

Analisis Perbandingan Kinerja Algoritma Apriori, FP-Growth dan Eclat dalam menemukan Pola Frekuensi pada Dataset INA-CBG'S

Comparative Analysis of The Performance of The Apriori, FP-Growth and Eclat Algorithms In Finding Frequency Patterns In The Ina-Cbg's Dataset

Eka Wahyu Pujiharto^{*1}, Kusrini², Asro Nasiri³

¹Program Studi PJJ Magister Teknik Informatika; Jl. Ring Road Utara, Condong Catur, Sleman, Yogyakarta, Telp: (0274) 884201-207, Fax: (0274) 884208

^{2,3}Program Studi PJJ Magister Teknik Informatika, FIK Universitas Amikom, Yogyakarta
e-mail: *eka.wahyu.pujiharto@students.amikom.ac.id, ²Kusrini@amikom.ac.id,

³Asro@amikom.ac.id

Abstrak

Setiap fasilitas kesehatan seperti rumah sakit, klinik dan puskesmas yang bekerjasama dengan BPJS wajib melakukan klaim pembayaran atas perawatan kesehatan terhadap pasien menggunakan tarif INACBG's (Indonesian - Case Based Groups). Tarif INACBG's merupakan paket layanan yang didasarkan kepada pengelompokan diagnosa penyakit yang menggunakan kode ICD-10. Penelitian ini bertujuan menemukan pola frekuensi pada dataset INA-CBG's terutama kombinasi diagnosa agar diketahui kombinasi diagnosa apa saja yang sering muncul untuk bahan evaluasi lebih lanjut oleh pihak manajemen fasilitas kesehatan. Penelitian ini membandingkan kinerja Algoritma Apriori, FP-Growth dan Eclat. Nilai Akurasi Lift Ratio dan Rule Asosiasi ketiga algoritma didapatkan nilai yang sama, tetapi untuk waktu komputasi dan pemakaian memori pada Algoritma Eclat lebih banyak daripada Algoritma Apriori dan Fp-Growth, maka dapat disimpulkan bahwa Algoritma FP-Growth dan Apriori lebih cocok untuk dijadikan solusi dalam menemukan pola frekuensi pada dataset INACBG's.

Kata kunci—Analisis Perbandingan, Apriori, FP-Growth, Eclat

Abstract

Every health facility such as hospitals, clinics and community health centers that collaborate with BPJS is required to make funding claims for health care for patients using INACBG's (Indonesian - Case Based Groups) rates. INACBG's tariff is a service package that is based on disease diagnosis groupings using ICD-10 codes. This research aims to find frequency patterns in the INA-CBG's dataset, especially combinations of diagnoses, in order to find out what combinations of diagnoses frequently appear for further evaluation by health facility management. This research compares the performance of the Apriori, FP-Growth and Eclat algorithms. The accuracy values of the Lift Ratio and Rule Association of the three algorithms obtained the same value, but the computing time and memory usage of the Eclat Algorithm is more than the Apriori and Fp-Growth Algorithms, so it can be concluded that the FP-Growth and Apriori Algorithms are more suitable to be used as solutions in found frequency patterns in INACBG's dataset.

Keywords—Comparative Analysis, Apriori, FP-Growth, Eclat

1. PENDAHULUAN

Pada bidang kesehatan, BPJS (Badan Penyelenggara Jaminan Sosial) menerapkan pola pembayaran Jaminan Kesehatan Nasional (JKN) Kesehatan menggunakan tarif INA-CBG's (Indonesia Case Based Groups) [1]. Setiap fasilitas kesehatan seperti rumah sakit, klinik dan puskesmas yang bekerjasama dengan BPJS wajib melakukan klaim pembayaran atas perawatan kesehatan terhadap pasien menggunakan tarif INACBG's (Indonesian - Case Based Groups) [2]. Tarif INACBG's berdasarkan paket pelayanan dengan cara mengelompokan diagnosis penyakit yang memiliki kelompok tarif.

Kendala yang terjadi pada fasilitas kesehatan dengan adanya sistem pembayaran klaim biaya pasien menggunakan INACBG's adalah fasilitas kesehatan mengalami kesulitan untuk menganalisa data klaim, terkait penegakan diagnosa, prosedur tindakan dan variabel lainnya, Tarif INACBGs menggunakan kode ICD-10 untuk sistem koding pada diagnosis penyakit.

Data mining adalah proses mengekstraksi informasi berguna dari database besar, mengubah dan mengekstraksinya menjadi informasi baru untuk membantu pengambilan keputusan [3]. Algoritma data mining menjadi sangat penting dalam menganalisis data untuk mengidentifikasi pola-pola kombinasi kode diagnosa dan kode prosedur ini. Pola-pola ini dapat mengungkap hubungan yang tidak terlihat secara langsung antara berbagai variabel dalam data. Algoritma untuk mengenali pola dari data yaitu algoritma asosiasi, ada tiga algoritma asosiasi yang umum digunakan diantaranya adalah Apriori, FP-Growth, dan ECLAT [4]. Perbedaan dasar dari ketiga algoritma adalah algoritma Apriori menggunakan pendekatan *generate and test* dengan cara membuat semua itemset kandidat kemudian menghitung dukungan untuk masing-masing itemset dan membuang yang tidak memenuhi ambang batas dukungan. Algoritma Eclat juga menggunakan pendekatan *generate and test*, tetapi lebih fokus pada menyusun itemset yang lebih panjang dari itemset yang lebih pendek. Sedangkan algoritma FP-growth menggunakan pendekatan *divide and conquer* yaitu dengan membangun struktur data yang disebut FP-tree dan kemudian mengeksplorasi struktur ini untuk menemukan pola asosiasi.

Telah dilakukan beberapa penelitian yang meneliti tentang algoritma asosiasi untuk menemukan pola frekuensi pada dataset, berikut ini tinjauan pustaka pada penelitian yang telah dilakukan oleh peneliti lain dan menjadi referensi dalam penelitian ini, yaitu:

- 1) Penelitian pertama oleh Idris, dkk [5]. Hasil literatur yang telah dilakukan diperoleh bahwa FP-Growth menggunakan memori yang lebih banyak daripada dua algoritma lainnya, tetapi memiliki performa terbaik. Algoritma Apriori-TID lebih cepat dan menggunakan memori yang lebih ringan.
- 2) Penelitian kedua oleh V Srinadh [6]. Hasil literatur yang telah dilakukan bahwa Fpgrowth menggunakan format data horizontal, yang mengharuskan pemindaian basis data beberapa kali dan menyebabkan waktu komputasi yang cukup lama. Eclat memiliki keunggulan dalam hal waktu komputasi. Eclat menghilangkan pemindaian basis data berulang dan mengkonsumsi waktu yang lebih sedikit daripada Apriori dan Fpgrowth tetapi sangat tergantung pada data masukan dan sumber daya yang tersedia. Jika diperhitungkan faktor lain selain waktu, hasilnya mungkin bervariasi tergantung pada faktor-faktor tersebut.
- 3) Penelitian ketiga oleh Thanathamathe dan Sawangarreera [7]. Hasil dari penelitian ini menyatakan bahwa Algoritma FP-Growth dan Eclat berhasil menemukan beberapa pola yaitu empat pola peningkatan pendapatan yang terkait kuat dengan kategori aset dan kewajiban dalam keuangan, sembilan pola penurunan pendapatan yang terkait kuat lainnya terkait dengan kewajiban, pendapatan, dan kategori beban dalam laporan keuangan, dan mengidentifikasi sepuluh akuntansi penting deskriptor yang mencerminkan laba masa depan.
- 4) Penelitian keempat oleh Syahrir, dkk [8]. Hasil dari penelitian ini adalah semakin besar dataset yang akan diolah, hasilnya akan lebih optimal jika menggunakan algoritma fp-growth RapidMiner, namun kurang maksimal jika dataset yang akan diolah merupakan

dataset kecil. Berbeda dengan menggunakan algoritma FP-growth Apriori dan Weka dimana rule yang dihasilkan kurang maksimal jika dataset yang digunakan besar dan optimal jika datasetnya kecil. Metode yang digunakan sebagai perbandingan adalah algoritma Apriori tradisional fp-growth dan TPQ-Apriori, Apriori Tradisional, FP-growth, dan TPQ-Apriori. Untuk pengujian rule, hasil rule TPQ-Apriori dibandingkan dengan algoritma Fp-Growth pada tools Rapidminer dan Weka serta algoritma Apriori pada tools Weka. Dari hasil pengujian, TPQ-Apriori mampu lebih optimal dan konsisten.

- 5) Penelitian kelima oleh Indradewi dan Putri [9]. Penelitian ini menyatakan bahwa Penerapan Association Rule menggunakan algoritma FP-Growth dengan prosedur yang telah dilakukan diperoleh 12 aturan asosiasi dimana memiliki nilai lift ratio > 1 yang digunakan sebagai acuan untuk mendapatkan pengetahuan-pengatahuan baru dari data yang diolah.
- 6) Penelitian keenam oleh Ritha, dkk [10]. Penelitian menyatakan bahwa Algoritma apriori dapat digunakan untuk menentukan aturan asosiasi di poliklinik penyakit dalam RSUD Bintan. Dalam penelitian ini, ditemukan 22 aturan asosiasi, yang memberikan ambang batas minimal support sebesar 6% dan minimal confidence sebesar 50%. Dan hanya 9 aturan asosiasi yang dapat digunakan, dan 13 aturan asosiasi yang diperoleh dari total 22 aturan asosiasi tidak dapat digunakan. Hal ini disebabkan karena informasi yang terdapat pada hasil aturan asosiasi 13 mengandung informasi yang kurang spesifik.
- 7) Penelitian ketujuh oleh Robu, dkk [11]. Penelitian ini membandingkan efisiensi algoritma Apriori dan Eclat dalam penambangan aturan asosiasi dengan mengevaluasi kinerjanya pada kumpulan data dunia nyata. Selain efisiensi kinerja, makalah ini juga mempertimbangkan faktor-faktor penting dalam lingkungan bisnis seperti kuantitas aturan yang dihasilkan, daya tarik aturan, kelemahan algoritma, dan efisiensi waktu yang diperlukan untuk menghasilkan aturan. Studi ini menjelaskan bahwa di dunia yang datanya terus dihasilkan, penting untuk mengubahnya menjadi pengetahuan yang berharga, terutama bagi perusahaan besar yang ingin memahami perilaku pelanggan. Penambangan aturan asosiasi penting untuk mengidentifikasi pola yang sering terjadi dalam kumpulan data besar, dengan fokus pada masalah Analisis Keranjang Pasar atau masalah Pembelajaran Aturan Asosiasi, yang bertujuan untuk mengidentifikasi kumpulan item yang paling sering muncul dalam kumpulan data transaksional.
- 8) Penelitian kedelapan oleh Islamiyah, dkk [12]. Makalah penelitian ini mengeksplorasi pentingnya data mining dalam menganalisis data transaksi dari minimarket untuk mendapatkan wawasan berharga tentang pola pembelian konsumen. Studi ini membandingkan algoritma Apriori dan algoritma FP-Growth untuk mengidentifikasi aturan asosiasi berdasarkan pola pembelian konsumen. Dengan menggunakan data primer dari transaksi penjualan, penelitian menemukan bahwa algoritma FP-Growth lebih cepat dan menghasilkan 10 aturan dibandingkan 11 aturan dari algoritma Apriori. Temuannya menunjukkan bahwa informasi ini dapat dimanfaatkan untuk mengelola penempatan barang dan merancang konsep pemasaran, sehingga berpotensi meningkatkan penjualan. Selain itu, penggunaan teknik data mining dan metode asosiasi di perusahaan ritel dapat membantu memahami kebutuhan konsumen dan merancang strategi pemasaran, yang pada akhirnya memungkinkan bisnis untuk mengoptimalkan penempatan barang dan stok. Makalah ini juga membahas tentang proses data mining, kelebihan algoritma FP-Growth dibandingkan algoritma Apriori, dan perbandingan waktu eksekusi kedua algoritma. Selain itu menguraikan berbagai tahapan dan langkah yang terlibat dalam penelitian menggunakan

metode CRISPDM, menekankan pentingnya memahami hubungan antara barang yang sering dibeli.

2. METODE PENELITIAN

Untuk melakukan penelitian dibutuhkan alur penelitian sebagai langkah-langkah yang harus dilakukan agar penelitian dapat berjalan dengan baik, adapun alur penelitian dari penelitian ini adalah sebagai berikut :

1. Tahap Awal

Alur dari penelitian ini berawal dari identifikasi masalah pada objek penelitian yang dilakukan serta studi pustaka yaitu dengan mengumpulkan dan menganalisa publikasi-publikasi yang telah dilakukan sebelumnya serta mengambil beberapa referensi teori dari beberapa buku yang telah terbit dan berkaitan dengan penelitian yang dilakukan sebagai rujukan untuk menentukan metode algoritma yang sesuai untuk penyelesaian permasalahan.

2. Pada tahap persiapan data, peneliti melakukan pengumpulan data klaim INACBG's pada Rumah Sakit X. Untuk data yang diambil adalah data pasien Rawat Inap selama satu tahun. Variabel yang ada pada data INACBG's adalah seperti yang tercantum pada *Tabel 1. Variabel Data INACBG's* sebagai berikut :

Tabel 1. Variabel Data INACBG's

No	Nama Variabel	Keterangan Variabel
1	KODE_RS	Kode Rumah Sakit
2	KELAS_RS	Kelas Rumah Sakit
3	KELAS_RAWAT	Kelas Perawatan
4	KODE_TARIF	Kode Tarif INACBG
5	ADMISSION_DATE	Tgl Masuk
6	DISCHARGE_DATE	Tgl Keluar
7	BIRTH_DATE	Tanggal Lahir
8	BIRTH_WEIGHT	Berat Lahir
9	SEX	Jenis Kelamin
10	DISCHARGE_STATUS	Status Pulang
11	DIAGLIST	List Diagnosa (ICD 10)
12	PROCLIST	List Prosedur (ICD 9)
13	INACBG	Kode Group Diagnosa
14	DESKRIPSI_INACBG	Keterangan Group Diagnosa
15	TARIF_INACBG	Tarif INACBG
16	TARIF_RS	Tarif Rumah Sakit
17	LOS	Length Of Stay
18	NAMA_PASIEN	Nama Pasien
19	MRN	Medical Record Number
20	UMUR_TAHUN	Umur dalam Tahun
21	UMUR_HARI	Umur dalam Hari
22	DPJP	Dokter Penanggung Jawab
23	SEP	No SEP

24	NOKARTU	No. Kartu BPJS
----	---------	----------------

Data yang sudah diambil kemudian dilakukan pre processing data yaitu melakukan pembersihan data dan transformasi data, setelah itu mengambil variabel Diaglist (list diagnose kode ICD 10) untuk dijadikan dataset yang dibutuhkan pada proses selanjutnya.

3. Penentuan Skenario

Pada tahap ini dilakukan proses pengolahan data terhadap dataset yang telah disiapkan sebelumnya, membuat model untuk masing-masing algoritma yang digunakan, dengan tahapan sebagai berikut :

- a. Mencari frequensi item untuk masing-masing item data.b.
- b. Mencari nilai support masing-masing item data.
- c. Menentukan nilai minimum support untuk memfilter data yang dibutuhkan untuk proses selanjutnya.
- d. Mencari asosiasi rule yang terbentuk dari hasil filter data.

4. Evaluasi

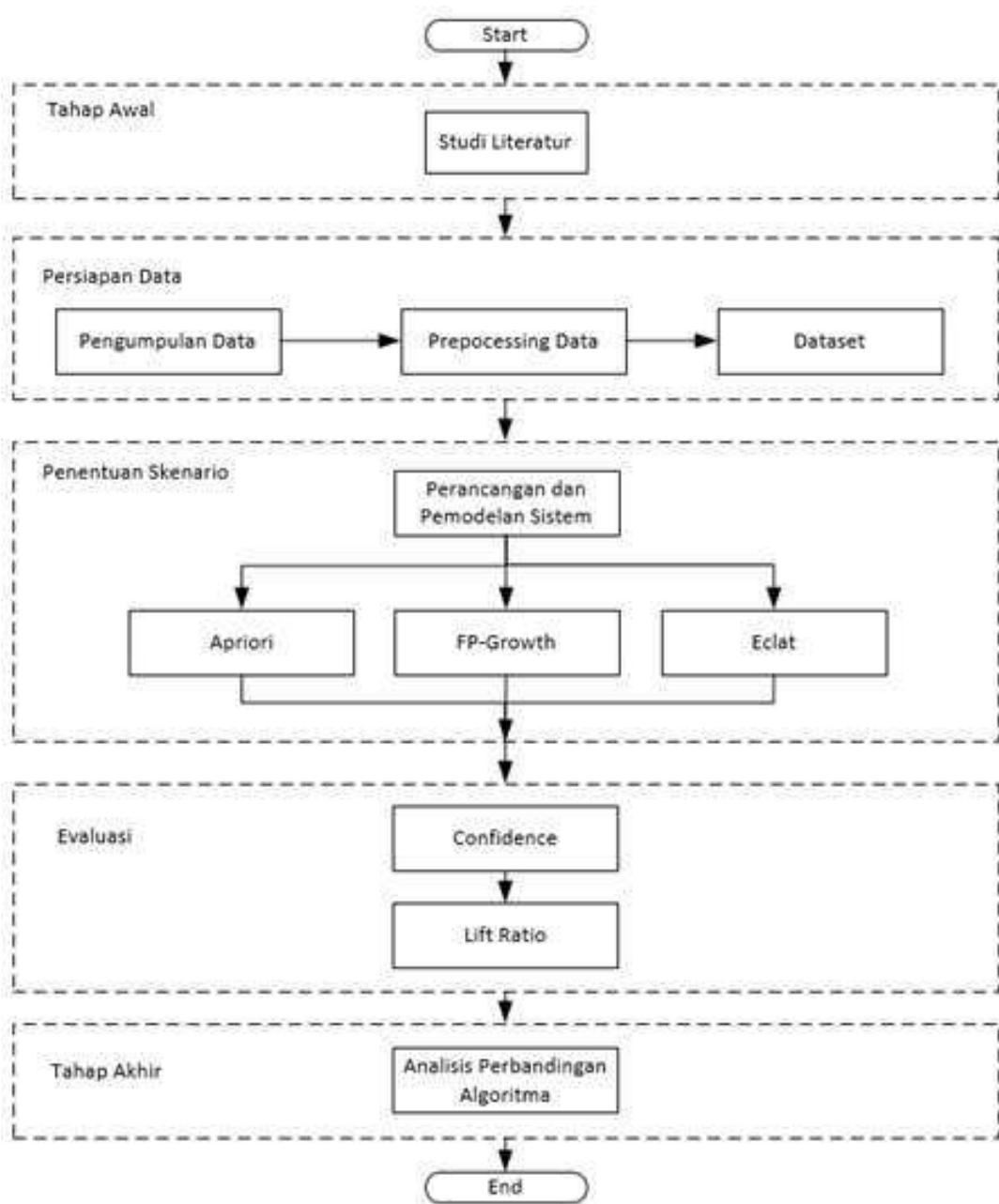
Pada tahap ini mengevaluasi asosiasi rule yang sudah ditemukan dengan mencari nilai confidence dan nilai lift ratio untuk setiap algoritma yang digunakan.

5. Tahap Akhir

Pada tahap akhir menganalisa asosiasi rule dengan tahapan :

- a. Mendeskripsikan antecedent dan consequent pada asosiasi rule yang terbentuk dalam bentuk kode ICD menjadi deskripsi diagnosa sesuai dengan Buku Kode Klasifikasi Penyakit dan Tindakan Medis ICD 10.
- b. Memilih asosiasi rule terbaik dengan nilai confidence $>60\%$ dan nilai lift ratio > 1 .
- c. Membandingkan asosiasi rule yang terbaik dari ketiga model algoritma. Hasil dari perbandingan ini maka ditemukan algoritma yang paling cocok untuk digunakan dalam menemukan pola frequensi dan aturan asosiasi pada dataset INACBG's.

Alur penelitian juga digambarkan pada *Gambar 1. Alur Penelitian* sebagai berikut :

*Gambar 1. Alur Penelitian*

3. HASIL DAN PEMBAHASAN

Proses perbandingan ini menggunakan dataset DIAGLIST (ICD10), dataset tersebut diperoleh dari data INACBG's yang memiliki 24 variabel data sebelumnya. Kami mengambil variabel DIAGLIST atau kelompok kode diagnose ICD10 sebagai dataset yang akan digunakan. Dataset DIAGLIST berukuran 14672 baris dan 1 kolom.

Tabel 2. Dataset DIAGLIST

DIAGLIST
K36
D10.5;R06.0
K85.9
R10.4
C43.9
K11.4;K04.1
S02.40;S06.6
K80.1
L02.4
D24
A09
... dst

Untuk mendapatkan aturan asosiasi pada dataset diperlukan 2 metodologi dasar dalam aturan asosiasi, yaitu analisis pola frekuensi tinggi dan pembentukan aturan asosiatif [13].

Peneliti menggunakan algoritma Apriori, FP-Growth dan ECLAT dengan memakai bahasa pemrograman Python untuk implementasinya.

Algoritma Apriori digunakan untuk menghitung jumlah aturan asosiasi antar item. Aturan asosiasi mendefinisikan bagaimana dua atau lebih item terkait satu sama lain [14]. Itemset yang sering muncul dipilih untuk menentukan aturan asosiasi.

Algoritma Apriori memiliki langkah-langkah sebagai berikut :

1. Menentukan support dari itemset (dengan nilai k = 1) pada database. Tahap ini menghasilkan himpunan kandidat.
2. Jika ditemukan support lebih kecil dari ambang batas (threshold) yang diberikan maka lakukan pemangkasan (pruning) kumpulan kandidat.
3. Buat himpunan ukuran k + 1 dengan menggabungkan himpunan item yang paling sering muncul., dan ulangi himpunan di atas sampai dengan itemset tidak dapat dibentuk kembali.

Algoritma Apriori memiliki kelemahan karena harus melakukan scan database setiap kali iterasi, sehingga waktu yang dibutuhkan semakin bertambah iterasi. Masalah ini diselesaikan dengan algoritma baru seperti FP-Growth [15].

Algoritma FP-Growth merupakan algoritma asosiasi yang dapat digunakan untuk mencari himpunan data yang paling sering muncul (frequent itemset) pada sebuah kumpulan data [16].

Algoritma FP-Growth memiliki langkah-langkah sebagai berikut :

1. Sediakan set data
2. Pengurutan didasarkan pada nilai frekuensi tertinggi dari item yang muncul
3. Membentuk FP-tree
4. Membentuk Conditional Pattern Base
5. Membentuk Conditional FP-tree
6. Membentuk Frequent Pattern Generated
7. Mencari frequency 2 Itemset
8. Mencari Support 2 Itemset dan Mencari Confidance 2 Itemset.

Algoritma FP-Growth memerlukan bantuan FP-Tree yang bertujuan untuk menemukan frequent itemset. Dengan menggunakan FP-Tree, Algoritma FP-Growth dapat langsung mengekstrak frequent itemset dari FP-Tree [17].

Algoritma ECLAT ditemukan oleh M. J. Zaki tahun 1997. Algoritma ECLAT melakukan pencarian item mulai dari yang sering muncul hingga item yang jarang muncul dengan dilakukan

tanpa memperhatikan urutan apapun, Proses pemindaian dilakukan satu kali tanpa pengulangan [18].

Algoritma ECLAT memiliki langkah-langkah sebagai berikut :

1. Langkah pertama adalah mengidentifikasi semua item dalam dataset dan membuat itemset awal yang hanya mengandung satu item.
2. Selanjutnya, algoritma akan menghitung dukungan (support) untuk setiap set item. dukungan (support) adalah jumlah transaksi yang mencakup itemset tersebut dibagi dengan jumlah total transaksi.
3. Setelah itu, itemset-itemset yang memiliki dukungan (support) di atas batas minimum akan dipertahankan, sedangkan itemset-itemset yang tidak memiliki dukungan akan dihapus.
4. Algoritma kemudian akan membentuk itemset baru yang lebih panjang dengan menggabungkan itemset yang dipertahankan.
5. Proses penggabungan itemset-itemset dilakukan secara berulang hingga tidak ada lagi itemset-itemset baru yang terbentuk.
6. Terakhir, algoritma ini akan menghasilkan semua itemset yang memenuhi batas minimum dukungan (support).

3.1. Analisis Pola Frekuensi Tinggi

Pada tahap ini dicari kombinasi item yang memenuhi syarat minimum dari nilai support pada database. Persentase item atau kombinasi item yang ada pada database disebut nilai support (penunjang).

Untuk menghitung frequensi dataset maka perlu membuat matrix itemset untuk menampilkan kemunculan kode ICD10 pada setiap transaksi seperti pada *Gambar 2. Matrix* dibawah ini :

	A010	A014	A049	A059	A069	A070	A084	A090	A099	A150	...	Z939	Z948	Z954	Z955	Z960	Z961	Z978	Z982	Z988	Z992
0	False	...	False																		
1	False	...	False																		
2	False	...	False																		
3	False	...	False																		
4	False	...	False																		
...
14667	False	...	False																		
14668	False	...	False																		
14669	False	...	False																		
14670	False	...	False																		
14671	False	...	False																		
14672 rows × 2151 columns																					

Gambar 2. Matrix Itemset

Pada *Gambar 2. Matrix Itemset* diatas diketahui ukuran dari dataset DIAGLIST memiliki 14672 baris dan 2151 kolom.

Langkah selanjutnya menghitung frequensi dataset dengan menghitung nilai support tiap-tiap itemset menggunakan algoritma apriori, fp-growth dan eclat. Berikut ini cara mencari nilai support sebuah item dengan rumus persamaan sebagai berikut :

$$\text{Support } (A) = \frac{\text{Jumlah Transaksi Untuk } A}{\text{Total Transaksi}} \quad (1)$$

Rumus nilai support pada rumus persamaan (1) menjelaskan bahwa nilai support adalah hasil bagi dari jumlah transaksi yang mengandung item A (satu item) dengan total jumlah keseluruhan transaksi.

Sedangkan untuk nilai support dari 2 item diperoleh dari rumus persamaan sebagai berikut

$$\text{Support } (A, B) = \frac{\text{Jumlah Transaksi untuk } A \text{ dan } B}{\text{Total Transaksi}} \quad (2)$$

Rumus persamaan (2) menjelaskan bahwa nilai support 2-itemsets diperoleh dengan cara membagi jumlah transaksi yang mengandung item A dan item B dengan total jumlah keseluruhan transaksi. Hasil dari perhitungan nilai support pada itemset terdapat pada tabel 3, sebagai berikut:

Tabel 3. Hasil Perhitungan Nilai Support pada Itemset

Algoritma	Support (\geq)	Jumlah Itemset N=1	Jumlah Itemset N=2	Jumlah Itemset N=3
Apriori	0.001 – 0.007	84	30	2
	0.008	67	20	1
	0.009	84	17	-
	0.01	84	12	-
	0.02	25	1	-
FP-Growth	0.001 – 0.007	56	12	-
	0.008	56	12	-
	0.009	56	12	-
	0.01	56	12	-
	0.02	51	1	-
Eclat	0.001	395	546	-
	0.009	58	17	-
	0.01	56	12	-
	0.02	25	1	-

Dari hasil perhitungan pada

Tabel 3. Hasil Perhitungan Nilai Support pada Itemset didapatkan beberapa kesimpulan yaitu sebagai berikut :

1. Semakin kecil nilai support maka dihasilkan jumlah itemset lebih banyak
2. Pada jumlah itemset N=1 dan nilai support 0.001 didapatkan Jumlah Itemset terbanyak adalah pada perhitungan dataset menggunakan algoritma Eclat dengan jumlah itemset sebanyak 395.
3. Jumlah itemset terbanyak pada itemset N=2 sebanyak 30 yang diperoleh dari perhitungan menggunakan algoritma Apriori dengan menggunakan nilai support 0.001.
4. Jumlah itemset N=3 didapatkan pada perhitungan menggunakan algoritma apriori dengan nilai support 0.001 sebanyak 2 itemset dan pada nilai support 0.008 sebanyak 1 itemset.

5. Perhitungan ketiga algoritma menghasilkan jumlah itemset yang sama saat nilai support 0.01 mendapatkan 12 itemset, dan pada saat nilai suppot 0.02 mendapatkan 1 itemset.

3.2. Pembentukan Aturan Asosiatif

Setelah semua pola frekuensi tinggi telah diidentifikasi, tahap selanjutnya adalah dicari aturan asosiasi yang memenuhi syarat minimum untuk confidence dengan cara menghitung nilai confidence aturan assosiatif A B.

Rumus persamaan di bawah ini digunakan untuk menghitung nilai confidence aturan A B sebagai berikut :

$$\text{Confidence } (A, B) = \frac{\text{Jumlah Transaksi untuk } A \text{ dan } B}{\text{Jumlah Transaksi } A} \quad (3)$$

Nilai confidence dihitung menggunakan rumus persamaan (3) yaitu hasil bagi jumlah transaksi yang mengandung item A dan item B dengan jumlah total transaksi yang hanya mengandung item A, seperti terlihat pada rumus di atas. Aturan asosiasi harus diurutkan menurut Support × Confidence untuk memilihnya. kemudian aturan diambil sebanyak n aturan yang memiliki hasil terbesar.

Lift Ratio digunakan untuk menghitung kekuatan aturan asosiasi (association rule) yang telah terbentuk dari nilai support dan confidence. Nilai lift ratio juga digunakan sebagai penentu apakah aturan asosiasi valid atau tidak valid [19].

Antecedent dan consequent adalah dua item yang saling mempengaruhi sebab akibat dari kedua item tersebut. Nilai dari expected confidence dihitung dengan rumus persamaan sebagai berikut :

$$\text{Expected Confidence} = \frac{\text{Jumlah Transaksi mengandung } B}{\text{Jumlah Transaksi}} \quad (4)$$

Ada 2 cara untuk menghitung Lift ratio yaitu sebagai berikut :

1. Membandingkan antara confidence untuk suatu aturan dibagi dengan expected confidence[9]. dengan rumus persamaannya sebagai berikut :

$$\text{Lift Ratio} = \frac{\text{Confidence}}{\text{Expected Confidence}} \quad (5)$$

2. Membandingkan nilai support A B dibagi dengan perkalian antara nilai support A dan nilai support B, dengan rumus persamaan sebagai berikut :

$$\text{Lift Ratio} = \frac{\text{Support } (A,B)}{(\text{Support } (A)*\text{Support } (B))} \quad (6)$$

Penulis akan menggunakan rumus persamaan (6) untuk menghitung lift ratio dalam penelitian ini. Jika nilai Lift Ratio lebih besar dari 1 mengandung arti adanya manfaat dari aturan tersebut. Jika lebih tinggi nilai Lift ratio, maka lebih besar kekuatan asosiasinya [20].

Dari tabel 3 kita pilih nilai support 0.01 untuk menampilkan aturan asosiasi yang terbentuk dengan menentukan nilai confidence > 0.6

Hasil dari perhitungan ketiga algoritma menghasilkan aturan asosiasi sebanyak 7 itemset dengan komposisi kode ICD yang sama, adapun hasilnya sebagai berikut :

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
18	(O821)	(Z370)	0.027331	0.047369	0.027058	0.990025	20.900210	0.025764	95.501244	0.978908
17	(O809)	(Z370)	0.014449	0.047369	0.013768	0.952830	20.114999	0.013083	20.195774	0.964218
14	(O429)	(Z370)	0.012336	0.047369	0.010564	0.856354	18.078302	0.009980	6.631776	0.956485
3	(C20)	(Z511)	0.017107	0.034283	0.012745	0.745020	21.731476	0.012159	3.787421	0.970588
23	(P929)	(P599)	0.014177	0.019766	0.010087	0.711538	35.998939	0.009807	3.398146	0.986202
7	(E870)	(J969)	0.014449	0.031625	0.010087	0.698113	22.074821	0.009630	3.207743	0.968696
21	(P071)	(P599)	0.016562	0.019766	0.010019	0.604938	30.605705	0.009692	2.481218	0.983617

Gambar 3. Hasil Aturan Asosiasi Menggunakan Algoritma Apriori

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
20	(O821)	(Z370)	0.027331	0.047369	0.027058	0.990025	20.900210	0.025764	95.501244	0.978908
12	(O809)	(Z370)	0.014449	0.047369	0.013768	0.952830	20.114999	0.013083	20.195774	0.964218
16	(O429)	(Z370)	0.012336	0.047369	0.010564	0.856354	18.078302	0.009980	6.631776	0.956485
23	(C20)	(Z511)	0.017107	0.034283	0.012745	0.745020	21.731476	0.012159	3.787421	0.970588
2	(P929)	(P599)	0.014177	0.019766	0.010087	0.711538	35.998939	0.009807	3.398146	0.986202
11	(E870)	(J969)	0.014449	0.031625	0.010087	0.698113	22.074821	0.009630	3.207743	0.968696
0	(P071)	(P599)	0.016562	0.019766	0.010019	0.604938	30.605705	0.009692	2.481218	0.983617

Gambar 4. Hasil Aturan Asosiasi Menggunakan Algoritma FP-Growth

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
6	(O821)	(Z370)	0.027331	0.047369	0.027058	0.990025	20.900210	0.025764	95.501244	0.978908
5	(O809)	(Z370)	0.014449	0.047369	0.013768	0.952830	20.114999	0.013083	20.195774	0.964218
1	(O429)	(Z370)	0.012336	0.047369	0.010564	0.856354	18.078302	0.009980	6.631776	0.956485
0	(C20)	(Z511)	0.017107	0.034283	0.012745	0.745020	21.731476	0.012159	3.787421	0.970588
4	(P929)	(P599)	0.014177	0.019766	0.010087	0.711538	35.998939	0.009807	3.398146	0.986202
3	(E870)	(J969)	0.014449	0.031625	0.010087	0.698113	22.074821	0.009630	3.207743	0.968696
2	(P071)	(P599)	0.016562	0.019766	0.010019	0.604938	30.605705	0.009692	2.481218	0.983617

Gambar 5. Hasil Aturan Asosiasi Menggunakan Algoritma Eclat

Setelah hasil aturan asosiasi yang terbentuk selanjutnya mendeskripsikan nilai antecedent dan consequent dalam bentuk kode ICD menjadi deskripsi diagnosa sesuai dengan Buku Kode Klasifikasi Diagnosa Penyakit ICD 10. menjadi sebagai berikut :

1. O821 : Delivery by emergency caesarean section
2. Z370 : Single live birth
3. O809 : Single spontaneous delivery
4. O429 : Premature rupture of membranes
5. C20 : Malignant neoplasm of rectum
6. Z511 : Encounter for antineoplastic chemotherapy and immunotherapy
7. P929 : Feeding problem of newborn
8. P599 : Neonatal jaundice
9. E870 : Hyperosmolality and hypernatraemia
10. J969 : Respiratory failure
11. P071 : Other low birth weight

Aturan asosiasi dengan nilai support 0.01 dan nilai confidence > 0.6, sebagai berikut :

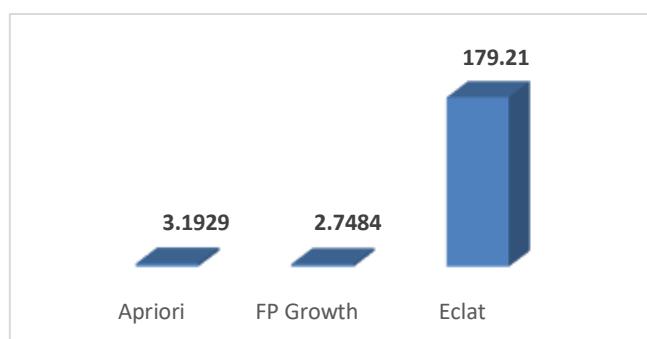
1. (O821) dan (Z370) dengan nilai confidence 0.99 dan nilai lift ratio 20.90
Jika pasien melahirkan dengan tindakan operasi caesar sebagian besar melahirkan 1 bayi
2. (O809) dan (Z370) dengan nilai confidence 0.95 dan nilai lift ratio 20.11
Jika pasien melahirkan dengan normal sebagian besar melahirkan 1 bayi
3. (O429) dan (Z370) dengan nilai confidence 0.85 dan nilai lift ratio 18.07
Jika pasien melahirkan bayi premature sebagian besar melahirkan 1 bayi
4. (C20) dan (Z511) dengan nilai confidence 0.74 dan nilai lift ratio 21.73
Jika pasien didiagnosa tumor sebagian besar dilakukan kemoterapi
5. (P929) dan (P599) dengan nilai confidence 0.71 dan nilai lift ratio 35.99
Jika pasien bayi baru lahir susah menelan sebagian besar terkena juga penyakit kuning
6. (E870) dan (J969) dengan nilai confidence 0.69 dan nilai lift ratio 22.07
Jika pasien didiagnosa Hiperosmolalitas dan hypernatremia biasanya mengalami gagal nafas
7. (P071) dan (P599) dengan nilai confidence 0.60 dan nilai lift ratio 30.60
Jika pasien bayi lahir dengan berat badan rendah sebagian besar terkena juga penyakit kuning

Aturan asosiasi dengan nilai support 0.02 dan nilai confidence > 0.6 adalah sebagai berikut :

1. (O821) dan (Z370) dengan nilai confidence 0.99 dan nilai lift ratio 20.90
Jika pasien melahirkan dengan tindakan operasi caesar sebagian besar melahirkan 1 bayi
- ### 3.3 Perbandingan Waktu Proses dan Kapasitas Memori

Untuk mendapatkan perbandingan dari kinerja ketiga algoritma yang digunakan, maka peneliti menghitung waktu proses dan kapasitas memori yang digunakan untuk melakukan perhitungan dalam pembentukan aturan asosiasi.

Perbandingan waktu proses dilakukan menggunakan *library time* pada python, proses yang dilakukan adalah mencari aturan asosiasi dengan nilai support 0.01 dan confidence 0.6 didapatkan hasil sebagai berikut :

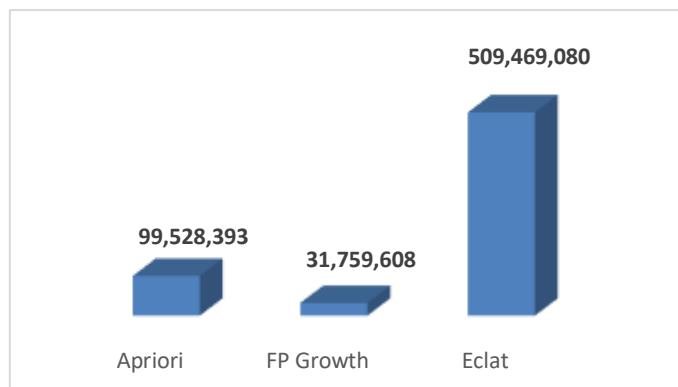


Gambar 6. Grafik Waktu Proses Algoritma

Pada Gambar 6. Grafik Waktu Proses Algoritma dapat disimpulkan bahwa untuk waktu proses perhitungan pembentukan aturan asosiasi dengan menggunakan algoritma Eclat membutuhkan waktu paling lama yaitu 179,21 detik , sedangkan FP-Growth adalah algoritma yang paling cepat prosesnya yaitu 2.74 detik.

Penggunaan memori dihitung dengan menggunakan *library tracemalloc* pada python, proses yang dilakukan adalah mencari aturan asosiasi dengan nilai support 0.01 dan confidence 0.6 setelah

diuji maka didapatkan perbedaan kapasitas penggunaan memori dari ketiga algoritma sebagai berikut :



Gambar 7. Grafik Penggunaan Memori pada Proses Algoritma

Pada *Gambar 7. Grafik Penggunaan Memori pada Proses Algoritma* dapat dilihat bahwa algoritma ECLAT membutuhkan kapasitas memori yang paling besar jika dibandingkan dengan kapasitas memori yang dibutuhkan algoritma apriori dan FP Growth untuk melakukan proses perhitungan pembentukan aturan asosiasi tersebut.

4. KESIMPULAN

Data yang digunakan dalam penelitian ini adalah data INACBG's kemudian dipilih satu variabel bernama DIAGLIST yang berisi kombinasi kode diagnose ICD10 untuk dijadikan sebuah dataset, ukuran dataset DIAGLIST memiliki 14672 baris dan 2151 kolom. Hasil dari proses pembentukan aturan asosiasi oleh algoritma Apriori, FP Growth dan ECLAT pada dataset DIAGLIST, dengan menentukan nilai support 0.01 dan confidence > 0.6 didapatkan hasil yang sama yaitu memperoleh 7 aturan asosiasi, adapun deskripsi dari aturan asosiasi tersebut adalah "Jika pasien melahirkan dengan tindakan operasi caesar maka sebagian besar melahirkan 1 bayi", "Jika pasien melahirkan dengan normal maka sebagian besar melahirkan 1 bayi", "Jika pasien melahirkan bayi premature maka sebagian besar melahirkan 1 bayi", "Jika pasien didiagnosa tumor maka sebagian besar dilakukan kemoterapi", "Jika pasien bayi baru lahir mengalami susah menelan maka sebagian besar terkena juga penyakit kuning", "Jika pasien didiagnosa Hiperosmolalitas dan hypernatremia maka biasanya mengalami gagal nafas", "Jika pasien bayi lahir dengan berat badan rendah maka sebagian besar terkena juga penyakit kuning". Aturan asosiasi yang dihasilkan dari frequent itemset dapat dipakai sebagai pendukung keputusan bagi fasilitas kesehatan dalam melakukan persiapan pelayanan yang berkaitan antara dua penyakit, meningkatkan pemantauan pelayanan pada kasus-kasus penyakit yang sering terjadi, dan meningkatkan persiapan obat, alat kesehatan dan sarana prasarana yang mendukung pelayanan tersebut. Hasil perbandingan dari waktu proses perhitungan pembentukan aturan asosiasi dengan menggunakan algoritma Eclat membutuhkan waktu paling lama yaitu 179,21 detik , sedangkan FP-Growth adalah algoritma yang paling cepat waktu prosesnya yaitu 2,7484 detik. Dalam proses pembentukan aturan asosiasi Algoritma ECLAT membutuhkan kapasitas memori yang paling besar jika dibandingkan dengan kapasitas memori yang dibutuhkan Algoritma Apriori dan FP-Growth untuk melakukan proses perhitungan pembentukan aturan asosiasi tersebut. Maka dari hasil analisis ketiga perbandingan tadi peneliti menyimpulkan bahwa Algoritma FP-Growth adalah algoritma yang paling cocok dan efisien dalam melakukan proses pembentukan aturan asosiasi untuk dataset DIAGLIST pada database INACBG;s.

5. SARAN

Penelitian ini hanya mengungkap hubungan antara kombinasi kode diagnose ICD10 pada variabel DIAGLIST di database INACBG's, penelitian ini dapat dikembangkan lagi dengan mengungkap semua variabel-variabel lain yang dimiliki database INACBG's. Belum banyak yang meneliti bidang datascience yang menggunakan database INACBG's sehingga masih dapat dikembangkan lagi pada penelitian-penelitian selanjutnya.

DAFTAR PUSTAKA

- [1] Kementerian Kesehatan Republik Indonesia, "INA-CBG'S untuk Optimalkan Pelayanan BPJS Kesehatan," 2014, [Online]. Available: <https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20140108/539538/ina-cbgs-untuk-optimalkan-pelayanan-bpjks-kesehatan/>
- [2] Peraturan Presiden Republik Indonesia Nomor 12, "Tentang Jaminan Kesehatan," 2013.
- [3] J. Suntoro, "Data Mining : Algoritma dan Implementasi dengan Pemrograman PHP," vol. PT Elex Me, 2019.
- [4] A. A. Khairul Amri, Alwis Nazir, Elin Haerani, Muhammad Affandes, Reski Mai Candra, "Penerapan Data Mining Dalam Mencari Pola Asosiasi Data Tracer Study Menggunakan Equivalence Class Transformation (ECLAT)," no. Tek. Inform. UIN Sultan Syarif Kasim Riau, 2022.
- [5] E. S. M. Andi Ilhamsyah Idris, Eliyah A M Sampetoding, Valian Yoga Pudya Ardhana, Irene Maritsa, Adrisumatri Sakri, Hidayatullah Ruslan, "Comparison of Apriori, Apriori-TID and FP-Growth Algorithms in Market Basket Analysis at Grocery Stores," vol. Hasanuddin, 2022, doi: DOI 10.30865/ijics.v6i2.4535.
- [6] Dr. V Srinadh, "Evaluation of Apriori, FP growth and Eclat association rule mining algorithms," *Int. J. Health Sci. (Qassim)*., doi: <https://doi.org/10.53730/ijhs.v6nS2.6729>.
- [7] S. S. Putthiporn Thanathamathe, "Discovering Future Earnings Patterns through FP-Growth and ECLAT Algorithms with Optimized Discretization," *Emerg. Sci. J.*, no. ISSN 2610-9182, 2022.
- [8] L. Z. A. M. Moch Syahrir, "Determination of the best rule-based analysis results from the comparison of the Fp-Growth, Apriori, and TPQ-Apriori Algorithms for recommendation systems," *Bumigora Univ. Indones. J. Manaj. Teknol. dan Inform.*, 2023, doi: <http://dx.doi.org/10.31940/matrix.v13i2.52-67>.
- [9] T. A. P. Karina Intan Indradewi, "Penerapan Association Rule Dalam Bidang Kesehatan Pada Rumah Sakit X," *Univ. Islam Indones. Semin. Nas. IENACO*, no. ISSN : 2337-4349, 2019.
- [10] W. P. Nola Ritha, Eka Suswaini, "Penerapan Association Rule Menggunakan Algoritma Apriori Pada Poliklinik Penyakit Dalam (Studi Kasus: Rumah Sakit Umum Daerah Bintan)," *J. Sains dan Inform.*, no. e-ISSN: 2598-5841.
- [11] V. Robu and V. D. Dos Santos, "Mining frequent patterns in data using apriori and eclat:

- A comparison of the algorithm performance and association rule generation," *2019 6th Int. Conf. Syst. Informatics, ICSAI 2019*, no. Icsai, pp. 1478–1481, 2019, doi: 10.1109/ICSAI48974.2019.9010367.
- [12] Islamiyah, P. L. Ginting, N. Dengen, and M. Taruk, "Comparison of Priori and FP-Growth Algorithms in Determining Association Rules," *ICEEIE 2019 - Int. Conf. Electr. Electron. Inf. Eng. Emerg. Innov. Technol. Sustain. Futur.*, pp. 320–323, 2019, doi: 10.1109/ICEEIE47180.2019.8981438.
- [13] luthfi taufiq E. Kusrini, "Algoritma Data Mining," no. Penerbit Andi, Yogyakarta, 2009.
- [14] trivusi.web.id, "Algoritma Apriori: Pengertian, Cara Kerja, Kelebihan, dan Kekurangannya," 2022, [Online]. Available: <https://www.trivusi.web.id/2022/08/algoritma-apriori.html>
- [15] Jeff Heaton, "Comparing dataset characteristics that favor the Apriori Eclat or FP-Growth frequent itemset mining algorithms," *SECON IEEE*, 2016.
- [16] D. Samuel, "Penerapan Stuktur FP-Tree dan Algoritma FP-Growth dalam Optimasi Penentuan Frequent Itemset," no. Program Studi Teknik Informatika, Sekolah Teknik Elektro dan Informatika, Institut Teknologi Bandung., 2008.
- [17] R. T. Kabir S, Ripon S, Rahman M, "Knowledge-based data mining using the semantic web. International Conference on Applied Computing," *Comput. Sci. Comput. Eng.*, 2014.
- [18] R. Mustofa and Irhamah, "Topic Discovery pada Jurnal-jurnal di IEEE Explore menggunakan Association Rule Mining dengan Pendekatan Closed Frequent Itemset," *Ejurnal.Its.Ac.Id*, vol. .8, no. 2, 2019.
- [19] M. Fauzy, K. R. S. W, and I. Asror, "Penerapan Metode Association Rule Menggunakan Algoritma Apriori pada Simulasi Prediksi Hujan Wilayah Kota Bandung," vol. 3Fakultas, 2015.
- [20] B. Santosa, "Data Mining Teori dan Aplikasi : Teknik Pemanfaatan Data untuk keperluan Bisnis," *Yogyakarta Graha Ilmu*, p. 2007, 2007.