be used in counseling and providing medicine for hypertension sufferers, especially at the Rafina Medical Center Clinic.

Keywords: Hypertension, Data Mining, Association Rule, Apriori, KDD

1. PENDAHULUAN

Hipertensi didefinisikan sebagai penyakit kronis yang dapat menyebabkan tekanan darah meningkat melebihi 140/90 mmHg. Tidak adanya gejala spesifik dari penyakit hipertensi dapat menyebabkan penyakit yang serius atau komplikasi pada penderita hipertensi, oleh karena itu hipertensi kerap dikatakan sebagai “*silent killer*”. Penderita hipertensi biasanya akan tersadar bahwa dirinya memiliki penyakit hipertensi jika gejala yang dirasakan sudah tidak dapat ditangani sendiri dan semakin parah sehingga mereka memeriksakan diri ke fasilitas pelayanan kesehatan. Gejala yang banyak ditemukan pada penderita hipertensi diantaranya adalah sakit kepala, pusing, mual, muntah, sakit pada bagian belakang leher (tengkuk), kelelahan, lemas, gelisah, sesak nafas, epitaksis, hingga terjadi penurunan kesadaran [1]. Melihat dari kasus sebelumnya, biasanya hipertensi hanya terjadi pada orang tua saja, namun saat ini banyak anak muda yang juga memiliki hipertensi. Faktor-faktor yang menjadi pemicu hipertensi yaitu jenis kelamin, usia, genetik, tingkat stres, obesitas, riwayat minum alkohol, riwayat merokok, asupan garam yang berlebihan, pola aktivitas fisik yang tidak teratur, penyakit diabetes mellitus dan penyakit ginjal, oleh sebab itu prevalensi kejadian hipertensi selalu tinggi. Kementerian Kesehatan Republik Indonesia pada tahun 2018 menyatakan bahwa kasus hipertensi mempunyai persentase sebesar 34,1% di Indonesia, dengan perkiraan total kejadian sebanyak 63.309.620 orang, selain itu jumlah kematian karena hipertensi dan komplikasinya di Indonesia sebanyak 427.218 kematian [2]. Pengukuran tekanan darah yang telah dilakukan pada penduduk berusia ≥ 18 tahun didapatkan hasil bahwa provinsi Kalimantan Selatan adalah provinsi yang mempunyai kasus hipertensi tertinggi yaitu sebesar 44,1% dan disusul oleh provinsi Jawa Barat yang berada di posisi kedua setelah Kalimantan Selatan.

Klinik Rafina Medical Center merupakan suatu fasilitas pelayanan kesehatan tingkat pertama di Kabupaten Purwakarta, Jawa Barat yang menjalankan Program Pengelolaan Penyakit Kronis (PROLANIS). Penyakit yang dapat ditangani dalam PROLANIS antara lain penyakit hipertensi, dan diabetes mellitus, namun semenjak pandemi COVID-19, kegiatan PROLANIS di Klinik Rafina Medical Center diberhentikan sementara. Dengan adanya PROLANIS, penderita hipertensi banyak mendapatkan edukasi tentang bahaya hipertensi dan cara menyikapinya, gejala dan komplikasi hipertensi, serta penderita hipertensi juga diberikan obat hipertensi rutin setiap bulan, dan juga dilakukan pemeriksaan tekanan darah berkala sehingga kondisi penderita hipertensi dapat terkontrol secara rutin. Penyakit hipertensi dari tahun 2020 sampai awal tahun 2022 di Klinik Rafina Medical Center, selalu termasuk ke dalam 10 diagnosa

terbanyak, artinya banyak pasien yang terdiagnosa hipertensi di klinik tersebut, maka dari itu informasi mengenai pola gejala hipertensi dibutuhkan agar pasien dapat diberikan penyuluhan mengenai gejala hipertensi, dan mengapa hipertensi harus ditindak lanjuti sesegera mungkin. Dengan adanya informasi tersebut, dapat memberikan pengetahuan baru mengenai gejala penyakit hipertensi bagi suatu fasilitas pelayanan kesehatan.

Data mining didefinisikan sebagai metode untuk menggali beberapa informasi yang menarik dan juga penting dari item atau variabel yang ada pada *big data*. Data lampau dapat digunakan untuk mencari solusi pada data yang akan datang menggunakan data mining [3]. Industri yang menghasilkan *big data* mengharuskan mencari hubungan yang menarik dan mendeteksi pola yang terkandung dalam item-item yang teridentifikasi [4]. Untuk mencari pola atau hubungan yang menarik pada suatu data, dapat digunakan metode *association rule*.

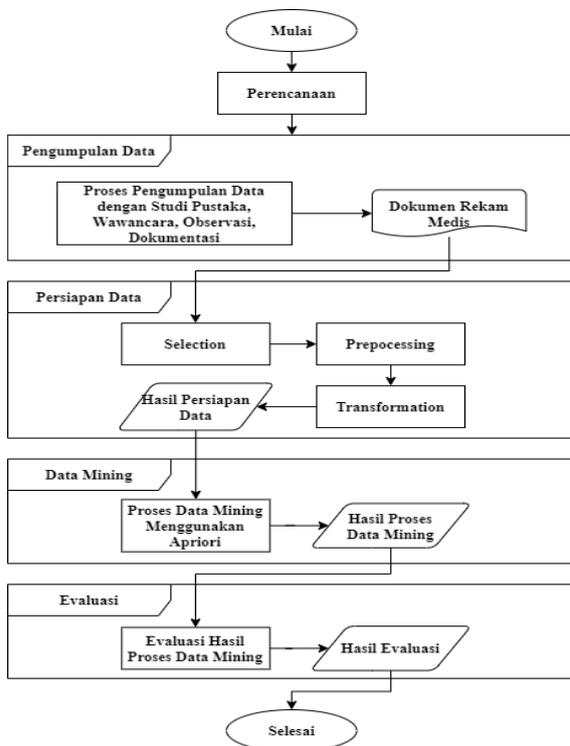
Association rule dapat membantu dalam menemukan pola yang sering terjadi, asosiasi, hubungan, dan korelasi dari kumpulan data dalam *database* transaksional [4]. Algoritma *association rule* yang banyak diterapkan untuk mengolah *big data* antara lain algoritma apriori dan FP-Growth. Algoritma apriori menghasilkan aturan asosiasi yang tingkat kekuatannya lebih tinggi jika dibandingkan dengan algoritma FP-Growth [5], maka dari itu pada penelitian ini algoritma apriori digunakan sebagai algoritma yang dapat menunjang metode *association rule*. Aturan asosiasi yang dihasilkan dari algoritma apriori dapat berguna untuk analisis data klinis, dan dapat menemukan informasi mengenai pola gejala pada pasien [6]. Informasi yang dihasilkan dari aturan asosiasi juga dapat menghasilkan pengetahuan baru. Salah satu pengetahuan baru yang dapat ditemukan dari algoritma apriori yaitu dapat mengatur dan mentata letak obat secara berurutan untuk mempermudah dalam mencari keberadaan obat [7].

Software Orange merupakan *software* yang dapat digunakan untuk proses *data mining* termasuk untuk metode *association rule*, maka dari itu pada penelitian ini digunakan *software Orange* sebagai *software* untuk mendapatkan pola yang penting dan menarik.

Berdasarkan penjelasan di atas, tujuan dari penelitian ini yaitu untuk menganalisis metode *association rule* untuk identifikasi pola gejala hipertensi menggunakan algoritma apriori. Hasil penelitian ini dapat memberikan pengetahuan baru pada fasilitas pelayanan kesehatan untuk mengetahui gejala apa saja yang dapat muncul pada penderita hipertensi, serta dapat digunakan dalam pengaturan tata letak dan penyediaan obat bagi gejala hipertensi yang aman dikonsumsi oleh penderita hipertensi, khususnya di Klinik Rafina Medical Center.

2. METODE PENELITIAN

Metode dari penelitian ini tertera pada gambar 1.



Gambar 1. Metode Penelitian

Tahapan-tahapan dari metode pada penelitian ini yaitu:

2.1. Perencanaan

Pengidentifikasi masalah dan penentuan apakah permasalahan yang akan diteliti dapat dijadikan sebagai bahan penelitian atau tidak dilakukan pada tahap perencanaan, setelah itu akan dikaji tujuan penelitian guna menentukan hasil yang akan dicapai dalam penelitian.

2.2. Pengumpulan Data

Sumber data dikumpulkan menggunakan 4 metode diantaranya:

a. Studi Pustaka

Dilakukan pengumpulan data dengan metode studi pustaka setelah tahap perencanaan. Metode ini didefinisikan sebagai metode untuk mengumpulkan data yang berasal dari artikel, buku, jurnal, dan sumber lainnya sebagai referensi untuk mendukung penelitian. Pada metode ini, dikumpulkan data yang berkaitan dengan hipertensi, *Knowledge Discovery in Database (KDD)*, *association rule*, *apriori*, dan *data mining*.

b. Wawancara

Wawancara didefinisikan sebagai metode untuk mengumpulkan data dengan mengajukan beberapa pertanyaan pada narasumber yang berada di tempat penelitian secara langsung. Wawancara dilakukan kepada Ibu Daniah, S.E. selaku pengelola Klinik Rafina Medical Center untuk mengetahui permasalahan yang sebenarnya terjadi di tempat

penelitian, serta untuk memperluas informasi yang diperoleh.

c. Observasi

Observasi merupakan metode untuk mengumpulkan data dengan mengamati proses pengolahan data rekam medis di tempat penelitian secara langsung yaitu di Klinik Rafina Medical Center, serta untuk mendapatkan bukti yang dapat mendukung hasil penelitian.

d. Dokumentasi

Metode dokumentasi dilakukan setelah wawancara dan observasi selesai. Metode ini melakukan dokumentasi untuk menganalisis dokumen atau memperoleh informasi dari sumber tertulis yang terdapat pada tempat penelitian.

Pengumpulan data yang telah dilakukan menghasilkan 2 sumber data yaitu:

a. Data Primer

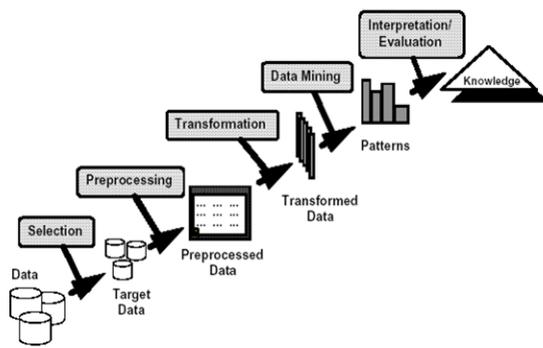
Data primer pada penelitian ini dihasilkan dari observasi dan wawancara dengan pengelola Klinik Rafina Medical Center mengenai ruang lingkup klinik, PROLANIS, dan informasi lain mengenai kejadian hipertensi di Klinik Rafina Medical Center.

b. Data Sekunder

Data sekunder pada penelitian ini dihasilkan dari metode dokumentasi seperti struktur organisasi, sejarah, serta dokumen rekam medis yang menjadi sumber utama dari penelitian ini. Data rekam medis yang dibutuhkan pada penelitian ini yaitu data rekam medis pasien yang memiliki riwayat hipertensi.

Didapatkan 3.204 data rekam medis pasien yang memiliki riwayat hipertensi dari bulan Januari 2020 hingga Maret 2022 dari pengumpulan data yang telah dilakukan. Data sekunder juga didapatkan melalui metode studi pustaka dengan mengumpulkan referensi yang dapat mendukung penelitian ini, seperti dari artikel, jurnal, serta buku yang berkaitan dengan hipertensi, *Knowledge Discovery in Database (KDD)*, *association rule*, *apriori*, dan *data mining*.

Setelah data terkumpul, selanjutnya data akan dianalisis. Proses analisis data dalam penelitian ini menggunakan metode *Knowledge Discovery in Database (KDD)*. KDD merupakan metode analisis data yang banyak diterapkan pada penelitian untuk *data mining*. KDD terdiri dari 5 tahap yaitu *selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation* yang tertera pada gambar 2.



Gambar 1. Tahapan KDD [8]

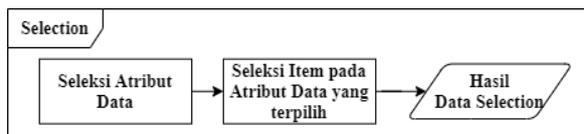
Adapun tahapan-tahapan dari KDD diantaranya:

2.3. Persiapan Data

Ada beberapa tahapan KDD yang dilakukan pada tahap persiapan data diantaranya sebagai berikut.

a. Selection

Tahap *selection* merupakan tahapan untuk menyeleksi data yang dibutuhkan dalam penelitian. Adanya tahap *selection* dapat membuat proses pengolahan data yang dilakukan dapat sesuai dengan tujuan dari penelitian. Pada tahap *selection*, dilakukan beberapa tahapan yang tertera pada gambar 3.

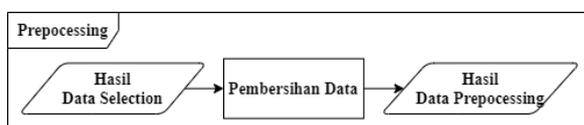


Gambar 2. Tahapan Selection

Berdasarkan gambar 3. diketahui bahwa pada tahapan *selection* dilakukan seleksi atribut data (untuk menyeleksi atribut atau variabel data yang menjadi sumber data utama dalam penelitian), serta seleksi item pada atribut atau variabel data yang terpilih.

b. Preprocessing

Tahap preprocessing dapat dilakukan dengan membuang data ganda, memeriksa konsistensi data, serta mengoreksi kesalahan-kesalahan yang ada dalam data. Dilakukan beberapa tahapan pada tahap preprocessing yang tertera pada gambar 4.

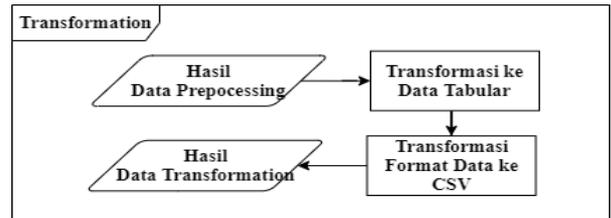


Gambar 3. Tahapan Preprocessing

Berdasarkan gambar 4. Diketahui bahwa dilakukan pembersihan atau penghapusan data pada hasil tahap *selection*.

c. Transformation

Tahap *transformation* merupakan tahap penyesuaian data yang digunakan sehingga data tersebut dapat diproses atau diolah menggunakan *data mining*. Dilakukan beberapa tahapan pada tahap *transformation* yang tertera pada gambar 5.



Gambar 4. Tahapan Transformation

Berdasarkan gambar 5. diketahui bahwa data yang telah di-*selection* dan *preprocessing*, kemudian dilakukan transformasi ke dalam data tabular, serta format data tersebut ditransformasi menjadi format CSV.

2.4. Data Mining

Dilakukan proses pengolahan data menggunakan teknik *data mining* pada tahap ini. Adapun alur dari tahapan data mining dalam penelitian ini tertera pada gambar 6.



Gambar 5. Tahapan Data Mining

Hasil dari data transformation, kemudian diolah menggunakan algoritma apriori pada data mining. Data mining merupakan alat yang memiliki tujuan untuk mendeteksi pola baru dalam suatu big database yang bermanfaat dalam pengambilan keputusan di masa depan [9].

Berdasarkan tugasnya, data mining terbagi menjadi beberapa kelompok [7], yaitu:

- Deskripsi: Pola atau tren pada data terkadang hanya ingin dideskripsikan oleh peneliti.
- Estimasi: Serupa dengan klasifikasi kecuali pada jenis variabel pada targetnya adalah numerik. Estimasi digunakan untuk memperkirakan suatu kondisi yang belum pernah ada, serta dinyatakan dalam bentuk kuantitatif. Contoh algoritma estimasi diantaranya Regresi Linier, Confidence Interval Estimations, Support Vector Machine, Neural Network dan lain-lain.
- Prediksi: Serupa dengan estimasi dan klasifikasi, kecuali hasil dari prediksi tersebut akan ada di masa depan. Prediksi digunakan untuk memperkirakan kejadian yang belum pernah terjadi. Contoh algoritma prediksi diantaranya adalah K-Nearest Neighbor (KNN), Decision Tree, dan lain-lain.
- Klasifikasi: Klasifikasi digunakan untuk menemukan metode yang dapat menjelaskan atau membagi konsep dari data agar dapat memprediksi kelas dari objek-objek yang labelnya belum diketahui. Contoh algoritma klasifikasi diantaranya adalah ID3, C4.5, Naïve Bayes, dan lain-lain.
- Pengklasteran: Pengklasteran dilakukan pada data dengan karakteristik yang serupa, yang kemudian

- data tersebut akan dikelompokkan atau diidentifikasi secara bersamaan. Contoh algoritma pengklasteran diantaranya adalah K-Means, K-Medoids, dan lain-lain.
- f. Asosiasi: Asosiasi digunakan untuk mencari atau mendeteksi item yang dapat muncul secara bersamaan pada satu waktu. Sehingga sering disebut market basket analysis dalam dunia bisnis, misalnya ketika orang membeli pasta gigi, biasanya mereka juga membeli sikat gigi. Contoh algoritma asosiasi diantaranya adalah Apriori, *Frequent Pattern Growth* (FP-Growth), Chi Square, dan lain-lain.

Association rule adalah metode data mining untuk mendeteksi pola dataset yang akan diteliti. Penerapan aturan asosiasi pada data mining bertujuan mendeteksi informasi dari item yang saling terhubung dalam bentuk aturan asosiasi [10]. Aturan asosiasi didapatkan dari hasil perhitungan yang terdiri dari 2 ukuran [11], yaitu:

- a. *Support*: Berdasar pada keseluruhan data transaksi, support bertujuan untuk mengetahui seberapa tinggi tingkat dominasi dari itemset pada data transaksi tersebut. Support akan menentukan apakah pencarian nilai confidence pada suatu itemset layak dilakukan atau tidak. Support juga dapat digunakan untuk mengetahui tingkat dominasi item tunggal. Cara mencari support pada 1-itemset tertera pada persamaan (1) berikut [11]:

$$Support(A) = \frac{Jumlah\ transaksi\ mengandung\ A}{Total\ Transaksi} \times 100\% \tag{1}$$

Adapun cara mencari *support* pada 2-itemset tertera pada persamaan (2) berikut [12]:

$$Support(A \cap B) = \frac{Jumlah\ transaksi\ mengandung\ A\ dan\ B}{Total\ Transaksi} \times 100\% \tag{2}$$

- b. *Confidence*: Bertujuan untuk mengukur seberapa kuat hubungan antaritem dalam suatu dataset. Cara mencari support tertera pada persamaan (3) berikut:

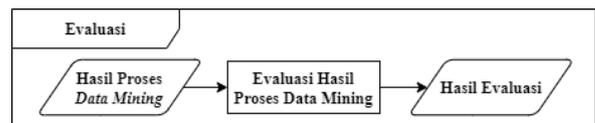
$$Confidence = P(B|A) = \frac{Jumlah\ transaksi\ mengandung\ A\ dan\ B}{Total\ Transaksi\ mengandung\ A} \times 100\% \tag{3}$$

Algoritma apriori didefinisikan sebagai algoritma *data mining* yang sering digunakan dalam metode *association rule*. Algoritma apriori berperan dalam penemuan pola frekuensi tinggi. Pola frekuensi tinggi tersebut merupakan pola dari item yang frekuensinya berada di atas ambang batas tertentu dalam suatu *database* [13]. Adapun tahap dari apriori antara lain sebagai berikut [14]:

- a. Pembentukan kandidat itemset. Kombinasi (k-1)-itemset yang telah didapatkan dari iterasi sebelumnya dapat membentuk kandidat itemset.
- b. Pehitungan *support* dari setiap kandidat k-itemset. Untuk mengukur jumlah transaksi yang mempunyai item, dibutuhkan *support* dari tiap-tiap kandidat yang didapatkan dengan meneliti *database* yang akan digunakan. Cara mencari *support* dapat dilakukan menggunakan perhitungan pada persamaan (1) dan persamaan (2).
- c. Analisis pola frekuensi tinggi. Pola frekuensi tinggi ditetapkan dari kandidat k-itemset yang melebihi nilai minimum *support*.
- d. Apabila pola frekuensi tinggi tidak didapatkan lagi, maka keseluruhan proses akan dihentikan. Dan jika tidak, maka k harus ditambah 1 kemudian proses kembali lagi ke tahap 1.

2.5. Evaluasi

Berikut tahapan evaluasi yang tertera pada gambar 7.



Gambar 6. Tahapan Evaluasi

Tahap evaluasi bertujuan untuk memeriksa apakah pola yang dihasilkan sesuai atau bertentangan dengan hipotesis yang ada. Akan dilakukan pengukuran akurasi dari suatu informasi yang dihasilkan pada tahap evaluasi. Pengukuran akurasi lift ratio digunakan untuk unsupervised learning [14]. Lift ratio merupakan tahapan uji akurasi untuk menguji kekuatan atau kevalidan dari aturan asosiasi yang sudah didapatkan [15]. Sebuah aturan asosiasi dikatakan valid atau kuat, apabila nilai lift ratio-nya > 1, maka dari itu dapat dikatakan bahwa pada transaksi tersebut satu item dan item lainnya benar-benar dibeli secara bersamaan [16]. Cara mencari lift ratio tertera pada persamaan (4) berikut:

$$Lift\ Ratio = \frac{Support(A \cap B)}{Support(A) * Support(B)} \times 100\% \tag{4}$$

Berdasarkan penjelasan di atas, dapat diketahui bahwa dilakukan evaluasi dengan menggunakan pengukuran lift ratio dalam penelitian ini.

3. HASIL DAN PEMBAHASAN

Metode dari penelitian ini tertera pada gambar 1.

3.1. Selection

Sumber data dalam penelitian ini adalah data pasien yang memiliki 136iwayat hipertensi dari bulan Januari 2020 hingga Maret 2022 yang berjumlah 3.204 data, di mana pada data rekam medis terdapat 10 atribut yang terdiri dari tanggal pelayanan, nama pasien, jenis kelamin, 136iwayat136, PRB,

PROLANIS, tanggal lahir, keluhan, terapi, dan tekanan darah.

a. Seleksi Atribut Data

Hanya 1 atribut saja yang digunakan dalam penelitian ini dari 10 atribut yang terdapat pada data rekam medis, yaitu atribut keluhan. Atribut keluhan merupakan atribut yang berisi keluhan atau gejala yang dirasakan pasien pada saat berobat yang menjadi pertimbangan dokter dalam mengambil 137iwayat137.

b. Seleksi Item dari Atribut Data

Atribut keluhan mempunyai 131 item yang berisi keluhan atau gejala yang dialami pasien hipertensi. Dari 131 item, hanya dipilih 27 item saha yang memiliki jumlah kasus ≥ 50 , atau frekuensi kemunculannya $\geq 2\%$ dari total 3.204 data rekam medis pasien hipertensi.

Adapun hasil dari data selection tertera pada tabel 1.

Tabel 1. Contoh Hasil Data Selection

no	keluhan
1	Nyeri Kepala
2	Nyeri Kepala, Nyeri Badan
3	Nyeri Kepala
4	Nyeri Ulu Hati, Mual, Overweight
5	Riw. Darah Tinggi Keluarga, Nyeri Badan, Nyeri Ulu Hati, Gatal Seluruh Badan
6	Kesemutan, Nyeri Tumit Kaki
7	Nyeri Badan, Nyeri Kepala
8	Nyeri Bahu, Nyeri Dada
9	Nyeri Kepala, Demam
10	Nyeri Kepala

3.2. Preprocessing

Pada tahap sebelumnya hanya dipilih 27 data dari atribut keluhan yang menyebabkan terdapat data yang tidak memiliki item sehingga dari 3.204 data rekam medis pasien yang memiliki 137iwayat hipertensi, dilakukan pembersihan pada 237 data sehingga pada penelitian ini hanya digunakan 2.967 data. Adapun hasil dari preprocessing tertera pada tabel 2.

Tabel 2. Contoh Hasi Data Preprocessing

no	keluhan
1	Nyeri Kepala
2	Nyeri Kepala, Nyeri Badan
3	Nyeri Kepala
4	Nyeri Ulu Hati, Mual, Overweight
5	Riw. Darah Tinggi Keluarga, Nyeri Badan, Nyeri Ulu Hati, Gatal Seluruh Badan
6	Nyeri Badan, Nyeri Kepala
7	Nyeri Kepala, Demam
8	Nyeri Kepala

3.3. Transformation

Tahap transformation dibagi menjadi 2 proses yaitu:

a. Transformasi ke Data Tabular

Data yang telah di-preprocessing kemudian ditransformasi ke dalam format tabular agar dapat memudahkan pembacaan dan pengolahan data pada proses data mining. Adapun hasil dari transformasi data ke format tabular tertera pada gambar 8.

	A	B	C	D	E	F	G	H
1	Kasus	Nyeri Kepala	Nyeri Badan	Batuk	Flu	Demam	Nyeri Ulu Hati	Mual
2	1	1						
3	2	1	1					
4	3	1		1		1		
5	4						1	1
6	5		1				1	
7	6	1						
8	7	1	1					
9	8		1					
10	9	1				1		
11	10	1						
12	11							
13	12	1						
14	13	1						
15	14							
16	15							1

Gambar 7. Transformasi ke Data Tabular

b. Transformasi Format Data ke CSV

Data yang telah ditransformasi menjadi data tabular, masih dalam format .XLSX atau Excel 2016. Selanjutnya data ditransformasikan 137iwayat137 format Comma Delimited (.CSV) agar data dapat diproses di Software Orange.

	A	B	C	D	E	F	G	H
1	Kasus,"Nyeri Kepala","Nyeri Badan","Batuk","Flu","Demam","Nyeri Ulu Hati","Mual","C							
2	1,"1"							
3	2,"1","1"							
4	3,"1","1","1"							
5	4,"1","1","1","1"							
6	5,"1","1","1","1","1"							
7	6,"1","1","1","1","1","1"							
8	7,"1","1","1","1","1","1","1"							
9	8,"1","1","1","1","1","1","1","1"							
10	9,"1","1","1","1","1","1","1","1","1"							
11	10,"1","1","1","1","1","1","1","1","1","1"							
12	11,"1","1","1","1","1","1","1","1","1","1","1"							
13	12,"1","1","1","1","1","1","1","1","1","1","1","1"							
14	13,"1","1","1","1","1","1","1","1","1","1","1","1","1"							
15	14,"1","1","1","1","1","1","1","1","1","1","1","1","1","1"							
16	15,"1","1","1","1","1","1","1","1","1","1","1","1","1","1","1"							
17	16,"1","1","1","1","1","1","1","1","1","1","1","1","1","1","1","1"							

Gambar 8. Transformasi Format Data ke CSV

3.4. Data Mining

Diterapkan metode data mining association rule sebagai metode untuk mengolah data menggunakan algoritma apriori pada penelitian ini. Tahapan yang dilalui pada proses data mining menggunakan algoritma apriori adalah sebagai berikut.

a. Pola Frekuensi Tinggi

Tahap ini bertujuan untuk mencari kombinasi item yang berada di atas nilai minimum support.

Iterasi-1 mulai dilakukan pada tahap ini untuk membentuk kandidat 1-itemset (C1) dari data keluhan tersebut dan hitung jumlah support-nya dengan nilai minimum support yang ditentukan sebesar 3%. Adapun hasil dari seleksi iterasi-1 (Large-Itemset 1) tertera pada tabel 3.

Tabel 3. Large-Itemset 1

Itemset	jumlah gejala	support (%)
Nyeri Kepala	1219	41,1
Nyeri Badan	183	6,2
Batuk	435	14,7
Flu	298	10,0
Demam	267	9,0
Nyeri Ulu Hati	508	17,1
Mual	644	21,7
Gatal Seluruh Badan	102	3,4
High Glucose	597	20,1
Pegal	206	6,9

Itemset	jumlah gejala	support (%)
Nyeri Pinggang	126	4,2
Nyeri Kaki	122	4,1
Nyeri Lutut	160	5,4
Nyeri Tengukuk	282	9,5
Pusing	200	6,7
Pusing Berputar	213	7,2
High Cholesterol	226	7,6
Kembung	116	3,9
Meriang	134	4,5
Nyeri Tenggorokan	115	3,9
High Uric Acid	100	3,4

1. Kombinasi 2-Itemset

Dilakukan proses *cross item* L2 untuk membentuk kandidat C2 (memiliki 2-itemset) pada iterasi-2, lalu akan dihitung nilai *support*-nya dengan nilai minimum *support* 3%. Adapun hasil dari seleksi iterasi-2 (Large-Itemset 2 tertera pada tabel 4.

Tabel 4. Large-Itemset 2

Itemset	jumlah gejala	support (%)
Batuk, Nyeri Kepala	117	3,94
Demam, Nyeri Kepala	90	3,03
Nyeri Ulu Hati, Nyeri Kepala	194	6,54
Mual, Nyeri Kepala	248	8,36
High Glucose, Nyeri Kepala	227	7,65
Nyeri Tengukuk, Nyeri Kepala	246	8,29
Nyeri Kepala, Batuk	117	3,94
Flu, Batuk	207	6,98
Demam, Batuk	125	4,21
Batuk, Flu	207	6,98
Demam, Flu	98	3,30
Nyeri Kepala, Demam	90	3,03
Batuk, Demam	125	4,21
Flu, Demam	98	3,30
Nyeri Kepala, Nyeri Ulu Hati	194	6,54
Mual, Nyeri Ulu Hati	232	7,82
Nyeri Kepala, Mual	248	8,36
Nyeri Ulu Hati, Mual	232	7,82
Pusing Berputar, Mual	119	4,01
Nyeri Kepala, High Glucose	227	7,65
High Cholesterol, High Glucose	104	3,51
Nyeri Tengukuk, High Glucose	90	3,03
Nyeri Kepala, Nyeri Tengukuk	246	8,29
Mual, Pusing Berputar	119	4,01
High Glucose, High Cholesterol	104	3,51
High Glucose, Nyeri Tengukuk	90	3,03

2. Kombinasi 3-Itemset

Dilakukan proses *cross item* L3 untuk membentuk kandidat C3 (memiliki 3-itemset) pada iterasi-3, lalu akan dihitung nilai *support*-nya dengan nilai minimum *support* 3%. Adapun hasil dari seleksi iterasi-3 (Large-Itemset 3) tertera pada tabel 5.

Tabel 5. Large-Itemset 3

itemset	jumlah gejala	support (%)
High Glucose, Nyeri Kepala, Nyeri Tengukuk	91	3,033
Nyeri Tengukuk, Nyeri Kepala, High Glucose	91	3,033
Nyeri Kepala, High Glucose, Nyeri Tengukuk	91	3,033
Nyeri Tengukuk, High Glucose, Nyeri Kepala	91	3,033

itemset	jumlah gejala	support (%)
Nyeri Kepala, Nyeri Tengukuk, High Glucose	91	3,033
High Glucose, Nyeri Tengukuk, Nyeri Kepala	91	3,033

3. Kombinasi 4-Itemset

Saat dilakukan proses *cross item* L4 untuk membentuk kandidat C4 (memiliki 4-itemset) pada iterasi ke-4, tidak didapatkan kandidat yang memenuhi nilai *138iwayat support* yaitu 3%, maka tidak ada satupun itemset yang terbentuk pada humpunan L4. Maka iterasi akan berhenti, dan pola frekuensi tinggi ada pada iterasi-3.

b. Pembentukan Aturan Asosiasi

Berdasar pada pola frekuensi tinggi yang telah terbentuk, selanjutnya dilakukan pemisahan item ke dalam 2 bagian yaitu Antecedent, dan Consequent yang bertujuan untuk mencari seluruh kemungkinan terbentuknya aturan asosiasi, lalu akan dicari nilai *confidence*-nya. Adapun hasil dari pembentukan aturan asosiasi dari semua kombinasi itemset dengan minimum *confidence* 70%, tertera pada tabel 6.

Tabel 6. Aturan Asosiasi

itemset	support item (%)	support antecedent (%)	confidence (%)
Nyeri Tengukuk => Nyeri Kepala	8,359	9,505	87,9
Nyeri Tengukuk, High Glucose => Nyeri Kepala	3,033	3,067	98,9
Nyeri Tengukuk => Nyeri Kepala	3,033	3,067	98,9

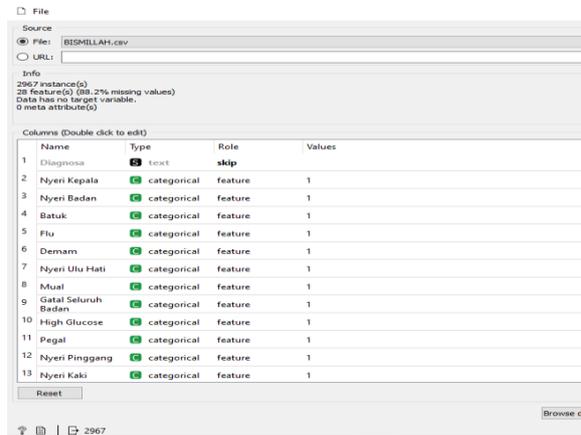
Berdasarkan tabel 6. Diketahui bahwa terdapat aturan asosiasi yang memenuhi nilai minum *support* 3%, dan nilai *138iwayat confidence* 70%. Berdasarkan aturan tersebut terdapat aturan yang menunjukkan keterlibatan kombinasi 2 itemset dan kombinasi 3 itemset, maka dari itu terdapat 3 aturan asosiasi yang dihasilkan oleh algoritma apriori menggunakan perhitungan manual:

- Jika penderita hipertensi merasakan gejala Nyeri Tengukuk, maka akan merasakan gejala Nyeri Kepala dengan nilai *support* sebesar 8,359% dan nilai *confidence* sebesar 87,9%.
- Jika penderita hipertensi merasakan gejala Nyeri Tengukuk dan mempunyai *138iwayat* High Glucose, maka akan merasakan gejala Nyeri Kepala dengan nilai *support* sebesar 3,033%, dan nilai *confidence* sebesar 98,9%.
- Jika penderita hipertensi mempunyai *138iwayat* High Glucose dan merasakan gejala Nyeri Tengukuk, maka akan merasakan gejala Nyeri Kepala dengan nilai *support* sebesar 3,033%, dan nilai *confidence* sebesar 98,9%.

Data Mining dengan Software Orange

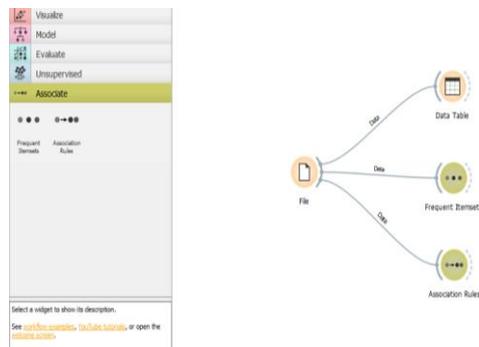
Data dapat diolah pada aplikasi *Orange* setelah data ditransformasikan ke dalam bentuk tabular, serta format data sudah ditransformasikan ke dalam bentuk

CSV. Dibutuhkan 4 operator untuk mendapatkan aturan asosiasi menggunakan algoritma apriori di aplikasi Orange yaitu operator *File*, *Data Table*, *Frequent Itemset*, dan *Association Rule*. Operator *File* digunakan untuk memasukkan data yang akan diolah yang tertera pada gambar 10.



Gambar 9. Input Data ke Software Orange

Hubungkan operator *File* dengan operator *Data Table* (untuk melihat dataset yang telah kita masukkan dalam bentuk tabel), operator *Frequent Itemset* (untuk mencari pola frekuensi tinggi), dan operator *Association Rules* (untuk membentuk aturan asosiasi) yang tertera pada gambar 11.



Gambar 10. Model Data Menggunakan Software Orange

a. Analisis Pola Frekuensi Tinggi

Masukkan nilai minimum *support* pada operator *Frequent Itemset* agar pola frekuensi tinggi dapat terbentuk. Pada penelitian ini, digunakan nilai minimum *support* 3%, dan menghasilkan pola frekuensi tinggi yang dilihat pada gambar 12.

Gambar 11. Hasil Frequent Itemset pada Software Orange

Berdasarkan gambar 12. diketahui bahwa pola frekuensi tinggi berhenti di iterasi-3 atau hanya mempunyai maksimal 3 item yang membentuk pola frekuensi tinggi, yaitu item Nyeri Kepala, High Glucose, dan Nyeri Tengku.

b. Pembentukan Aturan Asosiasi

Masukkan nilai minimum *support* dan nilai minimum *confidence* pada operator *Association Rules* untuk membentuk aturan asosiasi. Digunakan nilai minimum *support* sebesar 3%, dan nilai minimum *confidence* sebesar 70% pada penelitian ini sehingga terbentuklah aturan asosiasi yang tertera pada gambar 13.

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent		
0.084	0.879	0.095	4.323	2.141	0.045	Nyeri Tengku=1	→	Nyeri Kepala=1
0.030	0.989	0.031	13.396	2.407	0.018	High Glucose=1, Nyeri Tengku=1	→	Nyeri Kepala=1

Gambar 12. Hasil Association Rule pada Software Orange

Berdasarkan gambar 13. diketahui bahwa terdapat aturan asosiasi yang memenuhi nilai minimum *support* 3%, dan nilai minimum *confidence* 70%. Terdapat aturan yang menunjukkan keterlibatan kombinasi 2 itemset dan kombinasi 3 itemset dari aturan asosiasi tersebut, maka dari itu terdapat 2 aturan asosiasi yang dihasilkan oleh algoritma apriori menggunakan *software Orange* yaitu:

1. Jika penderita hipertensi merasakan gejala Nyeri Tengku, maka akan merasakan gejala Nyeri Kepala dengan nilai *support* sebesar 0,084 atau 8,4%, dan nilai *confidence* sebesar 0,879 atau 87,9%.
2. Jika penderita hipertensi mempunyai riwayat *High Glucose* dan merasakan gejala Nyeri Tengku, maka akan merasakan gejala Nyeri Kepala dengan nilai *support* sebesar 0,030 atau 3%, dan nilai *confidence* sebesar 0,989 atau 98,9%.

3.5. Evaluasi

Pengukuran Lift Ratio akan digunakan pada penelitian ini. Sebuah aturan asosiasi dikatakan valid atau kuat jika mempunyai nilai lift ratio > dari 1 [10], artinya dalam transaksi tersebut gejala A dan gejala B banyak dirasakan oleh pasien sebagai gejala hipertensi. Adapun hasil dari perhitungan *lift ratio* tertera pada tabel 7.

Tabel 7. Nilai *Lift Ratio* dari *Association Rule*

Itemset	support A \cap B	lift ratio	hasil
Nyeri Tengku => Nyeri Kepala	8,359	2,14	Valid
Nyeri Tengku, High Glucose => Nyeri Kepala	3,033	2,41	Valid
High Glucose, Nyeri Tengku => Nyeri Kepala	3,033	2,41	Valid

Berdasarkan tabel 7. dapat diketahui bahwa terdapat 3 aturan asosiasi yang dihasilkan yaitu Nyeri Tengku \rightarrow Nyeri Kepala, Nyeri Tengku, High Glucose \rightarrow Nyeri Kepala, dan High Glucose, Nyeri Tengku \rightarrow Nyeri Kepala. Dari 3 aturan asosiasi tersebut didapatkan nilai *lift ratio* > 1, maka aturan asosiasi tersebut dinyatakan valid.

Evaluasi Pemodelan Data Mining Menggunakan Software Orange

Adapun nilai *lift ratio* pada aturan asosiasi yang muncul menggunakan *software Orange* tertera pada gambar 14.

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent
0.084	0.879	0.095	4.323	2.141	0.045	Nyeri Tengku=1 - Nyeri Kepala=1
0.030	0.989	0.031	13.396	2.407	0.018	High Glucose=1, Nyeri Tengku=1 - Nyeri Kepala=1

Gambar 13. Nilai *Lift Ratio* dari *Association Rule* pada *Software Orange*

Berdasarkan gambar 14. diketahui bahwa terdapat 2 aturan asosiasi yang dihasilkan yaitu Nyeri Tengku \rightarrow Nyeri Kepala, dan High Glucose, Nyeri Tengku \rightarrow Nyeri Kepala. Dari 2 aturan asosiasi tersebut didapatkan nilai *lift ratio* > 1, maka aturan asosiasi tersebut dinyatakan valid.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, disimpulkan bahwa peneliti berhasil menganalisis data rekam medis pasien yang memiliki riwayat hipertensi dari rentang waktu bulan Januari 2020 sampai bulan Maret 2022 dengan jumlah kasus sebanyak 2.967 data. Berdasarkan pengolahan dan pengujian yang telah dilakukan, algoritma apriori mampu digunakan untuk menentukan gejala hipertensi yang paling sering dirasakan oleh penderita hipertensi dengan melihat kecenderungan gejala yang dirasakan penderita hipertensi. Hasil dari penelitian menggunakan *software Orange* dengan nilai minimum support 3%,

dan minimum confidence 70% yaitu didapatkan 2 aturan asosiasi. Dari 2 aturan asosiasi tersebut, aturan asosiasi yang mempunyai nilai confidence dan lift ratio tertinggi yaitu High Glucose, Nyeri Tengku \rightarrow Nyeri Kepala artinya jika penderita hipertensi mempunyai riwayat High Glucose, dan merasakan gejala Nyeri Tengku, maka penderita hipertensi 98,9% akan merasakan gejala Nyeri Kepala dengan nilai lift ratio sebesar 2,41 yang artinya bahwa aturan asosiasi tersebut valid. Dapat diketahui gejala-gejala yang paling sering dirasakan secara bersamaan oleh penderita hipertensi dari hasil aturan asosiasi yang telah diperoleh. Informasi ini dapat memberikan pengetahuan baru pada fasilitas kesehatan yaitu mengetahui gejala dan penyakit apa saja yang dapat ditimbulkan oleh penyakit hipertensi, serta dapat digunakan dalam pengaturan tata letak dan penyediaan obat bagi gejala hipertensi yang aman dikonsumsi oleh penderita hipertensi, khususnya di Klinik Rafina Medical Center.

5. DAFTAR PUSTAKA

- [1] N. Luh, A. Astari, and K. Primadewi. 2022. "Pengaruh Terapi Relaksasi Tarik Nafas Dalam Terhadap Penurunan Tekanan Darah Pada Lansia Penderita Hipertensi Di Desa Tihingan Banjarnegara Klungkung," *J. Med. Usada*, vol. 5, no. 1, p. 51.
- [2] F. G. Nonasri. 2020. "Karakteristik Dan Perilaku Mencari Pengobatan (Health Seeking Behavior) Pada Penderita Hipertensi," *J. Med. Utama*, vol. 02, no. 01, pp. 402–406.
- [3] E. Lucyana Hutahaean, M. Safii, and B. Efendi Damanik. 2020. "Implementasi algoritma apriori pada sistem persediaan barang," *JIKO (Jurnal Inform. dan Komputer)*, vol. 3, no. 3, pp. 173–180, doi: 10.33387/jiko.v3i3.2192.
- [4] M.Kavitha and D. S. Subbaiah. 2020. "Association Rule Mining using Apriori Algorithm for Extracting Product Sales Patterns in Groceries," *Int. J. Eng. Res. Technol.*, vol. 8, no. 3, pp. 5–8, [Online]. Available: www.ijert.org.
- [5] R. Takdirillah. 2020. "Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan," *Edumatic J. Pendidik. Inform.*, vol. 4, no. 1, pp. 37–46, doi: 10.29408/edumatic.v4i1.2081.
- [6] M. Tandan, Y. Acharya, S. Pokharel, and M. Timilsina. 2021. "Discovering symptom patterns of COVID-19 patients using association rule mining," *Comput. Biol. Med.*, vol. 131, no. December 2020, p. 104249, doi: 10.1016/j.compbiomed.2021.104249.
- [7] R. R. Rerung. 2018. "Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk," *J. Teknol. Rekayasa*, vol. 3, no. 1, p. 89, doi: 10.31544/jtera.v3.i1.2018.89-98.
- [8] U. Ependi and A. Putra. 2019. "Solusi Prediksi

- Persediaan Barang dengan Menggunakan Algoritma Apriori (Studi Kasus: Regional Part Depo Auto 2000 Palembang),” *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 2, p. 139, doi: 10.26418/jp.v5i2.32648.
- [9] S. Al Syahdan and A. Sindar. 2018. “Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota,” *J. Nas. Komputasi dan Teknol. Inf.*, vol. 1, no. 2, doi: 10.32672/jnkti.v1i2.771.
- [10] M. Afdal and M. Rosadi. 2019. “Penerapan Association Rule Mining Untuk Analisis Penempatan Tata Letak Buku Di Perpustakaan Menggunakan Algoritma Apriori,” *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, p. 99, doi: 10.24014/rmsi.v5i1.7379.
- [11] H. Sofyan. 2019. “Implementation of Data Mining with Association Rule in Decision Making for Product Purchase Correlation Using the Apriori Algorithm,” *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699.
- [12] C. N. Dengen, K. Kusriani, and E. T. Luthfi. 2019. “Penentuan Association Rule Pada Kelulusan Mahasiswa Menggunakan Algoritma Apriori,” *J. Rekayasa Teknol. Inf.*, vol. 3, no. 1, p. 20, doi: 10.30872/jurti.v3i1.2256.
- [13] F. A. Sianturi. 2018. “Penerapan Algoritma Apriori Untuk Penentuan Tingkat Pesanan,” *Mantik Penusa*, vol. 2, no. 1, pp. 50–57, [Online]. Available: <http://ejournal.pelitanusantara.ac.id/index.php/mantik/article/view/330>.
- [14] W. Nengsih. 2017. “Analisa Akurasi Permodelan Supervised Dan Unsupervised,” *Sebatik 1410-3737*, vol. 23, no. 2, pp. 285–291, [Online]. Available: <https://jurnal.wicida.ac.id/index.php/sebatik/article/view/771>.
- [15] A. R. Riszky and M. Sadikin. 2019. “Data Mining Menggunakan Algoritma Apriori untuk Rekomendasi Produk bagi Pelanggan,” *J. Teknol. dan Sist. Komput.*, vol. 7, no. 3, pp. 103–108, doi: 10.14710/jtsiskom.7.3.2019.103-108.
- [16] A. Setiawan and F. P. Putri. 2020. “Implementasi Algoritma Apriori untuk Rekomendasi Kombinasi Produk Penjualan,” *Ultim. J. Tek. Inform.*, vol. 12, no. 1, pp. 66–71, doi: 10.31937/ti.v12i1.1644.