

BAB 1 PENDAHULUAN

1.1 Latar Belakang Masalah

Penelitian ini berfokus pada pola komorbiditas pasien diabetes melitus tipe 2 (DMT2) dalam basis data MIMIC-IV v3.1. Fokus ini penting karena DMT2 jarang muncul sebagai penyakit tunggal. Pada banyak pasien, DMT2 hadir bersama penyakit kardiovaskular, gangguan ginjal, gangguan metabolik, infeksi, dan kondisi klinis lain yang dapat memperberat prognosis pasien.

DMT2 menjadi salah satu masalah kesehatan global yang terus meningkat. Prevalensi diabetes global naik dari 129 juta kasus pada tahun 1990 menjadi lebih dari 500 juta kasus pada tahun 2021, dengan proyeksi lebih dari 1,3 miliar kasus pada tahun 2050 [1, 2]. DMT2 menyumbang lebih dari 96% dari seluruh kasus diabetes dan berkaitan dengan peningkatan mortalitas, disabilitas, serta kebutuhan layanan kesehatan yang tinggi [1]. Beban ini menjadi lebih kompleks ketika pasien membutuhkan perawatan rumah sakit atau perawatan intensif.

Pada konteks klinis, pasien DMT2 sering memiliki lebih dari satu penyakit penyerta. Studi epidemiologi menunjukkan bahwa sebagian besar pasien DMT2 memiliki sedikitnya satu kondisi kronis lain, dan sebagian pasien memiliki tiga komorbiditas atau lebih [3]. Komorbiditas yang sering muncul meliputi hipertensi (I10), dislipidemia (E78), penyakit ginjal kronis (N18), gagal jantung (I50), dan gangguan metabolik lain [4]. Kombinasi penyakit tersebut membentuk pola klinis yang tidak cukup dijelaskan melalui daftar diagnosis tunggal.

Pendekatan konvensional dalam penelitian komorbiditas sering menggunakan ukuran agregat seperti *Charlson Comorbidity Index* (CCI). Ukuran ini berguna untuk menggambarkan beban komorbiditas total, tetapi tidak menunjukkan hubungan kemunculan bersama antardiagnosis secara eksplisit [5, 3]. Akibatnya, pola seperti keterkaitan antara penyakit ginjal kronis, gagal jantung, dan penyakit hipertensi dengan komplikasi ginjal tidak terlihat secara langsung. Padahal, pola seperti ini dapat membantu peneliti memahami struktur komorbiditas yang membedakan satu kelompok pasien dari kelompok lain.

Association rule mining (ARM) menawarkan pendekatan yang lebih sesuai untuk menemukan pola kemunculan bersama antarkode diagnosis. ARM dapat mengukur kekuatan aturan asosiasi melalui *support*, *confidence*, dan *lift*. Dalam penelitian kesehatan, ukuran tersebut membantu menilai seberapa

sering suatu kombinasi diagnosis muncul, seberapa kuat hubungan kondisional antaritem, dan seberapa besar asosiasi tersebut dibandingkan dengan ekspektasi independen. Namun, ARM tidak membuktikan hubungan sebab-akibat. ARM hanya menunjukkan pola asosiasi yang perlu ditafsirkan bersama konteks klinis dan analisis statistik tambahan.

Beberapa studi telah menggunakan ARM untuk mempelajari komorbiditas pada pasien diabetes atau populasi klinis lain. Namun, banyak studi masih memakai algoritma Apriori yang membutuhkan pemindaian basis data secara berulang dan dapat kurang efisien pada *dataset* besar [6, 7]. Algoritma FP-Growth (*Frequent Pattern Growth*) dapat mengatasi sebagian keterbatasan tersebut karena membangun struktur *FP-tree* untuk menemukan *frequent itemset* secara lebih efisien. Meski demikian, penerapan FP-Growth pada data ICD-10 berskala besar dari MIMIC-IV v3.1 dan pengaitannya dengan luaran klinis pasien DMT2 masih terbatas.

Keterbatasan lain dalam studi terdahulu terletak pada pemisahan antara penemuan pola komorbiditas dan analisis prognosis. Studi ARM dapat menemukan pasangan atau kombinasi diagnosis yang sering muncul, tetapi sering tidak menguji apakah pola tersebut berhubungan dengan perbedaan *survival*. Sebaliknya, studi *survival* pada pasien diabetes sering menggunakan satu indeks komorbiditas atau satu biomarker sebagai prediktor utama, tanpa memanfaatkan pola asosiasi antardiagnosis yang muncul dari data [3, 8]. Kondisi ini menciptakan kebutuhan untuk menggabungkan ARM, pembentukan kluster klinis, dan analisis *survival* dalam satu alur kerja yang koheren.

MIMIC-IV v3.1 menyediakan peluang untuk menjawab kebutuhan tersebut. Basis data ini mencakup data pasien, admisi, diagnosis ICD-10, data laboratorium, perawatan intensif, dan informasi mortalitas dari Beth Israel Deaconess Medical Center, Boston, Massachusetts [9]. Struktur data tersebut memungkinkan penelitian berbasis *real-world data* untuk membangun profil komorbiditas pasien DMT2, menambang aturan asosiasi, membentuk kluster klinis, serta menguji hubungan kluster dengan mortalitas melalui Kaplan-Meier dan regresi Cox.

Berdasarkan latar tersebut, penelitian ini mengembangkan alur kerja analitik yang mengintegrasikan tiga komponen utama. Komponen pertama ialah identifikasi pola asosiasi komorbiditas berbasis kode ICD-10 menggunakan FP-Growth. Komponen kedua ialah pembentukan kluster klinis berdasarkan aturan asosiasi yang kuat dan bermakna secara klinis. Komponen ketiga ialah evaluasi prognostik menggunakan analisis *survival* dan regresi Cox, disertai pemetaan profil keparahan

klinis melalui CCI, skor SOFA, biomarker laboratorium, dan pola medikasi intensif.

Kesenjangan utama yang menjadi dasar penelitian ini terletak pada belum terpadunya penambangan asosiasi komorbiditas, pembentukan kluster klinis, dan evaluasi prognosis pada pasien DMT2 dalam MIMIC-IV v3.1. Penelitian terdahulu telah memberi kontribusi penting dalam analisis pola penyakit, analisis komorbiditas, maupun analisis *survival*. Namun, sebagian besar masih memisahkan penemuan pola asosiasi dari evaluasi mortalitas dan profil keparahan klinis. Rincian penelitian terdahulu dan kesenjangan penelitian disajikan secara khusus pada Bab 2.

Berdasarkan kesenjangan tersebut, penelitian ini menempatkan ARM sebagai metode eksploratif untuk menemukan pola asosiasi, bukan sebagai alat inferensi kausal. Hasil ARM kemudian dievaluasi melalui ukuran kuantitatif, interpretasi domain klinis, dan analisis *survival*. Dengan posisi metodologis ini, penelitian dapat memberikan gambaran yang lebih sistematis tentang pola komorbiditas pasien DMT2 dan hubungan prognostiknya tanpa menyatakan hubungan sebab-akibat yang tidak dapat dibuktikan oleh desain penelitian ini.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, penelitian ini merumuskan masalah sebagai berikut:

1. Bagaimana pola asosiasi komorbiditas yang paling dominan pada pasien DMT2 berdasarkan kode ICD-10 dalam MIMIC-IV v3.1 jika diukur menggunakan *support, confidence, dan lift*?
2. Kluster komorbiditas apa saja yang dapat diidentifikasi dari aturan asosiasi yang kuat dan bermakna secara klinis, serta bagaimana hubungan kluster tersebut dengan perbedaan *survival* dan risiko mortalitas pasien?
3. Bagaimana profil keparahan klinis pada setiap kluster komorbiditas berdasarkan CCI, skor SOFA pada subpopulasi ICU, biomarker laboratorium, dan pola penggunaan medikasi intensif?

1.3 Batasan Permasalahan

Untuk menjaga fokus penelitian, penelitian ini menetapkan batasan sebagai berikut:

1. Sumber data penelitian ialah MIMIC-IV v3.1 yang diakses melalui Google BigQuery, khususnya *dataset path* `physionet-data.mimiciv_3_1_hosp` dan `physionet-data.mimiciv_3_1_icu`. Data berasal dari satu pusat layanan kesehatan, yaitu Beth Israel Deaconess Medical Center, Boston, Massachusetts, Amerika Serikat. Karena itu, generalisasi hasil ke rumah sakit lain perlu dilakukan secara hati-hati.
2. Populasi penelitian mencakup pasien dengan diagnosis DMT2 berdasarkan kode ICD-10 E11.x pada sedikitnya satu admisi. Penelitian tidak memasukkan pasien dengan diabetes tipe 1 (E10.x) dan diabetes tidak terspesifikasi (E14.x).
3. Profil komorbiditas dibangun dari kode diagnosis ICD-10 tiga digit. Penelitian mengecualikan kode faktor eksternal (Z), kode penyebab tambahan (Y), dan kode gejala tidak terspesifikasi (R) agar analisis berfokus pada diagnosis patologis yang relevan secara klinis.
4. Profil komorbiditas merepresentasikan beban komorbiditas kumulatif pasien dalam basis data. Desain ini tidak membedakan secara pasti antara penyakit yang sudah ada sebelum admisi dan kondisi yang muncul selama perawatan.
5. ARM digunakan untuk menemukan pola kemunculan bersama antardiagnosis. Penelitian ini tidak menafsirkan ARM sebagai bukti hubungan sebab-akibat. Kualitas aturan asosiasi diukur menggunakan *support*, *confidence*, dan *lift*. Penentuan ambang nilai dijelaskan pada Bab 3 berdasarkan ukuran kohort, stabilitas jumlah aturan, dan interpretabilitas klinis.
6. Karena ARM merupakan metode eksploratif tanpa label klasifikasi, penelitian ini tidak menggunakan *ground truth* seperti pada *supervised learning*. Validasi hasil dilakukan melalui kekuatan metrik ARM, konsistensi domain klinis, dan evaluasi prognostik menggunakan Kaplan-Meier serta regresi Cox.
7. Variabel independen utama ialah kluster komorbiditas yang dibentuk dari aturan asosiasi dan penilaian domain klinis. Model Cox utama menyesuaikan usia dan jenis kelamin. Variabel klinis tambahan, seperti BMI, berat badan atau indikator keparahan lain, digunakan sebagai analisis sensitivitas apabila kelengkapan data memenuhi kriteria analisis.
8. Skor SOFA hanya tersedia pada pasien yang pernah masuk ICU. Tingkat kehilangan data SOFA yang besar terutama mencerminkan *structural*

missingness, bukan sekadar data hilang acak. Karena itu, penelitian menafsirkan SOFA sebagai indikator keparahan pada subpopulasi ICU dan tidak melakukan imputasi multipel untuk seluruh kohort.

9. Mortalitas didefinisikan berdasarkan variabel *dod* (*date of death*). Pasien tanpa tanggal kematian dianggap tersensor kanan (*right-censored*) pada waktu observasi terakhir yang tersedia. Definisi ini mengikuti struktur data MIMIC-IV, tetapi dapat tidak menangkap seluruh kematian yang terjadi di luar cakupan pencatatan.
10. Model regresi Cox mengasumsikan *proportional hazards*. Penelitian menguji asumsi ini dan menafsirkan pelanggaran asumsi secara eksplisit. Jika ditemukan pelanggaran, hasil Cox diposisikan sebagai estimasi asosiasi prognostik yang perlu dibaca bersama analisis sensitivitas.
11. Penelitian ini tidak mengembangkan model prediksi individual berbasis *machine learning*, tidak melakukan inferensi kausal, dan tidak menganalisis data genomik, proteomik atau jalur biologis. Fokus penelitian terbatas pada pola asosiasi komorbiditas, pembentukan kluster klinis, profil keparahan, dan hubungan prognostik terhadap mortalitas.

1.4 Tujuan Penelitian

Sejalan dengan rumusan masalah, penelitian ini memiliki tujuan sebagai berikut:

1. Mengidentifikasi pola asosiasi komorbiditas yang paling dominan pada pasien DMT2 berdasarkan kode ICD-10 dalam MIMIC-IV v3.1 menggunakan algoritma FP-Growth, serta mengukur kualitas aturan asosiasi dengan *support*, *confidence*, dan *lift*.
2. Mengidentifikasi kluster komorbiditas berdasarkan aturan asosiasi yang kuat dan bermakna secara klinis, mendeskripsikan karakteristik klinis dan demografis setiap kluster, serta menganalisis hubungan kluster dengan *survival* dan risiko mortalitas menggunakan Kaplan-Meier dan regresi Cox.
3. Menganalisis profil keparahan klinis pada setiap kluster komorbiditas berdasarkan CCI, skor SOFA pada subpopulasi ICU, biomarker laboratorium, dan pola penggunaan medikasi intensif.

1.5 Manfaat Penelitian

Manfaat penelitian ini dirumuskan sesuai dengan fokus penelitian, yaitu penemuan pola komorbiditas DMT2, pembentukan kluster klinis, dan evaluasi prognosis berbasis data MIMIC-IV v3.1.

Bagi peneliti, penelitian ini menjadi sarana untuk memperdalam kemampuan dalam pengolahan data kesehatan berbasis *electronic health record*, analisis data sekunder, ARM, pembentukan kluster klinis, Kaplan-Meier, dan regresi Cox. Penelitian ini juga melatih integrasi antara pendekatan komputasional, statistik, dan interpretasi klinis.

Bagi pengembangan ilmu, penelitian ini dapat memperkaya literatur mengenai komorbiditas DMT2 berbasis data ICD-10 dan *real-world data*. Penelitian ini juga dapat menjadi rujukan metodologis bagi studi berikutnya yang ingin menggabungkan ARM, klasterisasi klinis, dan analisis *survival*.

Bagi tenaga kesehatan dan peneliti klinis, hasil penelitian ini dapat memberi gambaran tentang kelompok komorbiditas yang memiliki profil keparahan dan risiko mortalitas berbeda. Temuan tersebut dapat mendukung pengembangan strategi stratifikasi risiko pada pasien DMT2, khususnya dalam konteks perawatan kritis. Namun, hasil penelitian ini tidak dapat digunakan sebagai dasar tunggal pengambilan keputusan klinis karena desain penelitian bersifat observasional dan berbasis data sekunder.

Bagi masyarakat, penelitian ini dapat meningkatkan pemahaman bahwa DMT2 sering hadir bersama penyakit penyerta yang kompleks. Pemahaman ini penting agar pencegahan, deteksi dini, dan pengelolaan komorbiditas mendapat perhatian sejak awal perjalanan penyakit.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A