



## Penggunaan SQL Server dalam Pengolahan Data *Warehouse* yang Praktis dan Berkelanjutan

Arif Kuntadi<sup>1</sup>, Yudi Feriandi<sup>2</sup>

Jurnal Algoritma

Sekolah Tinggi Manajemen dan Informatika LIKMI Bandung

Jl. Ir. H. Juanda No. 96, Lebakgede, Kecamatan Coblong, Bandung, Jawa Barat 40132 Indonesia

Email : [info@likmi.ac.id](mailto:info@likmi.ac.id)

<sup>1</sup>arifkunt4di@gmail.com

<sup>2</sup>yudiferiandi@gmail.com

**Abstrak** – Untuk menampung data yang besar dibutuhkan data *Warehouse* yang terstruktur agar menghasilkan informasi yang lebih akurat dan menunjang pengambilan keputusan bagi pemangku kepentingan organisasi. Untuk membangun data *Warehouse* dapat digunakan piranti lunak seperti Microsoft SQL server yang memiliki keunggulan pada aspek data *security*, skalabilitas, kemampuan kolaboratif, serta *extendability*. Data serial kependudukan, tenaga kesehatan, ekonomi serta berbagai data primer maupun sekunder yang bersifat kontinyu dapat di unggah kedalam SQL Server sehingga dapat tersimpan secara terstruktur dan periodik. Data yang tersimpan kedalam basis data diolah dengan algoritma data *mining* tertentu untuk menghasilkan informasi, pengetahuan, dan wisdom. Pendekatan data *Warehouse* dan data *mining* ini dapat diterapkan pada berbagai sektor misalnya kesehatan. Penelitian ini bertujuan menerapkan pendekatan data *Warehouse* dan data *mining* pada analisis rasio kebutuhan tenaga kesehatan di Propinsi Jawa Barat Indonesia menggunakan algoritma *single moving average* dan pengklasifikasian semi-automatic menggunakan statistik empirik pada SQL Server dan SQL Server *Agent* serta visualisasi menggunakan Microsoft Power BI. Klasifikasi berdasarkan analisis *forecasting* data jumlah penduduk dan penentuan rasio gap tenaga kesehatan berhasil ditemukan polanya melalui data *mining* pada SQL Server secara efisien serta dengan mudah divisualisasikan menggunakan Power BI. Dengan demikian SQL server dapat menjadi salah satu tools praktis terpilih untuk implementasi data *mining* berkelanjutan.

**Kata Kunci** – *Data Mining*; *Data Warehouse*; Power BI; SQL Server; SQL Server *Agent*.

### I. PENDAHULUAN

*Data mining* (DM) adalah proses mengekstraksi pola informasi dan pengetahuan implisit, potensial, bermanfaat, namun belum diketahui sebelumnya dari analisis terhadap sekumpulan data acak, tidak terstruktur, memiliki *noise*, atau tidak lengkap. Sumber data tersebut harus bersifat riil, substantial dan tidak memiliki *noise* sehingga data kasar harus dilakukan pengolahan terlebih dahulu. Selain itu pengetahuan yang digali dari proses *data mining* haruslah diminati atau dibutuhkan, dapat diterima, dimengerti, dan dapat digunakan secara praktis oleh pengguna [1]. *Data mining* adalah langkah utama dalam proses *Knowledge Data Discovery* (KDD), yang terdiri dari implementasi metode analisis data dan algoritma *data discovery* untuk menemukan pola tertentu yang bermanfaat [2]. Selama ini seringkali data diolah hanya sebatas menghasilkan informasi yang bersifat satu arah dari penyaji data dan belum mampu memberikan pengetahuan dan pola pengetahuan tersembunyi yang dapat memberikan keputusan atau pengetahuan yang memiliki *wisdom* bagi penggunanya. Dengan pendekatan *data mining* maka adanya pengenalan pola-pola tertentu yang sebelumnya tersembunyi dari sekumpulan data pada akhirnya dapat menggiring pada lahirnya kearifan (*wisdom*) dari ditemukannya pengetahuan yang bermanfaat dan memberikan kearifan dalam pengambilan keputusan di sisi pengguna

misalnya dalam proses pengambilan keputusan manajemen [3]. Untuk pengambilan keputusan yang bersifat kontinyu, sebuah organisasi atau entitas dapat menyusun *data Warehouse* berkelanjutan sebagai sumber *data mining* untuk pengambilan keputusan organisasinya.

*Data Warehouse* adalah organisasi data yang terintegrasi, berorientasi subjek, memiliki kerangka waktu, dan bersifat *nonvolatile* yang berperan mendukung *Decision Support System (DSS)* organisasi [4], [5]. Berbagai tools *data mining* telah digunakan dan semakin berkembang dari tahun ke tahun. Kriteria dalam membandingkan piranti lunak *data mining* adalah portabilitas, keandalan, efisiensi, *human engineering*, pemahaman, modifikasi, harga, ketersediaan pelatihan, dan dukungan pengguna. Terdapat berbagai piranti lunak seperti WEKA, SQL Server, Microsoft Azure ML Studio, RapidMiner, H2O dan Apache Spark [2]. Salah satu tools yang dapat digunakan dalam mendukung *business intelligence* adalah SQL Server. SQL Server memiliki kemampuan untuk memfasilitasi tiga struktur hierarki *data mining*, yaitu SSRS (*SQL Server Reporting Services*), SSAS (*SQL Server Analyzing Services*), dan komponen *data mining (DM)* itu sendiri. Beberapa kelebihan SQL Server diantaranya adalah skalabilitas fitur akses dan keamanan, kemampuan integrasi dengan *business intelligence*, adanya *team ownership*, dan ekstensibilitasnya [1], [6]. Piranti lunak SQL Server Agent adalah layanan Microsoft yang menjalankan tugas administratif terjadwal/*schedule*, yang disebut pekerjaan/*job*. SQL Server Agent dapat digunakan untuk menjalankan pekerjaan T-SQL untuk membangun ulang indeks, menjalankan pemeriksaan kerusakan, dan mengumpulkan data agregat dalam instans DB SQL Server. Hal ini menunjang fungsi SQL Server Agent dalam proses *data mining* yang praktis dan berkelanjutan.

Salah satu penerapan SQL Server dalam *data mining* adalah dalam melakukan klasifikasi dari serangkaian pola dalam *data Warehouse*. Pada penelitian ini akan dilakukan studi kasus penggunaan SQL Server dalam mengklasifikasikan *data Warehouse* di bidang kesehatan. Penggunaan *data mining* dalam sektor kesehatan sendiri terus meningkat ditandai dengan dihasilkannya berbagai publikasi tentang *data mining* dalam mengamati pola morbiditas, mortalitas, maupun determinan kesehatan sebagaimana dilakukan oleh Nugroho [7], Prakarsya dan Arif [8], serta Sundari dan Ariani [9]. Namun, sayangnya di Indonesia pendekatan ini belum banyak dilakukan oleh para praktisi teknologi informasi maupun pemangku kepentingan secara masif, padahal pentingnya memahami sebuah data menjadi *wisdom* akan turut mewarnai kebijakan yang dihasilkan dalam mencapai tujuan pembangunan kesehatan berkelanjutan [3].

Studi kasus pada penelitian ini akan menyoroti tentang disparitas sumber daya manusia Kesehatan. Kesehatan merupakan hak azasi dan komponen vital dalam kehidupan manusia. *Sustainable Development Goals (SDGs)* sendiri menempatkan aspek kesehatan sebagai komponen penting untuk dicapai bersama oleh seluruh negara di dunia [10]. Disparitas sumber daya ini menyebabkan terjadinya ketidakmerataan dan ketidakadilan akses terhadap pelayanan kesehatan yang berdampak pada tidak samanya pencapaian derajat kesehatan antara satu wilayah dengan wilayah lainnya [11]. Data Kementerian Kesehatan RI tahun 2017 menunjukkan bahwa rasio dokter umum adalah 45 per 100.000 penduduk, perawat 170 per 100.000 penduduk dan bidan sebesar 162 per 100.000 penduduk di Indonesia. Selain itu kekurangan tenaga kesehatan apabila didasarkan pada standar minimal tenaga kesehatan adalah 49,632 orang yang diperparah dengan adanya maldistribusi 155,833 orang tenaga kesehatan di Puskesmas sebagai *gate keeper* pembangunan kesehatan. Bahkan sejak lebih dari satu dekade lalu disparitas sebaran dokter spesialis masih belum teratasi di Indonesia [12]–[14]. Berbagai fakta tersebut akan dijadikan sebagai dasar studi kasus yang tepat dengan penggunaan karakteristik dan keunggulan SQL Server dalam *data mining* yang berkelanjutan dan terintegrasi dengan pengambilan keputusan [3].

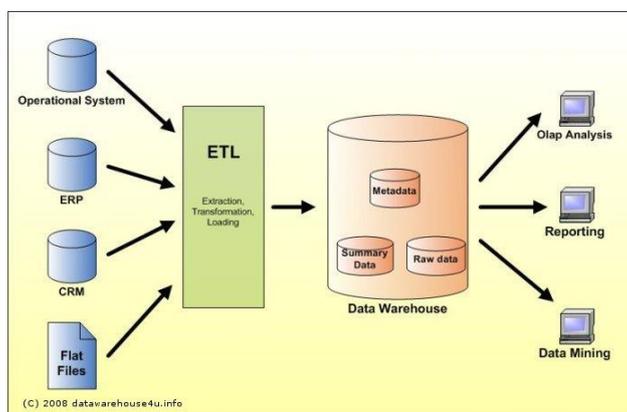
Oleh karena itu, penelitian ini bertujuan untuk melakukan *data mining* menggunakan SQL Server dan ekstensibilitas visualisasi hasil *data mining* menggunakan Microsoft Power BI dalam pengolahan data yang bersifat praktis dan kontinyu dengan studi kasus pada pengolahan data demografi dan data sumber daya manusia bidang kesehatan untuk mengklasifikasikan wilayah Kabupaten/Kota berdasarkan kesenjangan pemenuhan rasio tenaga kesehatan di masa depan. Melalui studi kasus yang nyata, diharapkan penelitian ini dapat memberikan penjelasan tentang aspek praktis dan keunggulan SQL Server dalam pengolahan *data Warehouse* yang berkelanjutan serta ekstensibilitasnya sebagai sumber basis data bagi piranti lunak pengolah dan visualisasi data lainnya.

## II. METODOLOGI PENELITIAN

### A. Data Warehouse

Menurut Inmon (2002) *data Warehouse* adalah sekumpulan data yang bersifat *integrated*, *subject-oriented*, *time variant* dan *nonvolatile* dalam mendukung pengambilan keputusan manajemen [5]. Untuk memenuhi kebutuhan organisasi susunan komponen harus diatur dengan cara tertentu untuk mendapatkan manfaat yang maksimal, hal ini bisa dilakukan dengan memberikan penekanan khusus pada suatu komponen atau memberikan dukungan komponen lain [4].

Pada Gambar 1, menjelaskan mengenai konsep *data Warehouse* sebagai konsep dasar dalam pengolahan data *mining*. Melalui konsep yang terstruktur akan memberikan hasil pengolahan data yang mudah khusus nya dalam mengelola data yang besar.



Gambar 1: Model konseptual *data Warehouse* [5]

### B. Data Mining

*Data mining* adalah proses pengumpulan dan pengolahan data yang bertujuan untuk mengekstrak informasi penting pada data. Proses pengumpulan dan ekstraksi informasi tersebut dapat dilakukan menggunakan perangkat lunak dengan bantuan perhitungan statistika, matematika, ataupun teknologi *Artificial Intelligence* (AI). *Data mining* sering disebut juga *Knowledge Discovery in Basis data* (KDD). *Data mining* dilakukan untuk memenuhi berbagai tujuan baik yang bersifat menjelaskan (*Explanatory*), konfirmasi (*Confirmatory*), maupun eksplorasi (*Exploratory*). *Data mining* memiliki beberapa metode, yaitu asosiasi, klasifikasi, regresi, dan klastering [15], [16].

### C. Store Procedure SQL Server

*Store Procedure* merupakan kumpulan perintah atau statement SQL yang disimpan didalam basis data dengan nama tertentu. Keuntungan penggunaan *Store Procedure* ini dapat menyembunyikan *Query SQL* didalam kode program. *Query SQL* akan ditulis dan dieksekusi disisi server basis data sehingga dapat meningkatkan kecepatan aplikasi. Dapat digunakan berulang-ulang dan dapat mengamankan nama tabel basis data yang digunakan. Dalam pengolahan dan proses data yang kompleksitasnya tinggi, *store procedure* membuat struktur proses lebih rapih. Hal ini sangat membantu proses dalam pembuatan dan pengolahan *data mining* sehingga dalam langkah demi langkah proses pengolahannya dapat terlihat dan terstruktur dengan rapih. Dalam membuat proses *data mining*, biasanya banyak *script* yang digunakan, apabila seluruh *script* di simpan dalam wadah-wadah yang mewakilkan proses *script* tersebut, akan memudahkan bagi administrator basis data atau *user* yang membuat *script* tersebut apabila ada kendala saat proses *data mining* [16], [17].

## D. Dimensional Modeling

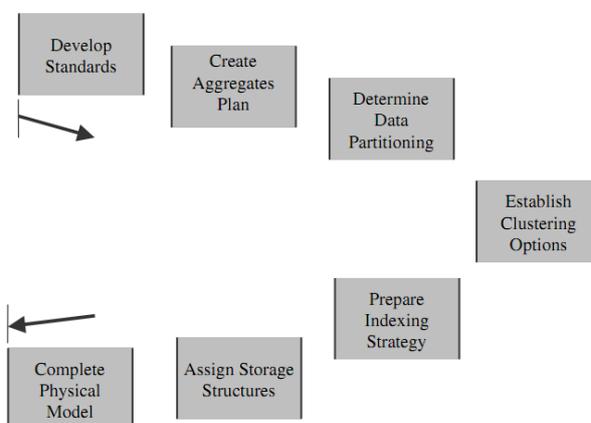
*Dimensional modeling* adalah teknik yang digunakan untuk menyajikan data analitik untuk men-deliver data yang dapat dimengerti pengguna bisnis. *Dimensional modeling* merupakan pemodelan data untuk menangkap data dan menjadikan informasi yang mendukung analisis yang kompleks dan pendekatan yang berbeda dari perancangan sistem operasional. Menurut Gupta & Krishna, *dimensional modeling* adalah sebuah teknik disain logika yang bertujuan untuk menampilkan data dalam bentuk standar, intuitif yang memungkinkan akses cepat. Berdasarkan Connolly & Begg, *dimensional modeling* menggunakan konsep model *Entity-Relationship* (ER) dengan beberapa batasan penting. Setiap model dimensi terdiri dari satu tabel dengan satu *composite primary key* yang disebut tabel fakta (*fact* tabel) dan memiliki kumpulan dari tabel yang lebih sederhana yang disebut tabel dimensi (*dimension* tabel). Tiap tabel dimensi memiliki *primary key* (*non composite*) yang akan berkorespondensi tepat satu dengan komponen pada *composite key* dalam tabel fakta [18].

## E. Perancangan Data Warehouse

Pada proses perancangan *data Warehouse* data demografi dan data kesehatan di wilayah Kabupaten/Kota Provinsi Jawa Barat, metode yang dilakukan menggunakan beberapa tahapan:

### 1. Desain Data Warehouse

Tahapan proses disain *data Warehouse* dimulai dari pengembangan standar sampai dengan *complete Physical Model*. Dalam hal ini ditentukan dimana setiap tabel mempunyai sejumlah kolom dan setiap kolom memiliki nama yang unik. Sebuah *physical data model* (disain basis data) adalah representasi dari desain data yang memperhitungkan fasilitas dan kendala sistem basis data yang diberikan manajemen. Dalam siklus hidup proyek itu biasanya berasal dari model data logis, meskipun mungkin *reverse-engineered* dari implementasi basis data yang diberikan. Pada proses proses ini disiapkan struktur data pada *SQL Server* dengan membentuk tabel-tabel sebagai penampung data yang akan di unggah dari sumber data yang diperoleh. Dipastikan struktur tabel di setiap kolomnya di sesuaikan tipe datanya agar tidak menyebabkan *invalid process* pada saat diunggah kedalam basis data [4].



Gambar 2: Proses disain *data Warehouse* dengan *physical data model*

### 2. Pemilihan Sumber Data / Dataset

Sebelum dilakukan analisis data, data yang digunakan harus benar-benar data yang valid dan dapat dipertanggungjawabkan, sehingga saat pengambilan keputusan, hasil yang diperoleh akan tepat sesuai dengan luaran yang diharapkan. *Dataset* yang digunakan pada penelitian ini bersumber dari laman *Open Data* Jabar dengan tautan <https://opendata.jabarprov.go.id> yang dihimpun menggunakan metode *crawling* pada laman resmi *Open Data* Jabar [19]. Sebagai sebuah studi kasus digunakan data pada tingkat Kabupaten/Kota di Jawa Barat dari aspek kependudukan dan data ketersediaan tenaga kesehatan serta tenaga medis pada tahun 2016 s.d 2021. *Dataset* yang digunakan adalah:

- a. Data Jumlah Penduduk Jawa Barat per Kabupaten/Kota periode rentang waktu antara tahun 2016 s.d 2021

b. Data Distribusi Tenaga Kesehatan di masing-masing Kabupaten/Kota di Propinsi Jawa Barat yang terdiri dari data jumlah Dokter Spesialis, Dokter, Dokter Gigi, Perawat, Bidan, Tenaga Gizi, dan Sanitarian dengan periode rentang waktu antara tahun 2016 s.d 2021.

Dari *dataset* tersebut didapatkan 189 baris data dan 14 Kolom yang terdiri dari kode kabupaten kota, nama kabupaten kota, jumlah penduduk tahun 2016 s.d 2021, dan tujuh kolom tenaga kesehatan sesuai jenisnya.

3. Identifikasi dan Penyesuaian Dimensi

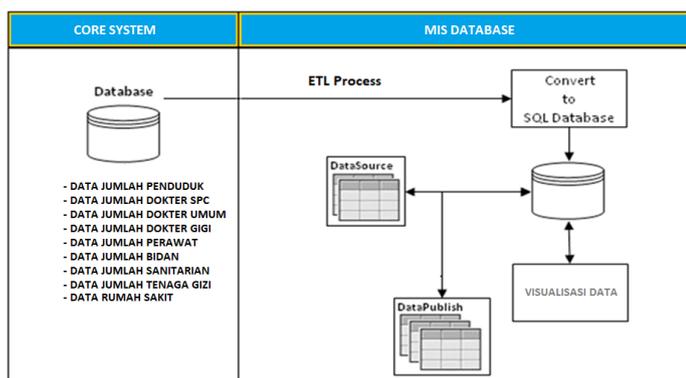
Pada tahap ini dilakukan identifikasi dan penyesuaian dimensi yang terkait dengan fakta yang ditampilkan dalam bentuk matriks. Adapun dimensi yang dihasilkan: profesi, kategori kota/kabupaten, kategori *gap* rasio tenaga kesehatan di masing-masing Kabupaten/Kota.

4. Pemilihan Fakta

Pada tahap ini dilakukan pemilihan fakta yang akan digunakan pada tabel fakta serta penyajian hubungan antara tabel fakta dan tabel dimensi. Fakta-fakta tersebut adalah Rata-rata pertumbuhan penduduk di setiap Kabupaten/Kota, “jumlah penduduk *forecast*”, “jumlah tenaga kesehatan *forecast*”, “*gap* tenaga kesehatan antara data aktual dengan data *forecast*”, serta “persentase *gap* tenaga kesehatan terhadap data *forecast*”. Data-data tersebut sebagai dasar dalam pembentukan *datamart* yang digunakan dalam proses analisis tenaga kesehatan dengan studi kasus pada data di kabupaten/kota khususnya di Jawa Barat.

5. Arsitektur Data Warehouse

Dalam implementasinya, proses pengolahan data dilakukan melalui SQL Server, dengan menggunakan mekanisme proses yang di tampung melalui *store procedure*, seluruh proses dapat dirangkum dengan rapi guna memudahkan dalam mengidentifikasi dan mengelempokkan proses. *Store procedure* yang telah dibuat kemudian dijalankan secara otomatis dengan menggunakan SQL Server Agent sehingga dapat termonitor proses harian pengolahan data dari impor data/*Convert to SQL* sampai dengan data yang siap disajikan (*Data Publish*) dan divisualisasikan lebih lanjut. *Data Source* berfungsi dalam menyimpan *data history* dari sumber data yang belum diolah. *Data publish* merupakan data hasil olahan yang berupa *datamart* yang akan digunakan dalam proses *data mining* sebagai visualisasi. Dalam hal ini, visualiasi diimplementasikan menggunakan piranti lunak Microsoft Power BI. Pada Gambar 3, dijelaskan mengenai arsitektur data Warehouse yang akan digunakan dalam pengolahan data untuk penelitian. Melalui arsitektur ini, memberikan kemudahan dalam mengelompokkan data yang bersifat Detail, dan yang bersifat Summry. Datasource bersifat data detail, sedangkan Datapublish bersifat summary.



Gambar 3: Arsitektur data Warehouse

6. Ekstraksi Data

Dalam pengolahan *data Warehouse* yang dilakukan dimulai dari konversi *dataset* mentah/*raw data* sampai menjadi *datamart* yang siap diolah dengan *data mining*. Proses *extracting data* dilakukan menggunakan SQL Server dengan mekanisme proses data yang sederhana. Dengan menggunakan pendekatan statistik empirik melalui penghitungan rata-rata dan standar deviasi menggunakan fungsi STDEVP pada SQL server sebagai dasar dalam pengklasifikasian wilayah dimana setiap kabupaten/kota dikategorikan menjadi dua klasifikasi yaitu [20]:

- a. Bila persentase gap tenaga kesehatan terhadap data *forecast* lebih dari nilai “0” maka diklasifikasikan “Berlebih”
- b. Bila persentase gap tenaga kesehatan terhadap data *forecast* kurang dari “0” maka dilakukan pengkategorian pada tiga subklasifikasi, yaitu:
  - 1) “Kurang”
  - 2) “Sangat Kurang”
  - 3) “Sangat Kurang Sekali” dengan pengkategorisasian sebagai berikut:

Tabel 1: Tabel klasifikasi wilayah berdasarkan gap persentase menggunakan pendekatan statistik empirik [20]

Rumus statistik empirik untuk pengklasifikasian	Kategori
Persentase gap tenaga kesehatan terhadap data <i>forecast</i> > 0	Berlebih
Persentase gap tenaga kesehatan terhadap data <i>forecast</i> < 0 dan persentase gap tenaga kesehatan terhadap data <i>forecast</i> ≥ (x̄+1. SD)	Kurang
Persentase gap tenaga kesehatan terhadap data <i>forecast</i> < 0 dan (x̄-1. SD) ≤ persentase gap tenaga kesehatan terhadap data <i>forecast</i> < (x̄+1. SD)	Kurang Sekali
Persentase gap tenaga kesehatan terhadap data <i>forecast</i> < 0 dan persentase gap tenaga kesehatan terhadap data <i>forecast</i> < (x̄-1.SD)	Sangat Kurang Sekali

Proses pengelompokan tersebut di lakukan dengan 2 proses *Store Procedure*:



Gambar 2: Proses *Store Procedure*

### III. HASIL DAN PEMBAHASAN

#### A. Analisis *Data Warehouse* dan Hasil *Data Mining*

Perancangan *Data Warehouse* yang baik dan terstruktur memberikan kemudahan dalam pengelolaan data yang bersifat *continue*. Dimulai dari proses penyusunan basis data, struktur dan relasi antar tabel, penentuan *index*, pemanfaatan *script*, pembuatan *store procedure* dan pembuatan *job* melalui *system Agent* yang dapat membuat proses menjadi cepat, merupakan bentuk satu kesatuan dalam penyusunan *data Warehouse* yang baik. Pada analisis ini, dibuat dua buah basis data:

1. *Data Source*, untuk menyimpan data yang bersifat *historical*
2. *Data Publish*, untuk meyimpan *datamart* sebagai salah satu implemasi *Data mining*

Dimulai dari penyusunan data yang telah dikumpulkan dari sumber laman *dataset* dengan format *.xls*, kemudian diterjemahkan kedalam bentuk tabel-tabel pada *SQL Server*. sebagai berikut:

Column Name	Data Type	Allow Nulls	Column Name	Data Type	Allow Nulls
id	varchar(20)	☑	id	varchar(20)	☑
kode_provinsi	varchar(10)	☑	kode_provinsi	varchar(10)	☑
nama_provinsi	varchar(250)	☑	nama_provinsi	varchar(250)	☑
kode_kabupaten_kota	varchar(50)	☑	kode_kabupaten_kota	varchar(50)	☑
nama_kabupaten_kota	varchar(250)	☑	nama_kabupaten_kota	varchar(250)	☑
jenis_kelamin	varchar(50)	☑	jumlah_dokter_spesialis	int	☑
jumlah_penduduk	float	☑	satuan	varchar(50)	☑
satuan	varchar(50)	☑	tahun	int	☑
tahun	int	☑			
kategori_usia	varchar(50)	☑			

Gambar 3: Tabel *dataset* demografi dan tenaga kesehatan (contoh: dokter spesialis) pada SQL Server

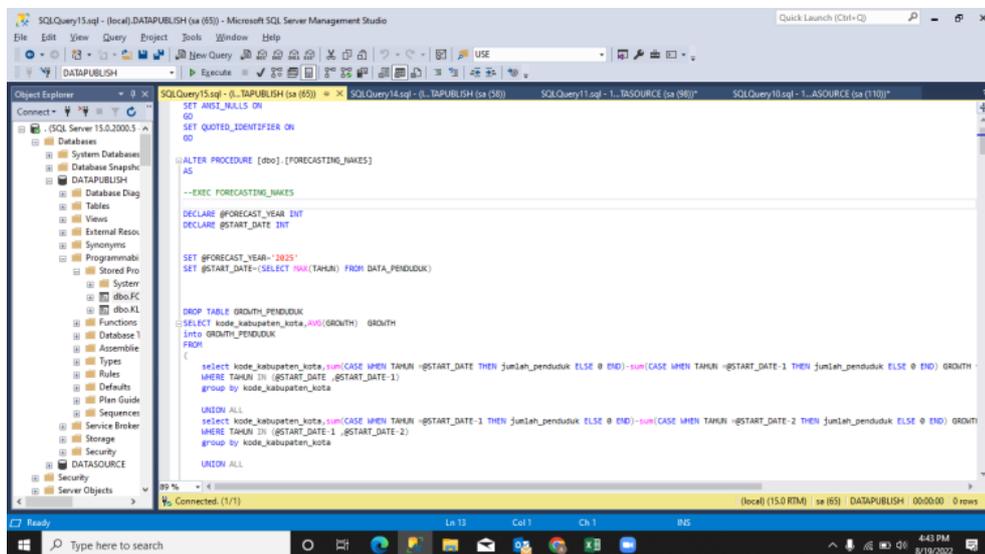
Kedua tabel di atas merupakan dua dari beberapa struktur tabel yang disimpan kedalam basis data yang merupakan *history* data dari setiap tahunnya. Dengan struktur yang sama seperti tabel Dokter Spesialis, dibuat juga tabel Dokter Umum, Dokter Nakes, Perawat, Bidan, Nakes Giji, dan Nakes Lingkungan didalam basis data yang telah dibuat sebelumnya, yaitu disimpan di *data Warehouse*. Setelah data berhasil diupload ke dalam *SQL Server*, dipastikan data yang digunakan telah dilakukan *cleansing* data, hal ini dilakukan untuk:

1. Mengantisipasi isi data yang tidak wajar, yang menyebabkan proses analisis menjadi *invalid*.
2. Memastikan isi data yang berelasi dengan tabel-tabel yang lain sehingga terbentuk kesesuaian relasi antar tabel

Proses berikutnya adalah dengan membuat *Script SQL* yang diterjemahkan dalam bentuk *Store Procedure* dengan membentuk 2 Procedure SQL:

1. *Forecasting\_Nakes*

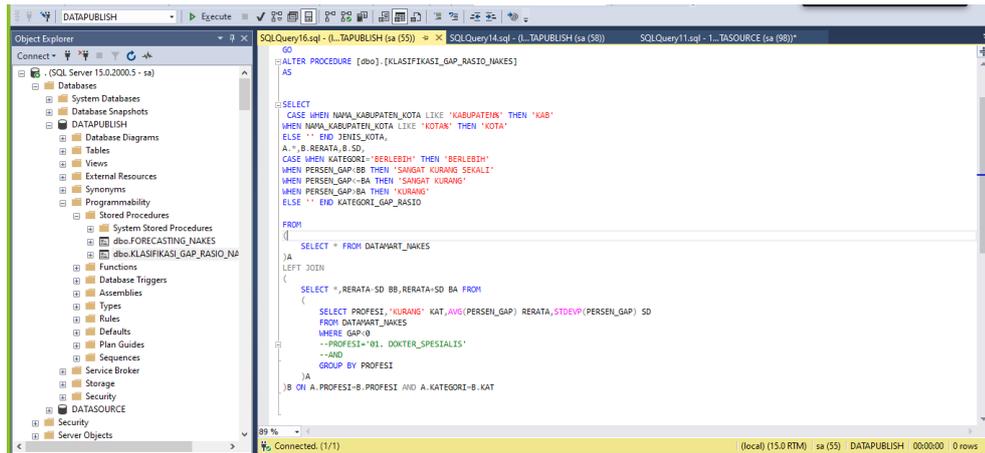
*Script* ini berisi dari proses pengolahan data dari yang sudah diunggah kemudian dilakukan *query* “Join” dengan tabel yang lainnya dan dibentuk *datamart* yang diberi nama *DATAMART\_NAKES*. Mulai dari menghitung rata-rata kenaikan jumlah penduduk, *forecasting* menggunakan pendekatan *single moving average* dan penghitungan data persentase (%) *gap* dari masing-masing data Kabupaten/ Kota



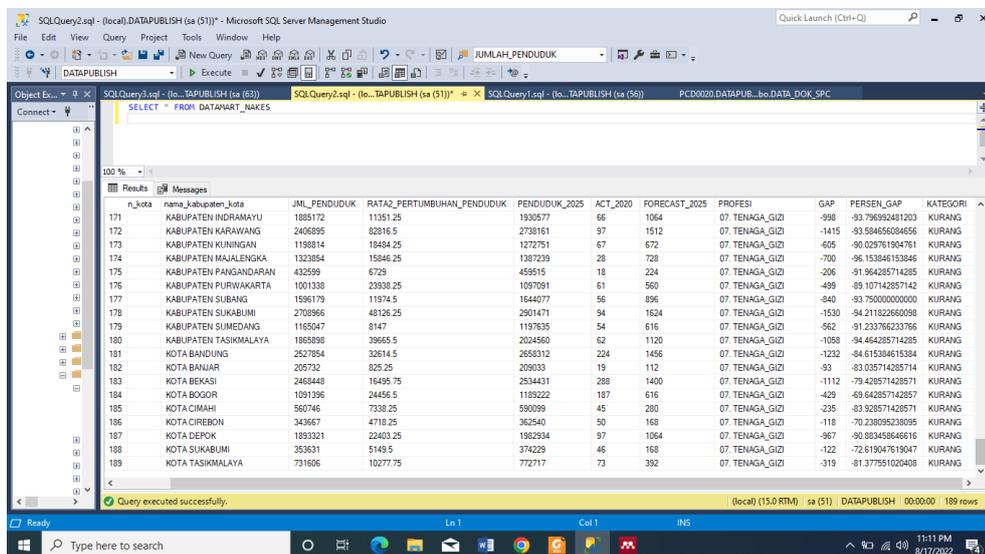
Gambar 4: Script *Forecasting\_Nakes* pada SQL Server

2. *Klasifikasi\_gap\_Rasio Nakes*

Procedure ini berisi *script*/perintah-perintah dalam menentukan klasifikasi data berdasarkan *gap* rasio. Perhitungan *gap* rasio merupakan implementasi metodologi statistik empirik dimana teknik ini rerata dan deviasi standar yang didapatkan dari data empirik dipakai sebagai referensi. Hasil proses ini menghasilkan data final yang siap digunakan untuk proses analisis data.



Gambar 5: Script algoritma klasifikasi gap rasio tenaga kesehatan pada SQL Server



Gambar 6: Hasil datamart pada SQL Server

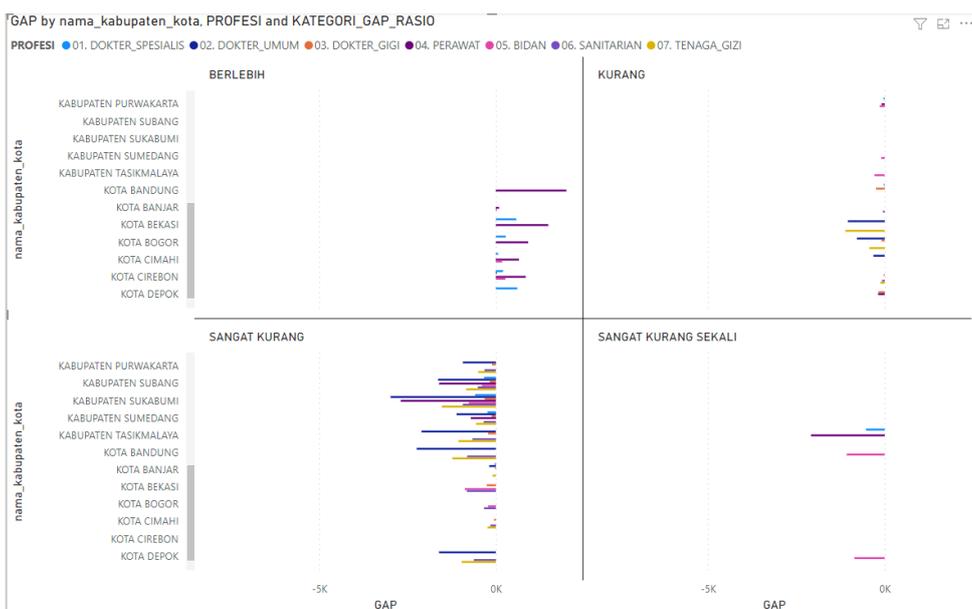
Untuk proses berikutnya setelah *datamart* sudah diperoleh, maka dilakukan analisis yang mendalam mengenai data tersebut menggunakan beberapa *tools*. Sebelum munculnya berbagai *tools data mining* dan visualisasi data, industri cenderung menggunakan Microsoft Excel, namun setelah meningkatnya kemampuan berbagai piranti lunak dalam membantu proses *data mining* dan visualisasinya maka *tools* seperti Rapid miner, Orange, Knime, dsb. banyak digunakan oleh para praktisi basis data maupun informatika. Pada penelitian ini digunakan *tools* Microsoft Power BI dalam memvisualkan hasil *data mining* dari SQL Server. Penyajian hasil *data mining* dari SQL Server berhasil mengklasifikasikan berdasarkan algoritma klasifikasi dengan pendekatan statistik empirik sebagaimana tersaji pada tabel 4. Selain itu, luaran *data mining* dapat divisualisasikan lebih lanjut secara lebih detail dan dinamis menggunakan *tools* Microsoft power BI sebagaimana tersaji pada Gambar 6, Gambar 9, dan Gambar 10.

JENIS_KOTA	PROFESI	BERLEBIH	KURANG	SANGAT KURANG	SANGAT KURANG SEKALI	Grand Total
KAB	01. DOKTER_SPESIALIS		3	13	2	18
	02. DOKTER_UMUM			18		18
	03. DOKTER_GIGI			18		18
	04. PERAWAT		2	12	4	18
	05. BIDAN	2	4	10	2	18
	06. SANITARIAN		1	16	1	18
	07. TENAGA_GIZI			18		18
KOTA	01. DOKTER_SPESIALIS	6	2	1		9
	02. DOKTER_UMUM	1	4	4		9
	03. DOKTER_GIGI		4	5		9
	04. PERAWAT	8	1			9
	05. BIDAN	5		2	2	9
	06. SANITARIAN		3	6		9
	07. TENAGA_GIZI		4	5		9

Gambar 7: hasil data mining klasifikasi wilayah Kabupaten/Kota menggunakan SQL Server

nama_kabupaten_kota	01. DOKTER_SPESIALIS	02. DOKTER_UMUM	03. DOKTER_GIGI	04. PERAWAT	05. BIDAN	06. SANITARIAN	07. TENAGA_GIZI		
KABUPATEN BANDUNG	-798	-3781	-400	-3953	-768	-1205	-1948	-12853	
SANGAT KURANG	-798	-3781	-400		-768	-1205	-1948	-8900	
SANGAT KURANG SEKALI				-3953				-3953	
KABUPATEN BANDUNG BARAT	-355	-1954	-206	-1955	-787	-620	-1007	-6884	
KABUPATEN BEKASI	-58	-2648	-383	-1719	-1002	-1163	-1824	-8797	
KABUPATEN BOGOR	-1092	-6513	-689	-6497	-3052	-2162	-3403	-23408	
KABUPATEN CIAMIS	-337	-1306	-141	-941	-250	-415	-680	-4070	
KABUPATEN CIANJUR	-645	-2552	-282	-3156	-833	-853	-1395	-9716	
KABUPATEN CIREBON	-464	-2527	-282	-1300	-400	-831	-1322	-7126	
KABUPATEN GARUT	-746	-3100	-328	-563	-805	-996	-1584	-8122	
KABUPATEN INDRAMAYU	-455	-1855	-206	-1432	-414	-585	-998	-5945	
KABUPATEN KARAWANG	-192	-2601	-286	-1301	-509	-897	-1415	-7201	
KABUPATEN KUNINGAN	-177	-1158	-131	-654	61	-372	-605	-3036	
KABUPATEN MAJALENGKA	-299	-1342	-143	-1173	-423	-421	-700	-4501	
KABUPATEN PANGANDARAN	-91	-392	-44	-186	77	-110	-206	-952	
KABUPATEN PURWAKARTA	-32	-933	-109	-86	-138	-321	-499	-2118	
KABUPATEN SUBANG	-334	-1634	-175	-1606	-395	-511	-840	-5495	
KABUPATEN SUKABUMI	-587	-2979	-315	-2693	-764	-937	-1530	-9805	
KABUPATEN SUMEDANG	-236	-1113	-121	-708	-97	-346	-562	-3183	
KABUPATEN TASIKMALAYA	-529	-2106	-220	-2085	-290	-663	-1058	-6951	
KURANG		-2106	-220		-290	-663	-1058	-290	
SANGAT KURANG			-220			-663	-1058	-4047	
SANGAT KURANG SEKALI	-529			-2085				-2614	
KOTA BANDUNG	-20	-2241	-243	1996	-1072	-813	-1232	-3625	
BERLEBIH				1996				1996	
KURANG	-20		-243					-263	
SANGAT KURANG		-2241				-813	-1232	-4286	
SANGAT KURANG SEKALI					-1072			-1072	
KOTA BANJAR	-28	-187	-22	90	26	-50	-93	-264	
KOTA BEKASI	577	-1042	-258	1483	-878	-820	-1112	-2050	
BERLEBIH	577			1483				2060	
KURANG		-1042						-1112	
SANGAT KURANG			-258		-878	-820		-1956	
KOTA BOGOR	278	-785	-92	916	-218	-335	-429	-665	
KOTA CIMAHI	62	-319	-50	651	177	-154	-235	132	
Total	-5789	-47370	-5423	-25249	-13137	-16564	-26203	-139735	

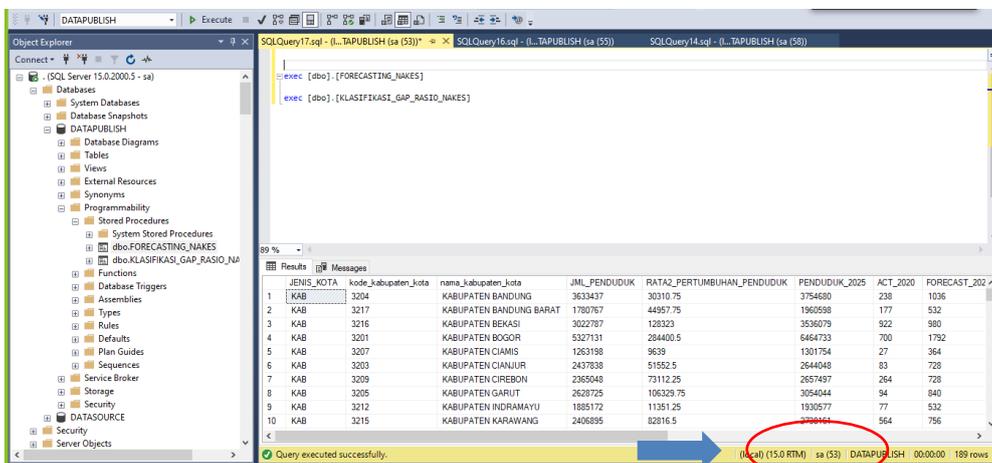
Gambar 8: visualisasi dinamis hasil data mining klasifikasi wilayah Kabupaten/Kota menggunakan Power BI



Gambar 9: Hasil visualisasi dinamis dengan tampilan matriks pada Power BI



Gambar 10: Hasil visualisasi dinamis dengan tampilan *field map* pada Power BI



Gambar 11: Waktu yang diperlukan untuk proses *data mining* (di dalam lingkaran merah yang ditunjukkan dengan anak panah)

## B. Pembahasan

*Data mining* telah memberikan paradigma baru dalam proses analisis data khususnya Big Data melalui pendekatan DKIW dan KDD yang menjadikan data tidak hanya diolah menjadi informasi, namun lebih lanjut dapat diidentifikasi pengetahuan dan pola yang tersimpan di dalamnya [3], [21] Pendekatan ini sangat bermanfaat bagi organisasi baik pemerintah maupun swasta. Salah satu kendala dalam *data mining* adalah penguasaan operator basis data/administrator basis data dan teknisi informatika di organisasi/institusi untuk menggunakan berbagai piranti lunak yang dapat digunakan dalam proses *data mining* [4], [22]. Selain itu, data yang tersimpan di dalam basis data organisasi yang bersifat kontinyu dapat diolah sedemikian rupa untuk menghasilkan kearifan bagi organisasi tersebut. Data yang bersifat kontinyu dapat dikelola dengan baik dan rapi melalui data warehouse manajemen yang terencana dan terstruktur dengan baik sehingga bisa menghasilkan *datamart* yang dapat dengan mudah dan menggunakan algoritma sederhana maupun kompleks dianalisis sesuai kebutuhan pemangku kepentingan organisasi. Peran berbagai jenis metode statistik pada proses *data mining* dan pemilihannya seringkali menjadi kendala dalam memahami dan mempraktikkan *data mining* pada *dataset* yang ada.

Hal ini memerlukan peningkatan kapasitas tenaga informatika dan komputer di organisasi tersebut yang bertugas mengelola dan menganalisis data baik yang bersumber dari internal maupun eksternal organisasi [2]. Instrumen *data mining* yang banyak digunakan saat ini diantaranya Rapid Miner, Knime, Orange, SQL Server,

dsb. *SQL Server Agent 2019* merupakan salah satu *tools* yang banyak digunakan di berbagai organisasi untuk pengelolaan basis data secara otomatis dalam proses pengelolaan *Warehouses*nya. Dengan *advanced query* untuk *data mining* dapat mentransformasi *SQL Server* menjadi *tools* dalam proses *data mining* yang kontinyu. Selain itu hasil *data mining* pada *SQL server* dapat menjadi *source* bagi berbagai *tools data mining* lainnya atau sebagai *source* visualisasi menggunakan bahasa pemrograman seperti *Phyton* dikarenakan memungkinkannya *script SQL Server* diintegrasikan kedalamnya [1], [2], [6]. Dikarenakan keterbatasan penguasaan metode dan *tools data mining*, masih banyak organisasi yang menggunakan *tools* sederhana seperti *Microsoft Excel* misalnya untuk mengolah data dalam jumlah besar. Meskipun memungkinkan, namun akan menghasilkan inefektivitas dan inefisiensi waktu dan sumber daya untuk mengolah data dalam jumlah besar apabila menggunakan *Ms Excel*. Oleh karena itu, dikarenakan *SQL Server* merupakan piranti lunak yang lazim dan dikuasai oleh sebagian besar tenaga *MIS (Management Information System)* organisasi, maka aspek kepraktisan dan efisiensi *SQL Server* dengan manajemen *data Warehouse* yang baik akan sangat menunjang pelaksanaan *data mining* di organisasi [6].

Implementasi hal di atas pada studi kasus penelitian menunjukkan bahwa penggunaan *SQL Server* dan *tools* visualisasi yang *user friendly* sebagai *tools data mining* dan visualisasi mampu memberikan input *wisdom* bagi pemangku kepentingan terkait pengambilan keputusan perencanaan penyediaan sumber daya manusia bidang kesehatan di Kabupaten/Kota di Propinsi Jawa Barat di masa mendatang. Sebagaimana penelitian tentang penggunaan *data mining* dalam *Decision Support System* kebijakan kesehatan di tingkat regional di Rusia oleh Rapakof dkk. (2021) [23]. Pada studi kasus ini waktu pengolahan data dengan *Ms. Excel* yang memerlukan pengolahan selama beberapa menit dapat disederhanakan dalam beberapa detik menggunakan *Script Query* Statistik Empirik pada *SQL Server*. Begitu pula hasil tabel *datamart* dari proses *data mining* dengan sangat mudah di ekspor pada *tools* visualisasi *Power BI* sehingga menghasilkan visualisasi yang spesifik dan dinamis. Dengan demikian, penggunaan *SQL Server*, *SQL Server Agent*, dan *tools* visualisasi *Power BI* pada penelitian ini dapat menjadi salah satu metode sederhana dan praktis dalam pengolahan *big data* yang bersifat kontinyu di berbagai sektor baik publik maupun swasta serta data internal maupun eksternal organisasi khususnya di dalam pengambilan keputusan analisis data yang bersifat klasifikasi.

#### IV. KESIMPULAN

Berdasarkan penelitian ini dapat disimpulkan bahwa penggunaan *SQL Server* dan *SQL Server Agent* dapat menjadi salah satu instrumen *data mining* yang praktis dan mendukung proses *data mining* berkelanjutan. Manajemen dan disain *data Warehouse* yang baik dengan *store procedure* yang dihasilkan mampu menyajikan *datamart* kontinyu yang dapat dimutakhirkan secara periodik dan berkelanjutan menyesuaikan dengan periodisasi *dataset* yang ingin dianalisis sebagai bahan pengambilan keputusan. Studi kasus terhadap pengolahan *dataset* kependudukan dan sumber daya manusia bidang kesehatan di kabupaten/kota Propinsi Jawa Barat menunjukan bahwa kombinasi antara *SQL Server* dan piranti lunak *data mining* dan visualisasi data lainnya seperti *Ms. Power BI* dapat menunjang proses bisnis dan perencanaan organisasi yang praktis dan berkelanjutan khususnya dalam mendukung pengambilan keputusan yang berkelanjutan. Keterbatasan pada penelitian ini adalah pada jumlah *dataset* yang tidak terlalu besar dikarenakan sulitnya *crawling* data yang mewakili seluruh kabupaten/kota di Indonesia serta banyaknya *missing data* sehingga harus dikaji validitas datanya. Selanjutnya dapat juga dilakukan *advance statistical query* pada *SQL Server* sesuai jenis algoritma *data mining* lainnya serta penggunaan *tools* visualisasi lainnya selain *Ms Power BI* untuk menunjukkan *extendability* dari basis data yang dihasilkan oleh *SQL Server*. Selain itu dapat pula dikaji tentang pengintegrasian *query data mining* pada *SQL Server* dalam *script* pemrograman pada aplikasi *front end* yang memungkinkan user untuk memvisualisasikan basis data kontinyu dari *datamart* yang dihasilkan.

#### DAFTAR PUSTAKA

- [1] Z. Zhang, G. Wang, L. Yang, M. Zhang, W. Zhao, and X. Xu, "Application of SQL server in data mining," 2010 Chinese Control Decis. Conf. CCDC 2010, pp. 521–524, 2010, doi: 10.1109/CCDC.2010.5498995.

- [2] D. Dušanka, S. Darko, S. Srdjan, A. Marko, and L. Teodora, "A Comparison of Contemporary Data Mining Tools," no. October, 2017, [Online]. Available: <http://www.iim.ftn.uns.ac.rs/is17>
- [3] M. Mulyadi, "Transisi Data dan Informasi dalam Pengembangan Ilmu Pengetahuan," *Pustakaloka*, vol. 10, no. 1, p. 67, 2018, doi: 10.21154/pustakaloka.v10i1.1237.
- [4] W. H. Inmon, *AM Building the*, 3rd ed. Canada: John Wiley & Sons, Inc., 2002.
- [5] S. Bakhri, Y. Nuryamin, P. Labu, and J. Selatan, "Rancangan Data Warehouse Untuk Penunjang Sistem Informasi Eksekutif Pada Yayasan Ummu'L Quro Di Depok," *Tek. Inform.*, vol. 4, no. 1, 2018.
- [6] Z. Ren, "Advanced studying on microsoft SQL server data mining," *CICC-ITOE 2010 - 2010 Int. Conf. Innov. Comput. Commun. 2010 Asia-Pacific Conf. Inf. Technol. Ocean Eng.*, pp. 87–89, 2010, doi: 10.1109/CICC-ITOE.2010.30.
- [7] O. Nugroho, B. A. Kusuma, T. Astuti, U. Khasanah, and R. Wahyudi, "Penerapan Data Mining untuk Mengidentifikasi Penyakit Diabetes Mellitus dengan Menggunakan Algoritme Iterative Dichotomiser 3 (ID3)," pp. 101–104, 2018.
- [8] L. Dempo, "Implementasi Data Mining Untuk Prediksi Penyebaran Virus Hiv / Aids Di Bandar Lampung," vol. 3, no. 2, pp. 18–26, 2020.
- [9] S. S. Sundari and N. Ariani, "Penerapan Data Mining Untuk Pengelompokan Penyakit Dengan Algoritma Fuzzy C-Means (Studi Kasus : UPT Puskesmas Salawu)," *J. VOI (Voice Informatics)*, vol. 8, no. 2, pp. 63–76, 2019.
- [10] M. I. Fadhlurrohmah, E. P. Purnomo, and A. D. Malawani, "Analysis Of Sustainable Health Development In Indonesia (Sustainable Development Goal's)," *J. Kesehat. Lingkung. Indones.*, vol. 19, no. 2, pp. 133–143, 2020, doi: 10.14710/jkli.19.2.133-143.
- [11] H. Maryani, L. Kristiana, A. Paramita, and N. Izza, "Disparitas Pembangunan Kesehatan Di Indonesia Berdasarkan : Disparity of Health Development in Indonesia Based on Healthy Family," *Bul. Penelit. Sist. Kesehat.*, vol. 23, no. 1, pp. 18–27, 2020, [Online]. Available: <file:///C:/Users/Asus/Downloads/2622-Article Text-21954-1-10-20200505.pdf>
- [12] I. Dwiprahasto, "Jurnal Manajemen Pelayanan Kesehatan Daftar Isi," *Indones. J. Heal. Serv. Manag.*, vol. 09, no. 06, pp. 94–101, 2006.
- [13] I. Nurlinawati and R. H. Putranto, "Faktor-Faktor Terkait Penempatan Tenaga Kesehatan di Fasilitas Pelayanan Kesehatan Tingkat Pertama Daerah Terpencil/Sangat Terpencil," *J. Penelit. dan Pengemb. Pelayanan Kesehat.*, vol. 4, no. 1, pp. 31–38, 2020, doi: 10.22435/jpppk.v4i1.3312.
- [14] Y. S. Romadhona and K. N. Siregar, "Analisis Sebaran Tenaga Kesehatan Puskesmas di Indonesia berdasarkan Peraturan Menteri Kesehatan Nomor 75 Tahun 2014 tentang Puskesmas," *J. Kesehat. Manarang*, vol. 4, no. 2, p. 114, 2018, doi: 10.33490/jkm.v4i2.99.
- [15] R. Setiawan, "Apa itu Data Mining dan Bagaimana Metodenya?" <https://www.dicoding.com/blog/apa-itu-data-mining/>
- [16] S. Jaya, "Perancangan Basis Data," *Penerbit Andi, Yogyakarta*, 2017.
- [17] William Assaf MSFT, "Create a stored procedure," 2022. <https://docs.microsoft.com/en-us/sql/relational-databases/stored-procedures/create-a-stored-procedure?view=sql-server-ver16>
- [18] R. Rudy, "Perancangan Dimensional Model dan Aplikasi Dashboard Bagi Unit Karir Perguruan Tinggi," *ComTech Comput. Math. Eng. Appl.*, vol. 4, no. 2, p. 677, 2013, doi: 10.21512/comtech.v4i2.2493.
- [19] "No Title." <https://opendata.jabarprov.go.id>
- [20] W. Widhiarso, "Pengategorian Data dengan Menggunakan Statistik Hipotetik dan Statistik Empirik Dampak Penggunaan Referensi Sebuah Tes Dua Strategi Pengategorian Data Perbedaan Kedua Strategi," pp. 1–3, 2014, [Online]. Available: <http://widhiarso.staff.ugm.ac.id/>
- [21] A. Targowski, "From Data to Wisdom," *Dialogue and Universalism*, vol. 15, no. 5, pp. 55–71, 2005, doi: 10.5840/du2005155/629.
- [22] J. Suntoro, "Data Mining," *Min. Massive Datasets*, vol. 2, no. January 2013, pp. 5–20, 2005, [Online]. Available: [https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book\\_part](https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part)
- [23] G. G. Rapakov, A. A. Sukonshchikov, A. N. Shvetsov, V. A. Gorbunov, and O. J. Kravets, "Applied aspects of data mining for decision support at the regional health system," *J. Phys. Conf. Ser.*, vol. 2094, no. 3, 2021, doi: 10.1088/1742-6596/2094/3/032006.