

PENERAPAN DATA TRANSFORMATION PADA DATABASE SISTEM INFORMASI MANAJEMEN RUMAH SAKIT

Gaswara Sani Guspa Prawira¹, Hari Setiaji²

Program Studi Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia

e-mail: ¹16523115@students.uii.ac.id, ²hari.setiaji@uui.ac.id

ABSTRAK

Dengan tema data mining, penelitian ini memiliki fokus tentang salah satu tahapan preprocessing, yaitu data transformation. Penelitian ini merupakan penelitian lanjutan dari penelitian yang dilakukan oleh Ainayya tentang data cleaning. Penelitian ini dilakukan sebagai proses akhir dari preprocessing agar data yang sudah dibersihkan pada data cleaning dapat ditransformasikan sehingga siap untuk dilakukan proses mining. Sampel dan sumber data dari penelitian ini adalah database dari rumah sakit yang ada di Jawa Tengah, Indonesia. Pada tahapan data transformation, mencari operasi-operasi yang dapat diterapkan pada database rumah sakit tersebut. Operasi yang diimplementasikan yaitu: Smoothing, Attribute Construction, Aggregation, Normalization, dan Discretization. Hasilnya penelitian ini menunjukkan bahwa operasi yang sangat disarankan pada database sistem rumah sakit adalah aggregation.

Kata Kunci: data transformation, preprocessing, data mining.

1. PENDAHULUAN

Dengan berkembangnya sistem informasi yang semakin pesat dan tingkat persaingan yang semakin ketat dalam usaha rumah sakit menyebabkan adanya daya saing bagi rumah sakit agar memenangkan persaingan yang ada [1]. Salah satu bentuk persaingan yang diciptakan adalah kualitas layanan. Rumah sakit harus berupaya meningkatkan kualitas jasa pelayanannya secara terus menerus. Para eksekutif rumah sakit menyadari bahwa data-data yang selama ini disimpan pada database dapat dikembangkan/diolah lagi menjadi data yang dapat membantu para eksekutif rumah sakit dalam mengambil keputusan secara cepat dan tepat demi mencapai tujuan bisnis mereka. Para eksekutif memerlukan teknologi masa kini yang dapat menganalisis dan mengelola data menjadi informasi yang berkualitas secara *real time* sehingga data tersebut digunakan untuk meningkatkan kualitas pengambilan keputusan. Teknologi yang menjadi solusi dari masalah tersebut adalah data mining.

Pada proses data mining memiliki beberapa tahapan, yaitu proses *preprocessing*, proses *mining*, evaluasi pola (*pattern evaluation*), dan presentasi pengetahuan (*knowledge presentation*) [2]. Pada tahapan awal data mining terdapat proses processing yang terdiri dari pembersihan data (*data cleaning*), integrasi data (*data integration*), seleksi data (*data selection*), dan transformasi data (*data transformation*). Tahapan ini sangat penting karena kualitas dari data yang masuk pada proses mining mempengaruhi kualitas output analisis yang dihasilkan [3].

Pada penelitian ini memiliki fokus pada proses data transformation dan juga merupakan penelitian lanjutan dari penelitian yang dilakukan oleh Ainayya tentang *data cleaning*. Oleh karena itu, penelitian ini merupakan proses akhir dari tahapan *preprocessing* agar data yang sudah dibersihkan pada proses *data cleaning* dapat ditransformasikan sehingga siap untuk dilakukan proses mining. Sampel dan sumber data yang digunakan adalah database dari sistem informasi manajemen rumah sakit yang berada di Jawa Tengah, Indonesia. Data dari *database* tersebut akan dilakukan pencarian operasi yang memungkinkan diterapkan pada data tersebut.

2. TINJAUAN PUSTAKA

2.1 Data Transformation pada Data Mining

Penelitian ini dilakukan oleh Junaedi dkk pada tahun 2011 dengan judul “Data Transformation pada Data Mining”. Pada penelitian ini membahas beberapa pendekatan yang terdapat pada data transformation dan melakukan analisis metode yang dimiliki untuk masing-masing pendekatan tersebut. Pendekatan atau operasi transformasi yang dibahas di sini, yaitu *Smoothing (Binning, Clustering, dan Regression)*, *Generalization (Histogram Analysis, Entropy-Based Discretization, dan Segmentation by Natural Partitioning)*, *Normalization (Min-max Normalization, Z-Score Normalization, dan Normalization by Decimal Scaling)*, *Aggregation (Roll-up)*, dan *Attribute Construction*.

2.2 Cross-company Customer Churn Prediction (CCCP) in Telecommunication: A Comparison of Data Transformation Methods

Penelitian ini dilakukan oleh Amin dkk pada tahun 2018 dengan judul “Cross-Company Customer Churn Prediction (CCCP) in Telecommunication: A Comparison of Data Transformation Methods”. Tujuan dari penelitian ini untuk memprediksi presentase pelanggan dengan menggunakan data dari perusahaan lain sebagai

sumber. Pada penelitian ini, mencari metode transformasi data pada kinerja model CCCP yang efektif. Metode yang diujicobakan adalah *log*, *z-score*, *rank*, dan *box-cox*. Hasilnya menunjukkan bahwa sebgaiian metode transformasi data dapat meningkatkan kinerja CCCP secara signifikan. Namun, metode transformasi data *z-score* tidak dapat mencapai hasil yang lebih baik dibandingkan dengan metode transformasi data lainnya dalam penelitian ini [4].

3. METODE PENELITIAN DAN KAKAS ANALISIS

3.1 Data Transformation

Data Transformation adalah tahapan di mana data ditransformasikan dan dikonsolidasikan ke dalam bentuk yang sesuai untuk *mining* [5]. Menurut Junaedi dkk. (2011), dalam data transformation, terdapat beberapa operasi/teknik untuk melakukan transformasi data, yaitu *smoothing*, *attribute construction*, *normalization*, *aggregation*, dan *discretization*.

3.1.1 Smoothing

Operasi *Smoothing* digunakan untuk mengatasi data bersifat *noise*/nilai yang tidak valid untuk proses *mining* dengan memperhatikan nilai-nilai tetangga. Salah satu metode yang akan digunakan adalah *Clustering*.

Clustering adalah proses mempartisi atau mengelompokkan serangkaian pola yang diberikan ke dalam cluster yang terpisah [6]. Metode *Clustering* berguna untuk menghilangkan *outliers/noise* (data yang terlalu menyimpang jauh dari data lainnya). Algoritma yang dipakai adalah k-Means.

3.1.2 Attribute Construction

Pada *attribute construction*, mengkontruksi atau menambahkan atribut baru untuk meningkatkan ketelitian/ketepatan proses *mining*.

3.1.3 Normalization

Normalization adalah proses pengelompokan atribut ke dalam hubungan yang terstruktur dengan baik dan bebas dari anomali [7]. *Normalization* digunakan untuk mentransformasi sebuah atribut numerik diskalakan dalam *range* yang lebih kecil seperti -1.0 sampai 1.0. Teknik yang digunakan untuk operasi ini adalah *Z-Score Normalization*.

3.1.4 Aggregation

Aggregation merupakan operasi untuk *summary* (peringkasan) yang digunakan untuk data numerik dengan menggunakan operasi *roll up*.

3.1.5 Discretization

Discretization adalah digunakan untuk mereduksi sekumpulan nilai yang terdapat pada atribut *continuous*, dengan membagi *range* dari atribut ke dalam interval. Operasi yang digunakan dalam *discretization* adalah *Binning*.

3.2 Tool

Alat Alat yang digunakan untuk membantu dalam menjalankan operasi-operasi yang telah disebutkan di atas adalah RapidMiner. RapidMiner adalah sistem yang mendukung desain dan dokumentasi dari keseluruhan proses penambangan data [8]. Alat ini menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran dalam, penambangan teks, dan analisis prediktif. Hal ini digunakan untuk bisnis dan komersial, juga untuk penelitian, pendidikan, pelatihan, *rapid prototyping*, dan pengembangan aplikasi serta mendukung semua langkah dalam proses pembelajaran mesin termasuk persiapan data, hasil visualisasi, validasi model, dan optimasi.

3.3 Sumber Data

Sumber data berasal dari *database* sistem informasi manajemen rumah sakit di Jawa Tengah, Indonesia. *Database* tersebut adalah *database* Oracle versi 10.2.0.1.0 dengan besar + 300 MB. Pada *database* tersebut terdapat 161 tabel tetapi tabel yang digunakan hanya yang berkaitan dengan pasien sehingga menyisakan 14 tabel.

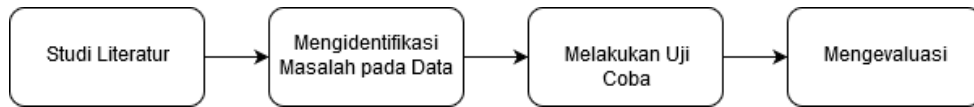
Tabel 1. Tabel *Database* Rumah Sakit

ASURANSI KEPESERTAAN VISIT	LOG BATAL KUNJUNGAN
ANTRIAN	PASIEN
RAWAT JALAN	HASIL PEMERIKSAAN LAB
BILLING	PEMBAYARAN

RAWAT DARURAT	PENDUDUK
KUNJUNGAN	PENJUALAN RESEP
KUNJUNGAN BPJS	VISIT

3.4 Tahapan Data Transformation

Pada tahap ini akan dilakukan analisis dan perancangan penelitian sebagai berikut:



Gambar 30. Tahapan Data Transformation

4. HASIL DAN PEMBAHASAN

A. Tabel ASURANSI_KEPESERTAAN_VISIT

NO_POLIS	ID	ID_VISIT	ID_ASURANSI_PRODUK
83619797	113466	72733	6
542373693	113467	72734	3
83803408	113468	72735	5

Gambar 31. Tabel asuransi_kepesertaan_visit

Kolom:

-id_asuransi_produk: number (foreign key)

Pada kolom ini dapat diketahui bahwa setiap id_visit yang masuk akan mendapatkan id_asuransi_produk sesuai kriteria masing-masing. Oleh karena itu, di sini dapat dilakukan operasi agregasi untuk mencari nilai dari id_asuransi_produk yang sering muncul.

B. Tabel BILLING

ID	ID_VISIT	WAKTU	ID_PEGAWAI_PETUGAS	STATUS PEMBAYARAN	TOTAL_TAGIHAN	JENIS
17006	8818	21-JUL-16 07...	171	1	102900	RJ
17007	8819	21-JUL-16 07...	171	1	81300	RJ
17008	8820	21-JUL-16 07...	171	1	87300	RJ

Gambar 32. Tabel billing

Kolom:

-waktu: timestamp

Setiap memasukkan data secara otomatis akan dicatat waktunya. Kolom ini cocok untuk dilakukan agregasi untuk mengelompokkan waktu yang lebih spesifik atau untuk mengetahui kapan waktu penginputan data yang paling ramai.

-id_pegawai_petugas: number (foreign key)

Kolom ini menjelaskan bahwa setiap menginput data maka akan tercatat identitas petugas yang bertanggung jawab. Dalam kolom ini dapat diagregasikan untuk melihat nilai atribut id_pegawai_petugas yang paling banyak.

-status_pembayaran: varchar2

Kolom ini berguna untuk memperlihatkan status pembayaran pada pasien apakah lunas atau belum. Dengan dilakukannya agregasi maka akan terlihat perbandingan antar nilai status pembayaran yang ada.

-total_tagihan: float

Terlihat bahwa kolom ini memuat total tagihan pasien yang harus dibayarkan sehingga sangat cocok untuk dilakukan penjumlahan seluruh nilai atribut kolom total_tagihan yang ada dengan agregasi.

-jenis: varchar2

Berisi tentang jenis layanan pasien apakah rawat jalan atau rawat inap. Kolom ini sama dengan kolom status_pembayaran, dapat diagregasikan untuk mencari perbandingan antar nilai atribut.

C. Tabel HASIL_PEMERIKSAAN_LAB

ID	TANGGAL	ID_VISIT
11502	14-OCT-16 11.55.12.000000 AM	20223
11503	14-OCT-16 12.02.11.000000 PM	20223
11506	14-OCT-16 12.31.25.000000 PM	20701

Gambar 33. Tabel hasil_pemeriksaan_lab

Kolom:

-**tanggal**: timestamp

Berisi tentang waktu terjadinya proses penginputan. Dilihat dari jenisnya kolom ini dapat dioperasikan untuk mencari waktu yang paling ramai dalam proses input.

D. Tabel KUNJUNGAN

ID	WAKTU	ID_PASIE	ID_LAYANAN	ID_PENDUDUK_DPJP	ID_PENDUDUK_P...	RENCANA_C...	STATUS
30667	25-JUN-16 0...	45595	1081	72994	147827	Bayar Sendiri	Masuk
30666	25-JUN-16 0...	9542	1081	72994	69225	Charity	Masuk
29078	16-JUN-16 0...	43448	1081	74807	135240	Charity	Masuk

Gambar 34. Tabel kunjungan

Kolom:

-**waktu**: timestamp

Sama seperti kolom waktu sebelumnya, berisi tentang pencatatan waktu pada proses penginputan. Oleh karena itu, kolom ini dapat dioperasikan untuk mencari kapan proses penginputan sering terjadi.

-**id_layanan**: number (*foreign key*)

Kolom ini berisi tentang layanan yang didapat oleh pasien. Jadi, kolom ini dapat dioperasikan untuk dapat melihat nilai atribut id_layanan yang paling banyak.

-**rencana_cara_bayar**: varchar2

Dapat dilihat bahwa kolom ini menjelaskan setiap pasien yang terdaftar dapat memilih cara pembayarannya. Untuk kasus ini, hal yang paling tepat adalah melakukan agregasi untuk melihat perbandingan cara pembayarannya.

-**status**: varchar2

Dari kolom ini dapat diketahui bahwa setiap pasien memiliki status yang berbeda-beda. Agregasi yang dilakukan di sini untuk melihat perbandingan status pada para pasien.

E. Tabel KUNJUNGAN BPJS

ID	ID_POLI_SEP	KODE_JENI...	NAMA_JENI...	KELAS_BPJS	NAMA_PRO...	TANGGAL_R...	ID_DIAGNOS...
2	13	22	PBI (APBD)	KELAS III	PEJAGOAN	12-FEB-18 0...	9925
3	12	3	PNS DAERAH	KELAS II	DR. R. SUNA...	12-FEB-18 0...	7583
4	8	22	PBI (APBD)	KELAS III	PEJAGOAN	19-FEB-18 1...	9925

Gambar 35. Tabel kunjungan_bpjs

Kolom:

-**id_pol_sep**: number (*foreign key*)

Kolom ini menjelaskan setiap pasien memiliki id_pol_sep-nya masing-masing. Agregasi yang dilakukan adalah untuk melihat nilai atribut id_pol_sep yang paling banyak sebagai perbandingan.

-**kode_jenis_bpjs**: number

Berisi tentang kode jenis bpjs yang digunakan oleh tiap pasien. Agregasi yang diterapkan untuk melihat perbandingan antar nilai atribut.

-**nama_provider**: varchar2

Terlihat bahwa kolom ini berisi nama provider si pasien. Di sini dapat diagregasikan untuk melihat perbandingan antar nilai atribut.

-id_diagnosa_awal: number (*foreign key*)

Kolom ini berisi kode diagnosis tiap pasien. Data ini dapat dioperasikan agregasi untuk melihat jumlah pasien menurut diagnosisnya.

5. KESIMPULAN

Di antara 5 operasi *data transformation* yang diujicobakan, sangat direkomendasikan operasi *aggregation*. Alasannya adalah data tersebut tidak memenuhi syarat untuk dilakukannya keempat operasi selain *aggregation*. Setiap tabel memiliki data dengan tipe data yang berbeda-beda dan banyak terdapat *foreign key* sehingga hanya dapat ditransformasikan untuk mengetahui jumlah keseluruhan (*summary*), perbandingan, dan presentase data pada kolom. Hal ini juga mendukung Peraturan Menteri Kesehatan Tentang Sistem Informasi Manajemen Rumah Sakit pasal 4 ayat 2 tentang sistem layanan dapat meningkatkan kecepatan pengambilan keputusan, akurasi, dan kecepatan identifikasi masalah dan kemudahan dalam penyusunan strategi dalam pelaksanaan manajerial.

DAFTAR PUSTAKA

- [1] K. Gunawan, "Kualitas Layanan dan Loyalitas Pasien (Studi pada Rumah Sakit Umum Swasta di Kota Singaraja – Bali)," 2002.
- [2] M. Ridwan, H. Suyono, and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Eecis*, vol. 7, no. 1, pp. 59–64, 2013.
- [3] H. Junaedi, H. Budianto, I. Maryati, and Y. Melani, "Data Transformation pada Data Mining," *Pros. Konf. Nas. Inov. dalam Desain dan Teknol.*, vol. 7, pp. 93–99, 2011.
- [4] A. Amin *et al.*, "Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods," *Int. J. Inf. Manage.*, vol. 46, no. September 2018, pp. 304–319, 2019.
- [5] J. Han, M. Kamber, J. Pei, and M. Kaufmann, "[DATA MINING: CONCEPTS AND TECHNIQUES 3RD EDITION] 2 Data Mining: Concepts and Techniques Third Edition," 2012.
- [6] K. Alsabti, "Electrical Engineering and Computer Science An efficient k-means clustering algorithm," 1997.
- [7] H. Lee, "Justifying database normalization: a cost/benefit model," *Inf. Process. Manag.*, vol. 31, no. 1, pp. 59–67, 1995.
- [8] Markus Hofmann Ralf Klinkenberg, *Rapid Miner Data Mining Use Cases and Business Analytics Applications*. 2009.