

Deteksi *Outlier* Transaksi Menggunakan Visualisasi-Olap Pada Data Warehouse Perguruan Tinggi Swasta

Outlier Detection of Transaction Using Olap-Visualization On The Data Warehouse of University

Gusti Ngurah Mega Nata

STIKOM BALI; Jln Raya Pupuran-Renon, telp/fax : (0361) 244445

Jurusan Sistem Informasi, STIKOM Bali, Bali

e-mail: mega@stikom-bali.ac.id

Abstrak

Mendeteksi outlier pada data warehouse merupakan hal penting. Data pada data warehouse sudah diagregasi dan memiliki model multidimensional. Agregasi pada data warehouse dilakukan karena data warehouse digunakan untuk menganalisis data secara cepat pada top level manajemen. Sedangkan, model data multidimensional digunakan untuk melihat data dari berbagai dimensi objek bisnis. Jadi, Mendeteksi outlier pada data warehouse membutuhkan teknik yang dapat melihat outlier pada data yang sudah diagregasi dan dapat melihat dari berbagai dimensi objek bisnis. Mendeteksi outlier pada data warehouse akan menjadi tantangan baru. Di lain hal, Visualisasi On-line Analytic process (OLAP) merupakan tugas penting dalam menyajikan informasi trend (report) pada data warehouse dalam bentuk visualisasi data. Pada penelitian ini, visualisasi OLAP digunakan untuk deteksi outlier transaksi. Maka, dalam penelitian ini melakukan analisis untuk mendeteksi outlier menggunakan visualisasi-OLAP. Operasi OLAP yang digunakan yaitu operasi drill-down. Jenis visualisasi yang akan digunakan yaitu visualisasi satu dimensi, dua dimensi dan multi dimensi menggunakan tool weave desktop. Pembangunan data warehouse dilakukan secara button-up. Studi kasus dilakukan pada perguruan tinggi swasta. Kasus yang diselesaikan yaitu mendeteksi outlier transaksi pembayaran mahasiswa pada setiap semester. Deteksi outlier pada visualisasi data menggunakan satu tabel dimensional lebih mudah dianalisis dari pada deteksi outlier pada visualisasi data menggunakan dua atau multi tabel dimensional. Dengan kata lain semakin banyak tabel dimensi yang terlibat semakin sulit analisis deteksi outlier yang dilakukan.

Kata kunci — Deteksi Outlier, Visualisasi OLAP, Data warehouse

Abstract

Detecting outliers in the data warehouse is important. Data in the data-warehouse has been aggregated and is a multidimensional models. Aggregation of the data warehouse for reporting quickly on top management level. Meanwhile, multidimensional modeling to view the data from different dimensions of business objects. So, Detect outliers in the data warehouse requires a technique that can be viewed outlier in the data that has been aggregated and can be viewed from different dimensions of business objects. Detecting outliers in the data warehouse will be a new challenge. On the other hand, Visualization On-line Analytical Process (OLAP) is an important task in presenting trend information (reporting) in the data warehouse. In this study, visualization OLAP is used for outlier detection transaction. Thus, in this study conducted an analysis to detect outliers using OLAP visualization. OLAP techniques were used that drill-down techniques. Data visualization is built from a single to multi-dimensional tables. visualization built using weave-desktop. Data warehouse development conducted button-up. The case studies conducted at universities. The problems solved are detected outlier payment

transactions student in each semester. Outlier detection in data visualization using one dimensional tables more easily analyzed than the outlier detection in data visualization using a two- or multi-dimensional tables. In other words, the more dimension tables that are involved the more difficult analysis of outlier detection is performed.

Keywords — *Outlier detection, OLAP Visualization, Data warehouse*

1. PENDAHULUAN

Data warehouse secara fisik adalah database yang menyimpan *history* data transaksi dalam model *multidimensional* dalam volum data yang besar [1][2]. Model *Multidimensional* adalah model database denormalisasi yang membagi tabel dalam dua tipe yaitu tabel fakta (*fact table*) dan tabel dimensi (*dimensional table*) [3]. Proses Menemukan informasi / *trend* yang menarik dari data warehouse yaitu dengan cara melihat dari berbagai dimensi data yang disebut *Online Analysis processing (OLAP)* [4]. Proses OLAP secara langsung menganalisis data multidimensional menggunakan bahasa *standard query language (SQL)* [4].

Dengan perkembangan penelitian teknologi informasi penggunaan visualisasi data untuk menampilkan hasil OLAP sudah banyak dilakukan, atau dikenal dengan Visualisasi-OLAP[5]. Namun, Visualisasi-OLAP umumnya digunakan untuk menampilkan report transaksi dalam bentuk visualisasi. Dalam data warehouse selain informasi trend, hal penting yang seharusnya dapat dilakukan yaitu deteksi *outlier* transaksi. Maka, dalam penelitian ini melakukan analisis untuk mendeteksi *outlier* menggunakan visualisasi-OLAP. Operasi OLAP yang digunakan yaitu operasi *drill-down*. Jenis visualisasi yang digunakan yaitu visualisasi satu dimensi, dua dimensi dan multi dimensi menggunakan tool *weave desktop*. Pembangunan data warehouse dilakukan secara *button-up*. Dengan mengambil sumber data yang tersedia yang selanjutnya digunakan sebagai sumber data warehouse. Subjek dari data warehouse yang dibangun yaitu transaksi pembayaran, dan dimensi yang digunakan bersumber dari keuangan, akademik dan data diri mahasiswa. Studi kasus dilakukan pada perguruan tinggi swasta yaitu STMIK STIKOM Bali. Kasus yang diselaikan yaitu mendeteksi *outlier* transaksi pembayaran mahasiswa pada setiap semester.

2. METODE PENELITIAN

Untuk melakukan analisis dalam mendeteksi outlier transaksi pembayaran mahasiswa pada data warehouse menggunakan visualisasi data, memiliki beberapa tahapan. Mulai dari studi pustaka (subbab 2.1), pengumpulan data (subbab 2.2), preprocessing sekaligus Ekstrak Transform Load (ETL) (subbab 2.3), Pemodelan Multidimensional (subbab 2.4), Online Analytic Processing (OLAP) (subbab 2.5) dan terakhir visualisasi data (subbab 3.6). Berikut adalah penjelasan dari masing – masing tahapan penelitian atau metode penelitian.

2.1 Studi Pustaka

2.1.1 Deteksi Outlier

Deteksi outlier juga dikenal sebagai deteksi anomaly merupakan proses mencari data objek dengan perilaku yang sangat berbeda dari keseluruhan data [4][6]. Objek data yang memiliki perilaku yang berbeda dari data yang kebanyakan tersebut disebut outlier atau anomaly. Deteksi outlier sangat penting diterapkan pada aplikasi deteksi penipuan, kesalahan perawatan medis, keamanan public, deteksi kerusakan industry, pengolahan gambar, pengawasan sensor, jaringan dan deteksi instruksi [4]. Deteksi outlier dan analisis klustering adalah dua *task* data mining yang sangat terkait. Pada lain hal, disebutkan bahwa data yang tidak outlier disebut data normal/inlier sedangkan data outlier disebut data abnormal. Pada umumnya jumlah data outlier lebih sedikit dibandingkan dengan data inlier/normal. Jika, probabilitas data outlier satu berbanding seribu data inlier/normal, maka akan berjumlah seribu data outlier jika jumlah data sudah mencapai satu juta.

2.1.2 Data Warehouse

Menurut wiliam H Imon, data warehouse merupakan sebuah subject-oriented, integrated, nonvolatile, dan time-variant dari data untuk mendukung keputusan manajemen [2]. Sebuah data warehouse merupakan database relasional yang didesain untuk query dan analisis dibandingkan untuk proses transaksi [3]. Data warehouse dibangun dengan mengintegrasikan data dari berbagai sumber yang heterogen [7]. Dan, secara fisik data warehouse adalah database yang menyimpan history data transaksi dalam model multidimensional dalam volum data yang besar [8]. Model Multidimensional adalah model database denormalisasi yang membagi tabel dalam dua tipe yaitu tabel fakta (fact table) dan tabel dimensi (dimensional table) [4]. Tujuan dibangunnya data warehouse yaitu untuk pelaporan dan analisis data untuk menemukan trend [6] [3]. Proses Menemukan informasi / trend yang menarik dari data warehouse yaitu dengan cara melihat dari berbagai dimensi data yang disebut Online Analysis processing (OLAP) [9]. Penjelasan lebih lanjut tentang Multidimensional dan OLAP disusun dalam subbab 2.1.1 dan 2.1.2 berikut.

2.1.3 Online Analytic Processing (OLAP)

Untuk sistem OLAP, waktu respon merupakan ukuran efektivitas. OLAP digunakan untuk menganalisis data multidimensi dari berbagai sumber dan perspektif. Tiga operasi dasar di OLAP adalah : *Roll-Up* (konsolidasi), *Drill-Down* dan *slicing & dicing* [4].

2.1.4 Visualisasi Data

Visualisasi data merupakan proses pemetaan data ke dalam bentuk visual [10]. Tujuan utama dari visualisasi data adalah untuk mengkomunikasikan informasi secara jelas dan efisien kepada pengguna lewat grafik informasi yang dipilih, seperti tabel dan grafik. Visualisasi yang efektif membantu pengguna dalam menganalisa dan penalaran

tentang data dan bukti. Ia membuat data yang kompleks bisa diakses, dipahami dan berguna.

2.2 Pengumpulan dan karakteristik Data

Data transaksi pembayaran dari perguruan tinggi sudah terkumpul. Data dalam bentuk format relasional. Data terdiri dari data transaksi pembayaran mahasiswa dari tahun 2010 sampai dengan data transaksi pembayaran 2016. Berikut pada Gambar 1 adalah tampilan data transaksi yang telah dikumpulkan:

STR_THN	BO	TGL_TRS	STR_NIM	STR	NUM_SU	NUM	NU	NUM_TOTAL	NUM_BAYAR	NUM_SISA
1	2005/2006	0	2005-09-16	050020...	01	1425000	0	1425000	713000	712000
2	2005/2006	0	2005-09-05	020010...	07	125000	0	125000	125000	0
3	2005/2006	0	2005-22-11	040020...	03	1150000	0	1150000	1150000	0
4	2005/2006	0	2005-09-15	020010...	07	1000000	0	1000000	1000000	0
5	2010/2011	0	2010-11-30	080010...	05	630000	0	630000	600000	30000
6	2010/2011	0	2010-12-13	070020...	07	750000	0	750000	500000	250000
7	2005/2006	0	2005-09-01	020010...	07	125000	0	125000	125000	0
8	2010/2011	0	2010-08-20	080010...	05	2015000	0	2015000	1135000	880000
9	2005/2006	0	2005-09-01	020010...	07	125000	0	125000	125000	0
10	2008/2009	0	2008-08-06	01	5380000	475000	0	4905000	4905000	0
11	2013/2014	0	2013-06-27	110030...	05	400000	0	400000	600	399400
12	2011/2012	1	2012-04-21	080010...	08	395000	0	395000	395000	0
13	2005/2006	0	2005-08-31	020010...	07	125000	0	125000	125000	0
14	2005/2006	0	2005-09-15	020010...	07	1050000	0	1050000	600000	450000
15	2005/2006	0	2005-09-07	030010...	05	155000	0	155000	155000	0
16	2015/2016	1	2016-01-04	140030...	04	1920000	0	1920000	1920000	0
17	2005/2006	0	2005-09-01	020010...	07	125000	0	125000	125000	0
18	2005/2006	0	2005-06-22	030010...	05	975000	0	975000	25000	950000
19	2006/2007	0	2006-08-08	050010...	03	250000	0	250000	250000	0
20	2012/2013	0	2012-07-21	090010...	07	1324988	0	1324988	1225000	99988
21	2005/2006	0	2005-07-01	030010...	05	155000	0	155000	155000	0
22	2005/2006	0	2005-09-05	020010...	07	125000	0	125000	125000	0
23	2005/2006	0	2005-09-05	020010...	07	125000	0	125000	125000	0

Gambar 1 Data transaksi pembayaran semesteran

Pada gambar 1 Menunjukkan isi dari data transaksi pembayaran. Field yang dipilih yaitu tahun ajaran, jenis semester, tanggal transaksi, nim mahasiswa, semester mahasiswa, total pembayaran, jumlah potongan, jumlah yang harus dibayar, jumlah pembayaran mahasiswa dan sisa / kekurangan pembayaran mahasiswa. Jumlah data pembayaran yang digunakan yaitu 181700 rows.

Data master mahasiswa yang digunakan mulai dari angkatan 2010 sampai angkatan 2016. Jumlah data mahasiswa yang digunakan yaitu 6983. Berikut pada Gambar 2 adalah data master mahasiswa yang digunakan:

str_nim	str_angkatan	bol_jk	bol_sts_nkh	str_kd_agm	bol_sts_bayar	str_sts_mhs	bol_mhs_transfer	
1	100010...	2010	1	0	4	0	A	0
2	100010...	2010	1	0	4	0	A	0
3	100010...	2010	1	0	1	0	A	1
4	100010...	2010	1	0	1	0	A	0
5	100010...	2010	1	0	4	0	A	0
6	100010...	2010	1	0	4	0	A	0
7	100010...	2010	1	0	4	0	A	0
8	100010...	2010	0	0	4	0	A	0
9	100010...	2010	1	0	1	0	A	0
10	100010...	2010	1	0	2	0	A	0
11	100010...	2010	1	0	1	0	A	0
12	100010...	2010	1	0	4	0	A	0
13	100010...	2010	1	0	1	0	A	0
14	100010...	2010	1	0	1	0	A	1
15	100010...	2010	1	0	4	0	A	0
16	100010...	2010	1	0	4	0	A	0
17	100010...	2010	1	0	4	0	A	1
18	100010...	2010	1	0	4	0	A	0
19	100010...	2010	1	0	4	0	A	0
20	100010...	2010	0	0	4	0	A	0
21	100010...	2010	1	0	4	0	A	0
22	100010...	2010	1	0	4	0	A	0

Query executed successfully. (local) (10.50 RTM) | psi-002\user (53) | ETL | 00:00:00 | 6983 rows

Gambar 2 data master mahasiswa

Field yang dipilih pada data master mahasiswa yaitu NIM (*str_nim*), angkatan (*str_angkatan*), jenis kelamin (*bol_jk*), status menikah (*bol_sts_nkh*), agama (*str_kd_agm*), status mahasiswa (*str_sts_mhs*), status mahasiswa (*bol_mhs_transfer*). Field jenis kelamin berisi 2 jenis values yaitu 1 (laki – laki), 0 (perempuan). Field Agama berisi 5 values dimana setiap angka mewakili 1 agama. Status bayar merupakan jenis pembayaran mahasiswa.

2.3 Preprocessing dan ETL

Proses preprocessing yaitu mempersiapkan data yang bersumber dari data transaksi sehari – hari yaitu OLTP kedalam data warehouse. Proses ini dikenal dengan istilah ETL (*Extrac transform dan load*). Berikut adalah penjelasan secara bertahap dari setiap proses yang dilakukan pada tugas ETL:

1. Extrac

Data yang diambil / ekstrak yaitu data transaksi pembayaran sebagai tabel fakta. Tabel mahasiswa sebagai tabel dimensi mahasiswa, dan informasi tahun ajaran serta informasi waktu masing – masing sebagai tabel dimensi tahun ajaran dan dimensi waktu. Data

2. Transform

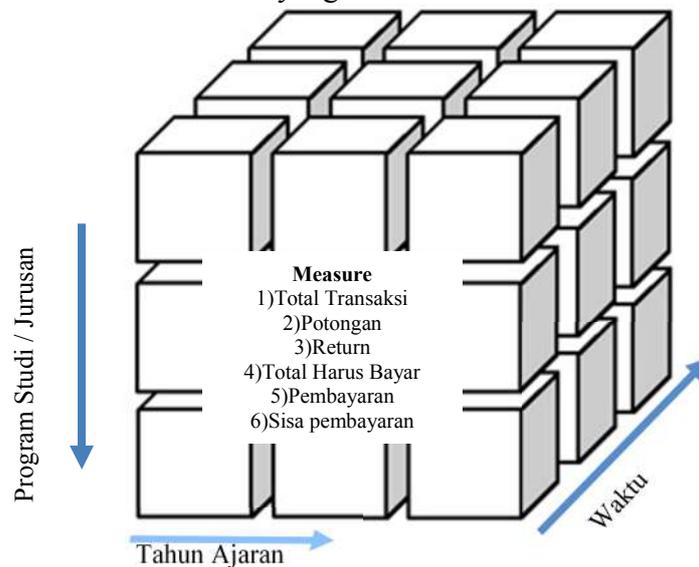
Agar format data yang dimasukkan ke data warehouse harus seragam dan membentuk model multidimensional maka proses transformasi perlu dilakukan. Pada tahap ini beberapa tabel dimensi yang ditransformasi yaitu sebagai berikut:

1. Tabel *dim_tahun_ajaran* adalah hasil tranformasi dari field *tahun_ajaran*, boolean semester (genap/ganjil) dan field semester romawi pada tabel transaksi pembayaran dan data master mahasiswa.

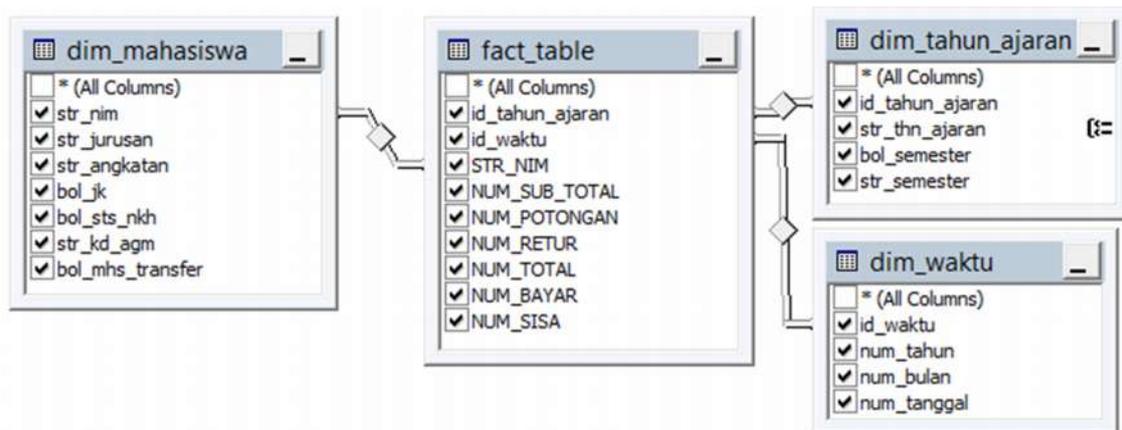
2. Tabel `dim_waktu` adalah hasil tranformasi dari field tanggal pembayaran mahasiswa pada tabel transaksi pembayaran yang dibagi menjadi `num_tahun`, `num_bulan` dan `num_tanggal`.
3. Load
Proses ini memuat data yang telah dipreprocessing dan ditranform kedalam model multidimensional. Setelah proses Load data warehouse sudah siap untuk dilakukan aktivitas OLAP dan visualisasi data.

2.4 Desain Multidimensional

Pekerjaan besar dalam tranformasi database relational menjadi model data warehouse yaitu membagi data kedalam tabel dimensi (dimensional table) dan tabel fakta (fact table) [3]. Dimensi merupakan entitas dasar yang stabil di lingkungan kita. Fakta merupakan pengukuran numerik atau pengamatan yang dikumpulkan oleh semua sistem pemrosesan transaksi dan sistem lainnya. Berikut pada Gambar 3 adalah model multidimensional dari data warehouse yang dianalisis:



Gambar 3 Desain Cube Transaksi Pembayaran

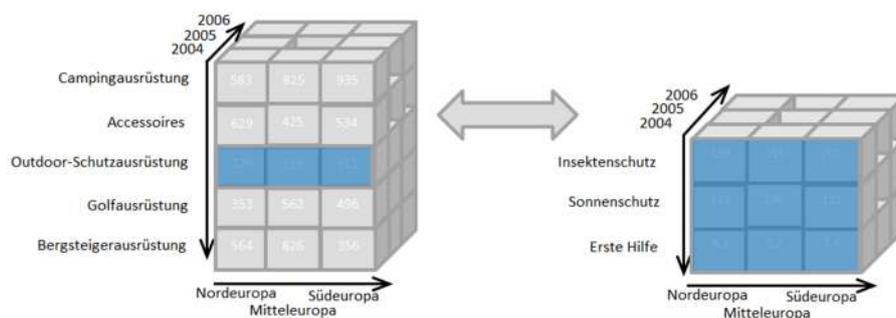


Gambar 4. Start Skema Multidimensional Keuangan Perguruan Tinggi

Model multidimensional dari Gambar 4 merupakan model multidimensional dari data warehouse perguruan tinggi swasta. Pada model tersebut terdapat 3 tabel dimensional yaitu *dim_mahasiswa*, *dim_tahun_ajaran* dan *dim_waktu*. Tabel fakta yang terbentuk hanya 1 yaitu *fact_table*. Field nilai pengukur (*measure*) dari tabel fakta yaitu *num_sub_total*, *num_potongan*, *num_retur*, *num_total*, *num_bayar*, dan *num_sisa*.

2.5 OLAP

Aktivitas OLAP dilakukan untuk menganalisis data didalam model kubus / model multidimensional. Data dalam model kubus tersebut akan diseleksi dan ditampilkan dalam bentuk tabel dua dimensi. Teknik OLAP yang digunakan untuk menampilkan data dari data warehouse yaitu dengan teknik *drill-down*. Teknik *drill-down* akan menampilkan informasi dari data global ke data yang lebih detail. Berikut adalah contoh teknik *drill-down*:



Gambar 5 Drill-down

Pada contoh diatas hirarki / sub dari field *Outdoor-schutztausrüstung* dipilih sehingga cube yang lebih detail muncul yang ditunjukkan pada Gambar 5. Pada penelitian ini proses *drill-down* dilakukan pada beberapa tabel dimensi. Berikut adalah informasi – informasi yang didapat menggunakan teknik *drill down* saat proses OLAP di data warehouse transaksi keuangan pada perguruan tinggi:

1. Menggunakan satu tabel dimensi
 Drill-down dimensi tahun ajaran

- Measure yang digunakan pada table fakta yaitu num_bayar
2. Menggunakan dua table
 - Drill-down dimensi mahasiswa yaitu Program studi
 - Drill-down dimensi tahun ajaran yaitu tahun_ajaran
 - Measure yang digunakan pada tabel fakta yaitu num_bayar dan num_potongan
 3. Program studi → tahun ajaran → semester (ganjil / genap)
 - Measure yang digunakan yaitu num_bayar (jumlah pembayaran), num_potongan (jumlah potongan), num_total (total yang harus dibayar)
 4. Menggunakan tiga tabel dimensi
 - Drill-down dimensi waktu tahun → bulan → tanggal
 - Drill-down dimensi mahasiswa yaitu program studi → Angkatan
 - Drill-down dimensi tahun ajaran yaitu tahun_ajaran → bol_semester (genap/ganjil) → str_semester mahasiswa (Romawi)
 - Measure yang digunakan pada table fakta yaitu total_pembayaran

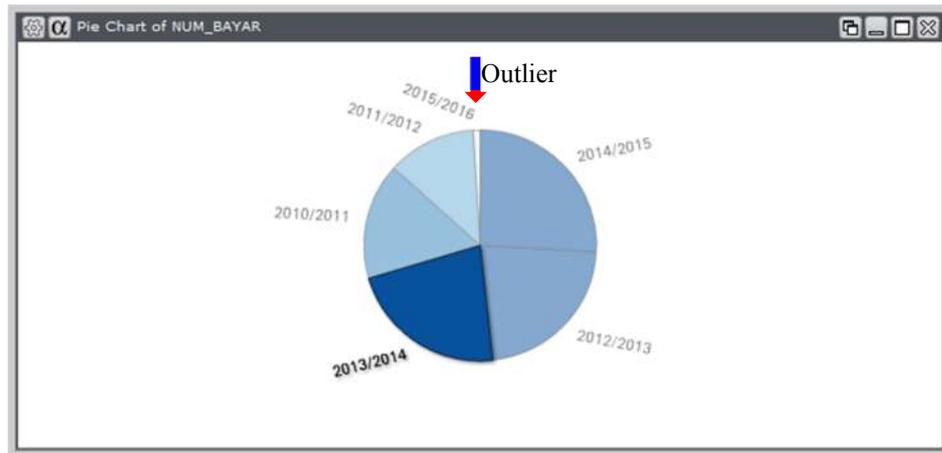
2.6 Visualisasi data

Visualisasi data dibentuk dari model data multidimensional transaksi seperti yang dijelaskan pada subbab 2.4. Sebelum data multidimensional tersebut divisualisasikan terdapat proses OLAP dengan teknik drill-down. Proses OLAP dilakukan bertahap, mulai dari menggunakan satu tabel dimensional, dua dimensional dan multi dimensional seperti yang telah dijelaskan pada subbab 2.5 yaitu proses OLAP. Deteksi outlier sendiri dilakukan pada hasil visualisasi data. Deteksi outlier dilakukan dengan cara melihat dan memahami informasi pada visualisasi. Suatu informasi dari data dikatakan outlier jika visualisasi yang terbentuk tidak merata atau sebagian kecil data divisualisasikan berjauhan dari data yang lain, atau memiliki ukuran, warna, posisi yang jauh berbeda dengan data yang lain.

Dalam penelitian ini jika nilai measure dari suatu hasil OLAP yang visualisasi sangat jauh berbeda dengan nilai measure yang lain maka informasi tersebut dikatakan tidak normal / outlier. Contoh nilai measure yang tidak normal / outlier yaitu seperti jumlah pembayaran terlalu rendah / terlalu tinggi, potongan yang terlalu besar / terlalu rendah atau nilai extream lainnya maka data tersebut dikatakan outlier.

1. Visualisasi satu dimensi

Dimensi yang digunakan pertama untuk melihat outlier pembayaran yaitu dimensi tahun ajaran dengan nilai measure pembayaran (field num_bayar). Field dimensi tahun ajaran yang digunakan yaitu field tahun ajaran saja. Tujuan dari visualisasi ini ingin membandingkan antar tahun ajaran dari tahun ajaran 2010/2011 sampai dengan 2015/2016. Nilai measure yang dibandingkan yaitu total pembayaran mahasiswa pada setiap tahun ajaran. Visualisasi yang dapat digunakan untuk menunjukkan perbandingan antar value satu dimensi yaitu pie chart atau grafik batang. Pada visualisasi yang pertama akan menggunakan pie chart yang ditunjukkan pada Gambar 6 berikut.



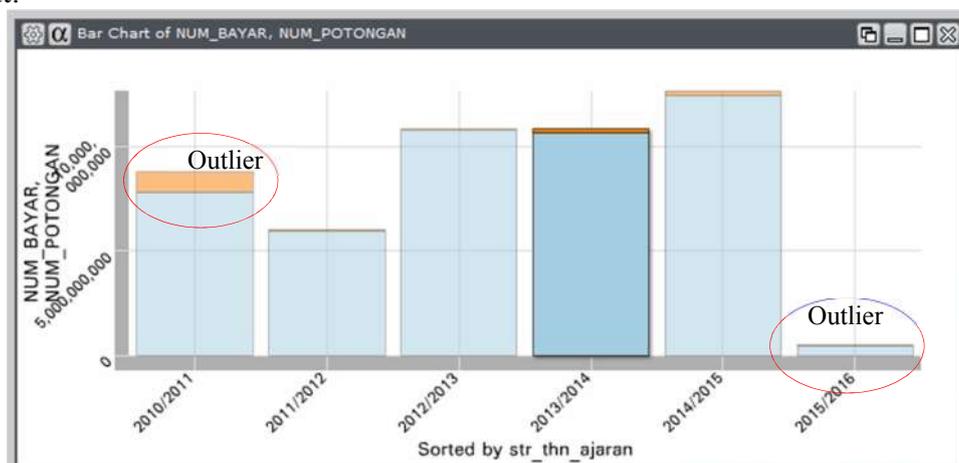
Gambar 6 Visualisasi pie chart dari jumlah pembayaran setiap tahun ajaran

Visualisasi data pie untuk menunjukkan informasi outlier dari jumlah pembayaran setiap tahun ajaran. Satu lingkaran pie chart bernilai 100 persen, setiap bagian / potongan dari grafik ini menunjukkan values dari sebuah label. Visualisasi data menggunakan pie chart hanya memiliki satu values pengukuran yaitu luas dari potongan pie chart.

Pada visualisasi data menggunakan pie chart diatas menunjukkan persentasi dari label tahun ajaran 2015/2016 memiliki luas paling kecil dibandingkan luas potongan pie chart yang lain. Jika, outlier adalah suatu objek bernilai sangat jauh berbeda dengan nilai objek kebanyakan (normal) maka, label tahun ajaran 2015/2016 dapat dikatakan sebagai objek / transaksi outlier.

2. Visualisasi satu dimensi dengan dua measure

Visualisasi kedua yaitu visualisasi katagori untuk memvisualisasikan perbandingan total pembayaran dan potongan dimensi yang digunakan yaitu tahun ajaran. Field dimensi tahun ajaran yang digunakan yaitu tahun ajaran sedangkan nilai measure yang digunakan yaitu pembayaran dan potongan pembayaran. Visualisasi yang digunakan yaitu grafik batang dengan dua nilai katagori ditunjukkan pada Gambar 7 berikut:

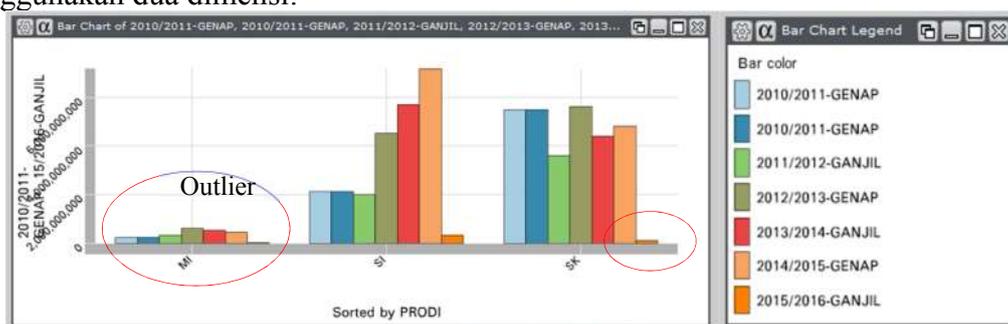


Gambar 7 Visualisasi satu dimensi pada pembayaran setiap tahun ajaran

Tujuan dari Visualisasi grafik batang diatas ingin mendeteksi outlier antara transaksi pembayaran dengan potongan yang didapat oleh mahasiswa pada setiap tahun ajaran. Pada grafik batang diatas warna biru pada barang grafik adalah nilai pembayaran sedangkan warna kuning tua (orange) adalah potongan yang didapat oleh mahasiswa pada tahun ajaran tersebut. Setiap tahun ajaran mulai tahun ajaran 2010/2011 sampai 2015/2016 terdapat potongan namun, tahun ajaran 2010/2011 memiliki potongan yang paling banyak dibandingkan dengan potongan di tahun ajaran yang lain. Jika dilihat dari perbandingan pembayaran tahun ajaran 2012/2013 sampai dengan 2015/2016 memiliki total pembayaran lebih tinggi namun potongannya jauh lebih kecil dari tahun ajaran 2010/2011. Berdasarkan analisis perbandingan nilai pada grafik batang diatas tahun ajaran 2010/2011 bisa dikatakan outlier pemberian potongan.

3. Visualisasi dua dimensi

Dengan menggunakan teknik drill-down proses OLAP pada data warehouse transaksi pembayaran maka, grafik pie chart dan grafik batang pembayaran sertiap tahun ajaran didetailkan menjadi tahun ajaran pada setiap tahun ajaran. Dimensi yang gunakan yaitu dimensi tahun ajaran field semester pada setiap tahun ajaran, dan dimensi mahasiswa yaitu field program studi mahasiswa. Nilai measure yang digunakan yaitu pembayaran (num_bayar). Berikut pada Gambar 8 adalah visualisasi grafik batang menggunakan dua dimensi:

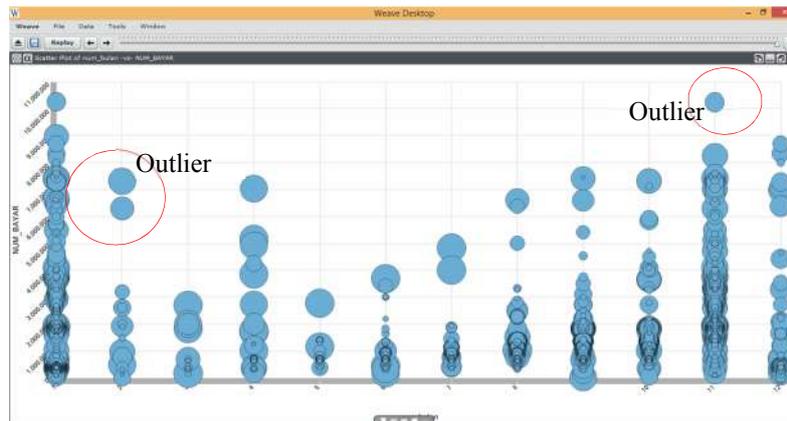


Gambar 8 Visualisasi pembayaran menggunakan dua dimensi

Deteksi outlier pada grafik batang menggunakan dua dimensi menjadi lebih kompleks, karena informasi banyak yang ditunjukkan. Secara global program studi MI memiliki nilai yang sangat jauh kecil dari program studi yang lain maka, nilai pembayaran pada program studi MI bisa dikatakan outlier dibandingkan program studi yang lain.

4. Visualisasi multi dimensional

Dimensi yang digunakan pada Proses OLAP berikut yaitu dimensi mahasiswa, dimensi tahun_ajaran, dan dimensi waktu, sedangkan measure nya yaitu pembayaran (num_bayar). Proses drill-down pada OLAP didetailkan hingga sampai report transaksi per mahasiswa. Demensi mahasiswa di drill-down yaitu program studi Sistem Informasi (SI), tahun ajaran di drill-down yaitu 2014/2015, dimensi waktu di drill-down yaitu bulan. Hasil dari OLAP terahir ini cukup sulit divisualisasikan dengan grafik batang. Maka, pada teknik yang digunakan yaitu dengan visualisasi *scatter plot* seperti berikut:



Gambar 9 Visualisasi pada multidimensional

Pada visualisasi scatter plot pada Gambar 9, garis x / label adalah urutan bulan dari dimensi waktu mulai dari bulan 1 yaitu januari hingga 12 yaitu desember. Garis Y menggunakan total pembayaran transaksi mahasiswa. Besar lingkaran yaitu jumlah mahasiswa yang membayar pada bulan tersebut (label/garis X) dan total pembayarannya sesuai garis Y (num_bayar). Deteksi outlier pada visualisasi scatter plot dilihat dari posisi titik lingkaran dan besar lingkaran. Titik lingkaran yang jauh dari lingkaran yang lain maka dapat dikatakan transaksi pembayaran yang diwakili oleh titik tersebut dikatakan outlier.

2.7 Deteksi Outlier

Deteksi outlier pada data yang sudah divisualisasikan bersifat objective. Deteksi outlier pada visualisasi seperti grafik batang, pie chart dan plot chart dianalisis melalui ukuran grafik, dan posisi. Ukuran dan (atau) posisi grafik yang paling berbeda dengan grafik yang lain bisa disebut outlier. Seperti pada gambar 5, luas label tahun ajaran 2015/2016 sangat sempit maka disebut outlier, dan lihat juga gambar 6 yaitu visualisasi satu dimensi pada pembayaran setiap tahun ajaran, tahun ajaran 2010/2011 memiliki potongan transaksi yang paling banyak. Pada gambar 8 yaitu visualisasi pada multidimensional, terdapat beberapa posisi point yang berjauhan dengan point yang ada di bawahnya.

3. HASIL DAN PEMBAHASAN

Visualisasi data dibangun melalui hasil OLAP dengan teknik *drill down*. OLAP pertama menggunakan satu tabel dimensional, kemudian dua tabel dimensional dan terakhir multidimensional. Visualisasi data menggunakan satu tabel dimensional divisualisasikan menggunakan pie chart atau grafik batang. Visualisasi data menggunakan dua tabel dimensional menggunakan grafik batang multi label. Visualisasi data menggunakan tiga dimensional menggunakan scatter plot. Deteksi outlier pada visualisasi data menggunakan satu tabel dimensional lebih mudah dianalisis dari pada deteksi outlier pada visualisasi data menggunakan dua atau multi tabel dimensional. Dengan kata lain semakin banyak tabel dimensi yang terlibat semakin sulit analisis deteksi outlier yang dilakukan.

4. KESIMPULAN

Kesimpulan yang didapat dari hasil penelitian ini adalah sebagai berikut:

1. Deteksi outlier pada visualisasi data bersifat objektive.
2. Semakin banyak tabel dimensi yang terlibat dalam membangun visualisasi data maka semakin sulit analisis deteksi outlier yang dilakukan.
3. Visualisasi data pada hasil OLAP dapat digunakan untuk mendeteksi outlier pada nilai measure dari tabel fakta.
4. Visualisasi hasil teknik drill-down untuk mendeteksi outlier dapat memberikan informasi yang bertahap mulai dari global sampai detail data penyebab outlier.

5. SARAN

Penelitian lebih lanjut dapat diterapkan dengan menggabungkan teknik data mining, OLAP, dan visualisasi data.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada STIKOM Bali atas dukungan financial terhadap penelitian ini, dan telah memberikan sebagian data untuk menjadi objek penelitian.

DAFTAR PUSTAKA

- [1]. Fatimah Sidi,dkk (2012), Data Quality comparative Model for data warehouse, IEEE
- [2]. William H. Inmon, (2005), Building the data warehouse fourth edition, Wiley Publishing
- [3]. Kimball Ralph, dkk, (2010), The Kimball Group Reader; Relentlessly Practical Tools for Data Warehousing and Business Intelligence, Wiley Publishing.
- [4]. Han Jiwei, Kamber, Pei., (2012), Data Mining concepts and techniques third edition. Morgan Kaufmann publishers
- [5]. Sri Yulianto Joko P, (2009), Olap Technology: Visualization Multidimensional Data On Agribusiness Information Resources In Indonesia, konferensi nasional system dan informatika 2009, Bali
- [6]. IAN H. witten, Eibe Frank, Mark A. Hall., (2011), Data Mining practical machine learning tools and techniques third edition. Morgan Kaufmann publishers

- [7]. Fred R. McFadden,(1996), Data Warehouse for EIS, some Issues and Impacts, IEEE
- [8]. Sai Satyanarayana Reddy,dkk,(2008), Advanced techniques for scientific data warehouses, International conference on Advanced Computer Control, IEEE
- [9]. Stefano Rizzi, (2007), OLAP Preferences : A Research Agenda, DOLAP'07, ACM
- [10]. Andy kirk, (2012), Data Visualization : A Successful Design Process. Packt publishing