

BAB IV

PERANCANGAN DATA WAREHOUSE

Pada bab ini dijelaskan profil perusahaan serta langkah-langkah yang akan dilakukan dalam perancangan *data warehouse* di PT. Indosat Tbk yaitu mengenai proses identifikasi, perencanaan secara teknik, penentuan sumber data yang digunakan dan proses persiapan data melalui proses ETL (*extract, transform, load*).

4.1 Profil Perusahaan

Perusahaan yang dijadikan objek penelitian oleh penulis dalam bentuk perseroan terbatas yaitu PT. Indosat Tbk. Perusahaan ini didirikan pada tahun 1967 sebagai perusahaan investasi asing yang didirikan untuk menyediakan layanan telekomunikasi secara internasional di Indonesia, memulai operasinya pada tahun 1969 dengan diresmikannya stasiun bumi Jatiluhur. Pada tahun 1980, pemerintah Indonesia mengambil semua saham dari PT. Indosat Tbk, yang kemudian menjadi SOE (*State-Owned-Enterprise*). Pada tahun 1994, PT. Indosat Tbk mendaftarkan sahamnya pada *Jakarta Stock Exchange* dan *Surabaya Stock Exchange*, sehingga menjadi SOE pertama yang sahamnya di tampilkan di luar negeri.

4.1.1 Latar Belakang Perusahaan

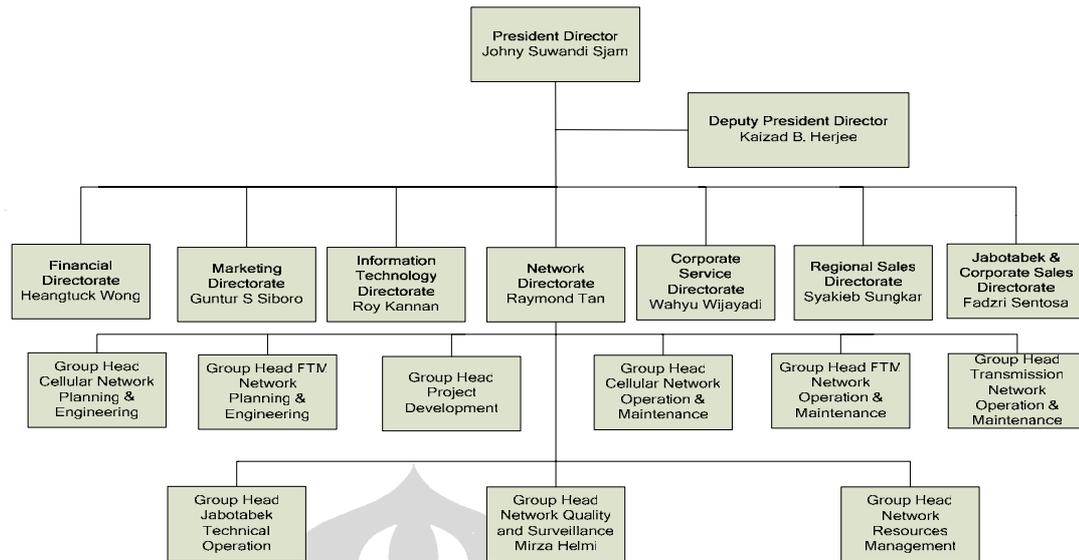
Pada tahun 1969 sampai dengan tahun 1990 PT. Indosat Tbk menyediakan layanan *switched* dan *non-switched* telekomunikasi internasional, termasuk telepon langsung internasional, jaringan komunikasi data, jalur sewa internasional dan layanan transmisi televisi internasional. Memasuki abad ke-21, supaya tetap menjaga tren global yang sedang berjalan, pemerintah Indonesia memutuskan untuk melakukan deregulasi pada sektor telekomunikasi nasional, membuka kompetisi terbuka kepada pasar. PT. Indosat Tbk mulai mengejar target utamanya yaitu bisnis selular pada pertengahan tahun '90. Pada tahun 2001, PT. Indosat Tbk membangun PT Indosat Multi Media Mobile (IM3), diikuti dengan kontrol penuh dari satelit Indonesia Palapa yang membuat grup PT. Indosat Tbk menjadi operator selular terbesar kedua di Indonesia. Pada akhir tahun 2002, pemerintah Indonesia menjual sahamnya sebesar 41,94% ke Singapore Technologies Telemedia Pte, Ltd. Kemudian pada tahun 2003, terjadi merger antara perusahaan Satelindo, IM3 dan Bimagraha menjadi PT. Indosat Tbk. Kemudian PT. Indosat Tbk menjadi perusahaan yang berfokus pada *Full Network Service Provided* (FNSP). Pada akhir tahun 2006, PT. Indosat Tbk menjadi operator kedua terbesar dengan 16.704.639 *subscriber*. Saat ini dalam mengelola usahanya, PT. Indosat Tbk, harus menghadapi persaingan ketat dengan operator lain, terutama dalam hal persaingan merebut pangsa pasar. Hal ini yang membuat PT. Indosat Tbk berusaha untuk meningkatkan *quality of service* nya untuk dapat merebut pangsa pasar di Indonesia..

4.1.2 Visi dan Misi Perusahaan

Visi dari PT. Indosat Tbk adalah menjadi perusahaan no 1 yang berfokus pada *cellular* atau *wireless*, jaringan telekomunikasi terintegrasi dan penyedia layanan di Indonesia, Sedangkan misi dari PT. Indosat Tbk adalah (1) untuk menyediakan dan mengembangkan produk yang inovatif dan berkualitas, layanan dan solusi yang menawarkan nilai yang terbaik untuk para pelanggan, (2) secara kontinu menumbuhkan nilai-nilai bagi *stakeholder*, (3) menyediakan kualitas hidup yang lebih baik dari para *stakeholder*.

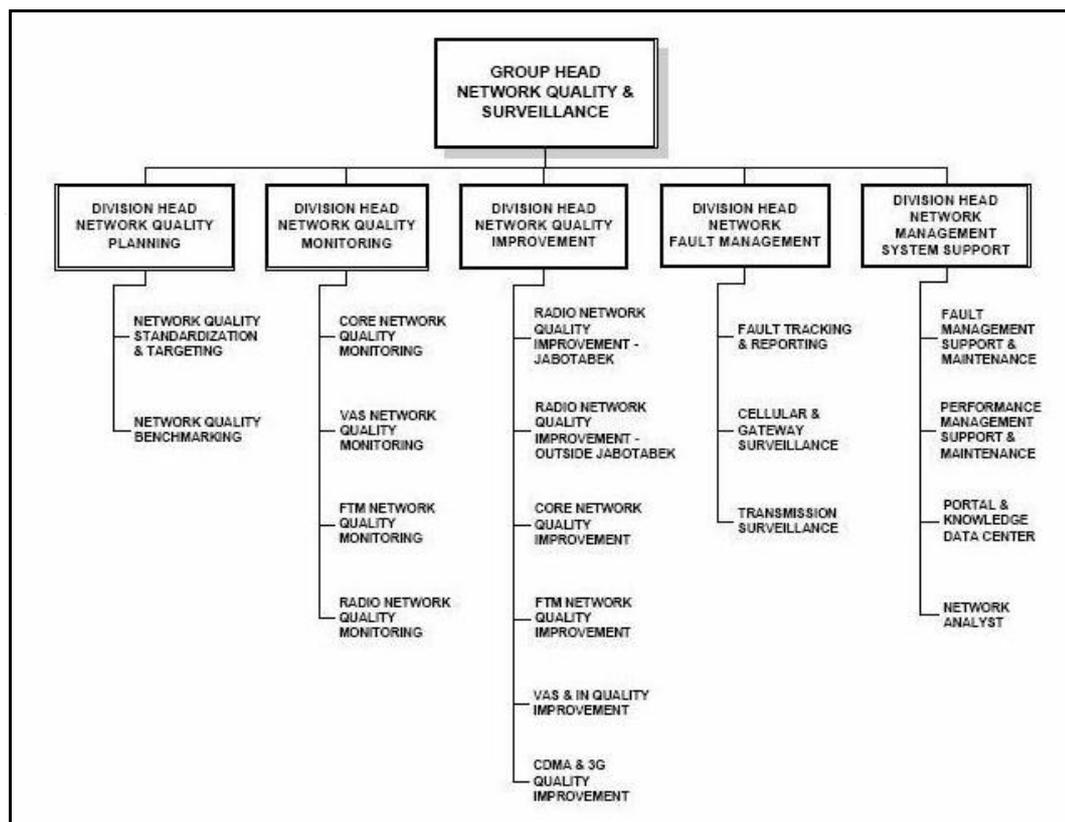
4.1.3 Struktur Organisasi Perusahaan

PT. Indosat Tbk memiliki struktur organisasi yang dibagi menjadi *directorate*, *sub directorate (group)* dan *division*. Berikut ini adalah struktur organisasi pada level *directorate*.



Gambar 4.1 – Struktur Organisasi Tingkat Direktorat

Berikut ini adalah struktur organisasi yang akan diperlihatkan struktur organisasi dalam ruang lingkup yang lebih kecil sesuai dengan ruang lingkup yang telah dibatasi oleh penulis yaitu pada *sub directorate (group) Network Quality & Surveillance (NQS)*. Untuk mengetahui lebih jelasnya dapat dilihat pada Gambar 4.2 berikut ini.



Gambar 4.2 – Struktur Organisasi di Group *Network Quality & Surveillance*

4.1.4 Proses Bisnis

Proses bisnis yang dilakukan oleh PT. Indosat Tbk adalah (1) mengembangkan suatu produk yang inovatif, beraneka ragam dan berkualitas, layanan dan solusi yang menawarkan nilai yang terbaik untuk para pelanggan dengan cara membuat dan menawarkan produk yang bervariasi dengan layanan jaringan telekomunikasi yang lengkap seperti *cellular*, *3G*, *fixed landline*, *fixed wireless*, *DLD*, *IDD*, *internet* dan *satelit*, (2) memasarkan suatu produk dengan harga, layanan dan konten yang bervariasi, sebagai contoh *Mentari 50*, *Super SMS 200*, *Mentari 5000*, *Mentari 1 rupiah* (3) berfokus pada layanan jaringan

telekomunikasi yang lengkap sesuai motto/nilai PT. Indosat Tbk yaitu Insan Gemilang (integritas, kerjasama, keunggulan, kemitraan, fokus pada pelanggan) dan Indosat *Brand Promise* (layanan personal, jaminan kehandalan, memberi lebih dari yang diharapkan, dan keunggulan teknologi). Untuk dapat menjalankan proses bisnis perusahaan dengan baik, PT. Indosat Tbk memiliki salah satu *sub directorate (group)* yang berperan penting dalam mendukung proses bisnis PT. Indosat Tbk yaitu *group Network Quality & Surveillance (NQS)*, dimana tugas dari *group* ini adalah untuk menjaga kualitas dan kinerja (*measurement data*) dari masing-masing *network element* yang terdapat di jaringan PT. Indosat Tbk, menurut suatu nilai standar yang dikenal dengan *Key Performance Indicator (KPI)*.

4.2 Perancangan Arsitektur Perusahaan

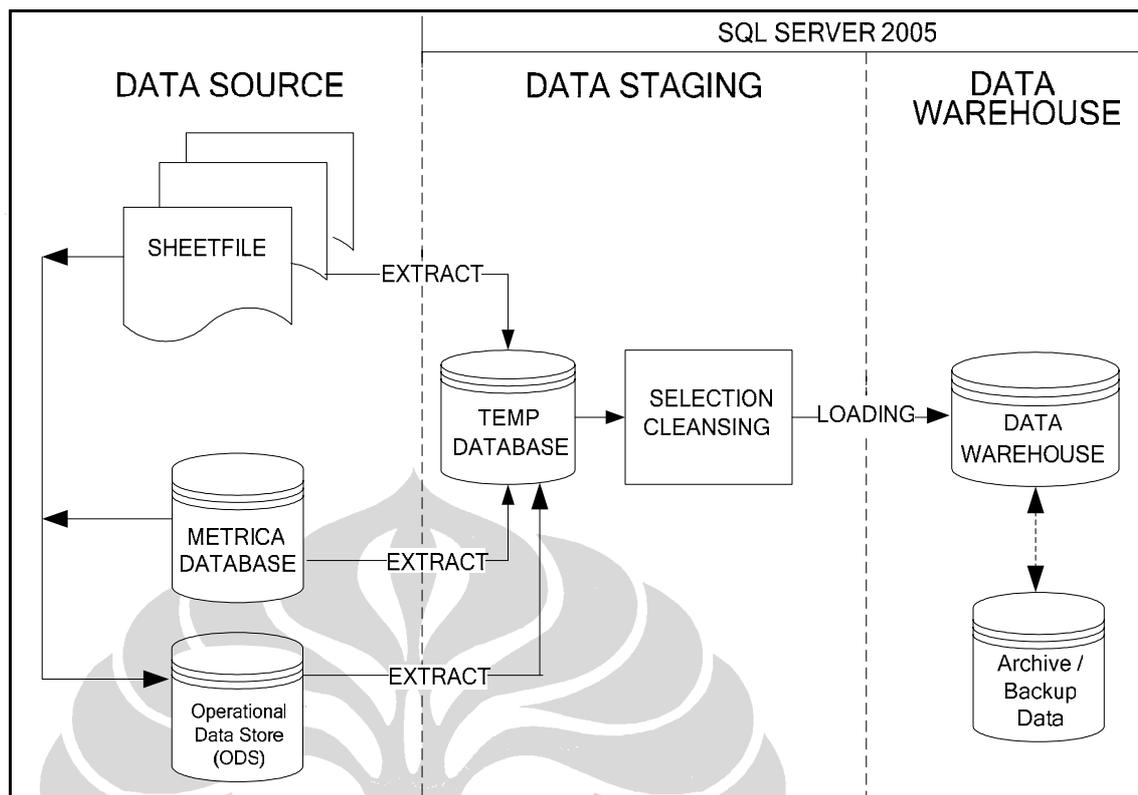
Proses perancangan arsitektur *data warehouse* adalah proses yang sangat rumit, untuk memulainya kita harus mendefinisikan kebutuhan dari pengguna yang paling dibutuhkan dan data mana yang harus lebih diutamakan. Pada tahap ini penulis akan mencoba merancang arsitektur *logical* dan arsitektur fisik untuk *data warehouse*. Arsitektur *logical* adalah rancangan tahapan alur data dari sumber data yang digunakan sampai *data warehouse* yang digunakan, sedangkan arsitektur fisik adalah gambaran teknis dari konfigurasi yang akan diterapkan pada *data warehouse*.

4.2.1 Arsitektur *Logical*

Pada perancangan arsitektur *logical* pada *data warehouse*, sumber data yang akan digunakan adalah sumber data yang diperoleh dari data operasional. Untuk lebih jelasnya dapat dilihat sebagai berikut:

- *Metrica*, yaitu *database* yang digunakan untuk penyimpanan data operasional dari elemen-elemen jaringan, yang datanya diperoleh dari OSS (*The Operation and Support Subsystem*)
- *Sheet File*, adalah data yang diperoleh dari laporan harian yang berisikan kualitas dan kinerja masing-masing elemen jaringan yang terdapat di jaringan PT. Indosat Tbk

Dari sumber-sumber tersebut, kemudian dilakukan proses *selection* yaitu proses pemilihan data yang diperlukan dari sumber tersebut, karena dalam pembuatan *data warehouse* tidak semua data akan digunakan. Proses yang akan dilakukan selanjutnya adalah data yang sudah diseleksi akan disatukan ke dalam suatu sistem *database* sementara yang terpisah. Kegunaan dari pemisahan *database* tersebut adalah agar tidak mengganggu kinerja dari masing-masing sistem *database* operasional yang ada. Data yang sudah terseleksi kemudian akan dilakukan proses *cleansing* yaitu proses pembersihan data dan proses transformasi. Jika diperlukan kedua sumber data yang telah melalui proses di atas dapat disimpan di *operation data store* (ODS) untuk keperluan analisis yang lain. Proses terakhir yang dilakukan adalah proses *loading* yaitu proses memasukkan data hasil dari tahap sebelumnya ke dalam *data warehouse* oleh *load manager*. Aliran data dari arsitektur *logical* tersebut dapat dilihat pada Gambar 4.3

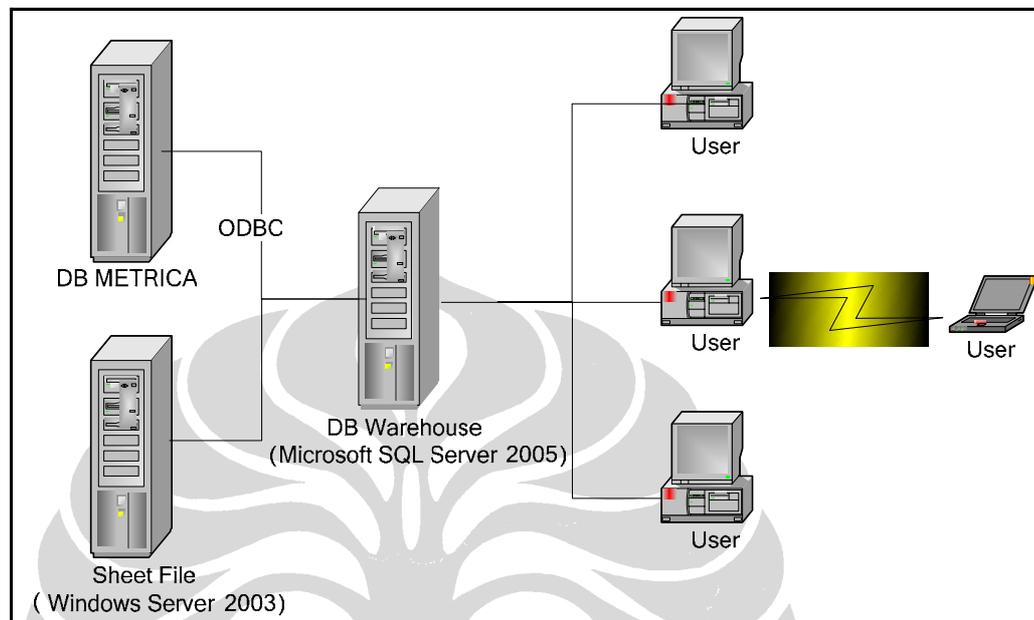


Gambar 4.3 – Arsitektur *Logical*

4.2.2 Arsitektur Fisik

Database yang digunakan adalah *metrica database* dimana *engine* RDBMS yang digunakan adalah *metrica engine* dijalankan pada *server database* yang menggunakan platform *unix*, sedangkan untuk data dalam bentuk *sheet file excel*, penyimpanan datanya diletakkan pada *platform Microsoft Windows Server 2003*. Terhadap kedua sumber tersebut akan dilakukan proses pemilihan data (*extract*) dan transformasi. Data yang sudah dibersihkan akan melalui proses selanjutnya yaitu proses *loading*. Pada proses ini data tersebut akan di *load* ke

dalam *data warehouse* yang menggunakan *engine Microsoft SQL Server 2005* dan menggunakan *platform Windows Server 2003*. Untuk lebih jelasnya dapat dilihat pada Gambar 4.4 berikut ini:



Gambar 4.4 – Arsitektur Fisik

4.3 Sumber Data

Sumber data yang akan digunakan adalah sumber data yang diperoleh dari sumber data internal yaitu *metrica database* dan *sheet file*. *Metrica Database* berisikan informasi mengenai *measurement counter*, sedangkan *sheet file* berisikan informasi mengenai kualitas dan kinerja dari masing-masing *network element* yang terdapat di jaringan PT. Indosat Tbk. Untuk lebih jelasnya dapat dilihat sumber data yang digunakan dapat dilihat pada Tabel 4.1 berikut ini:

| Metrica Database | |
|---------------------------|----------------------|
| Tipe Database | RDBMS |
| Engine | Metrica Engine |
| Platform | HP UX |
| Nama Database | Metrica |
| Jumlah Total Tabel | 30 |
| Jumlah Total Kolom | 60 |
| Pertambahan Data | 2,4 Giga Bytes / Day |
| Ukuran Data | 876 Giga Bytes |

Tabel 4.1 – Metrica Database

| Data Sheet File | |
|---------------------------|-------------------------------|
| Tipe Data | Sheet File |
| Engine | Microsoft Excel 2003 |
| Platform | Microsoft Windows Server 2003 |
| Jumlah Total Sheet | 5 / Day |
| Jumlah Total Kolom | 236 |
| Pertambahan Data | 29296 Record / Day |
| Ukuran Total Data | 22265 Mega Bytes |

Tabel 4.2 – Data Sheet File

Contoh dari *sheet file* yang digunakan sebagai sumber data dapat dilihat pada Gambar 4.5 berikut ini:

| Performance Analysis | | Traffic Observation | | | | | | | | | | Succ.Call Observation | | | | |
|----------------------|----------|---------------------|----------------|-------|---------|-----------------|-------------------|---------|--------|---------------|-----------|-----------------------|--------------|---------|---------|-----|
| ANALYSIS | | | | | | | | | | | | | | | | |
| Date | | | | | | | | | | | | | | | | |
| 31-Dec-07 | | | | | | | | | | | | | | | | |
| No. | BSS_Name | BTS_Name | CI | Area | Total | Total Traffic | HO Attempt | Call | BH TCH | Max Available | Dedicated | GOS | TCH | Avg TCH | Erlang | RA |
| | | | | | Traffic | (Normalization) | (Out/Intr a Cell) | Attempt | Erl | CS | PS | Availability | Holding Time | Drop | Minute | Su |
| 4 | | | | | | | | | | | | | | | | |
| 5 | 1 | AKSAN | HMARBELLA1 | 23118 | 31.07 | 31.07 | 652 | 1788 | 2.64 | 12 | 1 | 0.00% | 100.00% | 67.86 | 310.70 | 95 |
| 6 | 2 | AKSAN | HMARBELLA2 | 23119 | 23.93 | 23.93 | 597 | 1229 | 2.38 | 12 | 1 | 0.00% | 100.00% | 44.69 | 1425.90 | 100 |
| 7 | 3 | AKSAN | AKSAN1 | 23011 | 22.47 | 22.47 | 1680 | 1412 | 2.85 | 12 | 1 | 0.00% | 100.00% | 26.67 | 103.71 | 95 |
| 8 | 4 | AKSAN | AKSAN2 | 23012 | 35.23 | 35.23 | 1117 | 2878 | 4.92 | 12 | 1 | 0.80% | 100.00% | 32.45 | 117.43 | 100 |
| 9 | 5 | AKSAN | AKSAN3 | 23013 | 115.21 | 115.21 | 2374 | 7402 | 11.35 | 18 | 1 | 3.00% | 100.00% | 37.65 | 288.03 | 100 |
| 10 | 6 | AKSAN | BALEKAMBANG1 | 23301 | 2.44 | 2.44 | 137 | 241 | 0.34 | 12 | 1 | 0.00% | 100.00% | 21.82 | 36.50 | 95 |
| 11 | 7 | AKSAN | BALEKAMBANG2 | 23302 | 1.12 | 1.12 | 33 | 148 | 0.28 | 12 | 1 | 0.00% | 100.00% | 17.55 | 67.20 | 100 |
| 12 | 8 | AKSAN | BALEKAMBANG3 | 23303 | 2.63 | 2.63 | 243 | 190 | 0.58 | 12 | 1 | 0.00% | 100.00% | 19.71 | 52.60 | 100 |
| 13 | 9 | AKSAN | BATUKUDA1 | 22211 | 2.11 | 2.11 | 103 | 214 | 0.68 | 2 | 0 | 0.00% | 93.20% | 4.76 | 42.20 | 95 |
| 14 | 10 | AKSAN | BATUKUDA2 | 22212 | 0.12 | 0.12 | 6 | 13 | 0.08 | 2 | 0 | 0.00% | 96.20% | 2.70 | nd | 100 |
| 15 | 11 | AKSAN | BATUKUDA3 | 22213 | 0.24 | 0.24 | 25 | 23 | 0.12 | 2 | 0 | 0.00% | 96.67% | 2.54 | nd | 95 |
| 16 | 12 | AKSAN | BONJEN1 | 23461 | 22.98 | 22.98 | 1549 | 1684 | 2.51 | 12 | 0 | 0.00% | 100.00% | 28.80 | 153.27 | 100 |
| 17 | 13 | AKSAN | BONJEN2 | 23462 | 58.75 | 58.75 | 2350 | 4701 | 5.29 | 20 | 0 | 0.00% | 100.00% | 29.73 | 130.56 | 95 |
| 18 | 14 | AKSAN | BONJEN3 | 23463 | 12.42 | 12.42 | 1035 | 830 | 1.44 | 13 | 0 | 0.00% | 100.00% | 21.31 | 149.04 | 100 |
| 19 | 15 | AKSAN | BULUH1 | 23481 | 0.81 | 0.81 | 25 | 73 | 0.15 | 13 | 0 | 0.00% | 100.00% | 21.47 | nd | 100 |
| 20 | 16 | AKSAN | BULUH2 | 23482 | 0.05 | 0.05 | 4 | 10 | 0.03 | 13 | 0 | 0.00% | 100.00% | 3.18 | nd | 100 |
| 21 | 17 | AKSAN | BULUH3 | 23483 | 0.22 | 0.22 | 1 | 19 | 0.07 | 13 | 0 | 0.00% | 100.00% | 18.61 | nd | 100 |
| 22 | 18 | AKSAN | CAMARA1 | 23611 | 0 | 0 | 0 | 0 | 0 | 13 | 0 | 0.00% | 100.00% | 0.00 | nd | 100 |
| 23 | 19 | AKSAN | CAMARA2 | 23612 | 3.43 | 3.43 | 104 | 169 | 0.42 | 13 | 0 | 0.00% | 100.00% | 33.62 | nd | 100 |
| 24 | 20 | AKSAN | CAMARA3 | 23613 | 0.04 | 0.04 | 6 | 11 | 0.02 | 13 | 0 | 0.00% | 100.00% | 1.55 | nd | 95 |
| 25 | 21 | AKSAN | CARITAAKSAN1 | 24371 | 47.74 | 47.74 | 5007 | 3050 | 9.24 | 11 | 1 | 17.90% | 100.00% | 27.85 | 110.17 | 100 |
| 26 | 22 | AKSAN | CARITAAKSAN2 | 24372 | 5.16 | 5.16 | 907 | 449 | 0.97 | 11 | 1 | 0.00% | 100.00% | 13.30 | 154.80 | 100 |
| 27 | 23 | AKSAN | CARITAAKSAN3 | 24373 | 41.09 | 41.09 | 3091 | 3342 | 3.31 | 12 | 1 | 0.00% | 100.00% | 25.16 | 164.36 | 100 |
| 28 | 24 | AKSAN | CARITABAY1 | 23021 | 7.97 | 7.97 | 1885 | 839 | 1.05 | 11 | 1 | 0.00% | 100.00% | 12.95 | 95.64 | 100 |
| 29 | 25 | AKSAN | CARITABAY2 | 23022 | 131.83 | 131.83 | 16279 | 10337 | 14.39 | 10 | 1 | 39.00% | 100.00% | 19.32 | 118.06 | 100 |
| 30 | 26 | AKSAN | CARITABAY3 | 23023 | 108.1 | 108.1 | 15510 | 4614 | 19.62 | 11 | 1 | 53.00% | 100.00% | 17.63 | 129.72 | 95 |
| 31 | 27 | AKSAN | CARITAROAD1 | 23031 | 25.98 | 25.98 | 1015 | 1957 | 2.77 | 11 | 1 | 0.00% | 100.00% | 31.53 | 103.92 | 100 |
| 32 | 28 | AKSAN | CARITAROAD2 | 23032 | 2.98 | 2.98 | 397 | 316 | 0.31 | 11 | 1 | 0.00% | 100.00% | 15.27 | 18.28 | 100 |
| 33 | 29 | AKSAN | CARITAROAD3 | 23033 | 14.41 | 14.41 | 522 | 1005 | 2.14 | 11 | 1 | 0.00% | 100.00% | 31.06 | 144.10 | 100 |
| 34 | 30 | AKSAN | CERLANG1 | 23121 | 24.34 | 24.34 | 1519 | 1627 | 2.78 | 12 | 1 | 0.00% | 100.00% | 26.92 | 104.31 | 95 |
| 35 | 31 | AKSAN | CERLANG2 | 23122 | 31.88 | 31.88 | 860 | 2225 | 2.86 | 12 | 1 | 0.00% | 100.00% | 42.22 | 136.63 | 100 |
| 36 | 32 | AKSAN | CERLANG3 | 23123 | 62.08 | 62.08 | 1327 | 3661 | 6.47 | 12 | 1 | 3.30% | 100.00% | 52.75 | 206.93 | 95 |
| 37 | 33 | AKSAN | CIBADAK_PDGL1 | 23651 | 2.53 | 2.53 | 71 | 317 | 0.38 | 13 | 0 | 0.00% | 100.00% | 23.05 | 79.90 | 100 |
| 38 | 34 | AKSAN | CIBADAK_PDGL2 | 23652 | 3.06 | 3.06 | 112 | 386 | 0.44 | 13 | 0 | 0.00% | 100.00% | 20.44 | 91.50 | 100 |
| 39 | 35 | AKSAN | CIBADAK_PDGL3 | 23653 | 11.52 | 11.52 | 845 | 859 | 1.69 | 13 | 0 | 0.00% | 100.00% | 34.04 | 31.42 | 100 |
| 40 | 36 | AKSAN | CIBALUNG1 | 23271 | 5.31 | 5.31 | 85 | 375 | 0.83 | 6 | 0 | 0.10% | 99.23% | 42.53 | 79.65 | 100 |
| 41 | 37 | AKSAN | CIBALUNG2 | 23272 | 5.91 | 5.91 | 49 | 460 | 0.69 | 13 | 0 | 0.00% | 100.00% | 45.53 | 254.60 | 100 |
| 42 | 38 | AKSAN | CIBALUNG3 | 23273 | 3.78 | 3.78 | 67 | 495 | 0.67 | 13 | 0 | 0.00% | 100.00% | 46.02 | nd | 100 |
| 43 | 39 | AKSAN | CIBAWANG1 | 23691 | 19.23 | 19.23 | 623 | 1789 | 1.91 | 13 | 0 | 0.00% | 100.00% | 28.15 | 30.36 | 100 |
| 44 | 40 | AKSAN | CIBAWANG2 | 23692 | 39.68 | 39.68 | 843 | 3119 | 4.49 | 12 | 0 | 0.40% | 100.00% | 34.59 | 36.63 | 100 |
| 45 | 41 | AKSAN | CIBAWANG3 | 23693 | 11.14 | 11.14 | 994 | 879 | 1.21 | 13 | 0 | 0.00% | 100.00% | 20.62 | 60.76 | 100 |
| 46 | 42 | AKSAN | CIBITUNG_PDGL1 | 23531 | 0.52 | 0.52 | 7 | 103 | 0.24 | 13 | 0 | 0.00% | 100.00% | 36.57 | nd | 100 |

Gambar 4.5 – Contoh Sheet File

4.4 Data Staging

Sumber data yang digunakan untuk keperluan *data warehouse* cukup banyak dan cukup besar ukurannya, tetapi data yang akan digunakan untuk keperluan *data warehouse* tidaklah semuanya. Hanya data yang mendukung informasi yang dibutuhkan oleh pengguna yang akan digunakan. Tempat untuk melakukan seleksi informasi yang diperlukan dan mempersiapkan data untuk diproses lebih lanjut ke *data warehouse* adalah *staging area*, sedangkan data pada kondisi tersebut dinamakan *data staging*. Karena itulah pada tahap *data staging*, akan dilakukan proses *selection, filtering, editing, summarizing, combining* dan

loading data terhadap sumber data untuk mempersiapkan data dalam pemrosesan lebih lanjut ke *data warehouse*.

4.5 Proses ETL

Proses ETL (*extract, transform, load*) adalah proses yang digunakan dalam memproses *data* sebelum dimasukkan ke dalam suatu *data warehouse* yang akan dilakukan oleh *load manager*. Proses ini dilakukan untuk men-standarisasikan data yang akan digunakan pada *data warehouse* sebelum dimasukkan ke dalam *data warehouse*. Proses ini dilakukan karena tahapan ini merupakan proses yang tepat dilakukan karena proses ini adalah langkah kritis dalam pembuatan suatu *data warehouse*. Proses ETL ini dilakukan secara periodik dan otomatis.

4.5.1 Proses Extract

Ekstraksi dilakukan dari sumber data yang digunakan melalui proses pemilihan data yang kemudian disimpan pada *database* DWTemp. Penempatan DWTemp diletakkan pada penyimpanan *database*, mesin dan *platform* yang sama dengan *data warehouse*. Adapun daftar *database* metrica dan DWTemp dapat dilihat di lampiran 1 dan lampiran 2 beserta penjelasan atribut *metadata* untuk tabel DWTemp pada lampiran 3. Untuk lebih detailnya mengenai perbedaan karakteristik sumber data dan tujuan data dapat dilihat pada Tabel 4.3 berikut ini.

| EKSTRAKSI | | | |
|---------------------------|-----------------|------------------|-------------------------------|
| | <i>Source I</i> | <i>Source II</i> | <i>Destination</i> |
| <i>Engine</i> | Metrica Engine | ODBC Engine | Microsoft SQL Server 2005 |
| <i>Platform</i> | HP UX | Windows | Microsoft Windows Server 2003 |
| Sumber Data | DB Metrica | Excel | DWTemp |
| Jumlah Total Table | 30 | 5 | 7 |
| Jumlah Total Field | 60 | 236 | 31 |
| Jumlah Total Row | 482663880 | 8044398 | 8083913 |

Tabel 4.3 – Hasil Ekstraksi

Pada Tabel 4.4 berikut ini dapat dilihat juga contoh dari sejumlah *field* yang digunakan pada proses ekstraksi dari dua sumber data yaitu *source I* (DB Metrica) dan *source II* (Excel) yang kemudian akan diekstrak ke tempat tujuan yaitu DWTemp. Pemilihan *field* yang digunakan tersebut berdasarkan keperluan analisis terhadap parameter-parameter yang merupakan KPI utama yang merupakan tolak ukur dalam pengukuran kualitas dan kinerja jaringan.

| TABEL | Field | DIGUNAKAN | |
|-----------------------------|----------------------|-----------|-------|
| | | YA | TIDAK |
| HUAWEI | No. | | X |
| | Date | V | |
| | BSCID | V | |
| | CELLID | V | |
| | CELLNAME | V | |
| | Traffic (Erl) | V | |
| | TCH Attempt | | X |
| | TCH Seizure | | X |
| | TCH Block | | X |
| | Call Drop | V | |
| | BHTraffic | | X |
| | CSSR(%) | V | |
| | CDR(%) | V | |
| | SCR(%) | V | |
| | HOSR(%) | V | |
| | SDCCH_Block_Rate (%) | | X |
| | SDCCH_Drop_Rate(%) | | X |
| | TCH Block Rate (%) | | X |
| TCH Assign Failure Rate (%) | | X | |

| | | | |
|----------------------|---------------------------------------|---|---|
| | Erlang Minutes Per Drop | | X |
| | RACH Success Rate (%) | | X |
| | TCH Availability(%) | | X |
| | Cell Utilization (%) | | X |
| | TCH Available | | X |
| | SDCCH Attempt | | X |
| | SDCCH Seizure | | X |
| | SDCCH Block | | X |
| | SDCCH Drop | | X |
| | SDCCH Drop due to Radio Problem | | X |
| | SDCCH Drop due to MiscDrop | | X |
| | TCH Channel Activation NACKs (Call) | | X |
| | TCH Channel Activation Timeout (Call) | | X |
| | TCH Ass. Fail | | X |
| | RACH Attempt | | X |
| | RACH Success | | X |
| | Call Drop due to Radio Problem | | X |
| | Call Drop due to Misc Drop | | X |
| | Call Drop on A_if | | X |
| | Call Drop on Abis | | X |
| | O/G HO Fail due to rev. fail | | X |
| | Internal HO Fail due to rev. fail | | X |
| | Intracell HO Fail due to rev. fail | | X |
| | Successful_HO_Numbers | | X |
| | HO_Numbers | | X |
| | SDCCH Available | | X |
| | SDCCH Traffic (Total) | | X |
| | SDCCH Traffic (Max) | | X |
| | Total Hours | | X |
| Ref_Geography | No | | |
| | City | V | |
| | Province | V | |
| | Island | V | |
| | Region | V | |

Tabel 4.4 – Contoh *field* pada proses ekstraksi

Untuk lebih memahami istilah-istilah yang akan digunakan berikut ini dijelaskan mengenai beberapa komponen GSM (Gunnar Heine, 1999):

➤ *Mobile Station (MS)*

Perangkat fisik yang digunakan oleh pelanggan GSM. Terdiri dari dua bagian: (1) *Subscriber Identity Module (SIM)* yaitu kartu pintar yang membawa informasi khusus pelanggan yang digunakan MS. Fungsi utama adalah identifikasi pemakai MS dan untuk keamanan dan kerahasiaan prosedur. SIM juga menyimpan data lokasi aktual dan informasi personal pemakai seperti singkatan kode panggilan (direktori telepon), (2) *Mobile Equipment* yang menyediakan komunikasi radio dan pemrosesan yang diperlukan untuk mengakses jaringan GSM, antarmuka manusia dengan mesin yang memungkinkan pemakai mengakses layanan yang disediakan operator

➤ *Cell*

Adalah salah satu bagian kecil didalam suatu BTS, yang biasanya dicirikan sebagai *directional antenna* yang mengarah ke suatu area. konsentrasi satu *cell* ke suatu area ini dapat juga disebut satu sektor. Struktur tipikal dari suatu BTS memiliki tiga sektor atau tiga *cell* yang mengarah ketiga area yang berbeda dan berlawanan

➤ *Base Transceiver Station (BTS)*

Mencakup komunikasi radio didalam satu *cell*. BTS merupakan perangkat memancar dan menerima gelombang radio (termasuk antenna) dan yang berhubungan dengan pemrosesan sinyal. BTS biasanya dicirikan dengan tiang pemancar dimana ada 3 antenna pada bagian atas tiang tersebut

➤ *Base Station Controller (BSC)*

Switch kecil yang handal dengan kemampuan pemrosesan yang lebih baik.

BSC adalah yang mengkoneksikan sejumlah BTS melalui *abis-interface*.

Suatu BSC bertindak sebagai konsentrator lokal trafik dan melakukan *switching* lokal untuk *handover* antar sejumlah BTS

Penjelasan lebih rinci dari masing-masing deskripsi *field* (Tim Network Quality Improvement, 2006) yang digunakan dalam proses ekstraksi seperti pada Tabel 4.4 dapat dilihat berikut ini:

- *Date*: Tanggal dimana data pengukuran kinerja dihasilkan
- *CELLID*: Identitas dari *cell*
- *Cellname*: Nama dari suatu *cell*
- *BSCID*: Identitas dari suatu BSC
- *Traffic*: Secara umum *traffic* dapat diartikan sebagai perpindahan informasi dari satu tempat ke tempat lain melalui jaringan telekomunikasi. Besaran dari suatu *traffic* telekomunikasi diukur dengan satuan waktu, sedangkan nilai *traffic* dari suatu kanal adalah lamanya waktu pendudukan pada kanal tersebut dengan satuan *erlang*. *traffic* sendiri merupakan perhitungan statistik dan probabilitas, karena kita tidak mengetahui kapan suatu pembicaraan akan dilakukan dan kapan akan selesai. Salah satu tujuan perhitungan *traffic* adalah untuk mengetahui unjuk kerja jaringan (*Network Performance*) dan mutu pelayanan jaringan telekomunikasi (*Quality of Service*)

- CSSR (*Call Setup Success Rate*): Indikator ini merepresentasikan persentase dari panggilan yang berhasil memperoleh *traffic channel* dan tidak boleh terinterupsi dengan permasalahan putusnya koneksi jaringan, hilang sinyal, proses normal yang gagal dan jaringan yang *congest*
- CDR (*Call Drop Rate*): Adalah istilah yang umum untuk panggilan telepon selular dimana panggilan telepon terputus secara tidak diharapkan karena alasan teknis dalam bentuk persentase. Penyebab paling umum yang terjadi adalah ketika MS keluar dari area jaringan *wireless*. Suatu panggilan aktif tidak bisa dikelola ketika berada di area jaringan perusahaan yang berbeda ketika sedang digunakan menyebabkan diputusnya panggilan ketika sinyal tidak bisa dikelola antara MS dan jaringannya. Penyebab umum yang lain adalah ketika MS dibawa memasuki area dimana komunikasi *wireless* tidak tersedia, terinterupsi, intervensi atau *jammed*. Dalam perspektif jaringan, hal ini sama dengan MS keluar dari cakupan area. Terkadang panggilan terputus karena adanya *handoff* (perpindahan jaringan) antara *cell* di dalam jaringan penyedia jasa yang sama, hal ini bisa terjadi karena tidak seimbangnya antara *traffic* antara dua *cell* didalam zona BTS yang berdekatan. Jika *cell* pada BTS baru sudah mencapai maksimal kapasitasnya maka tidak bisa menerima tambahan *traffic* dari panggilan baru. Selama MS tidak bisa mencari alternatif lain untuk mengambil alih panggilan tersebut maka panggilan itu akan terputus. Panggilan telepon juga bisa terputus akibat MS lawan bicara dalam keadaan *lowbat* (baterai lemah) kemudian baterai habis sehingga menyebabkan *call drop*.

- *Call Drop*: indikator ini tidak jauh berbeda maknanya dengan *call drop rate* yaitu suatu kondisi dimana panggilan telepon terputus secara tidak diharapkan karena alasan teknis, yang berbeda adalah *call drop* tidak dalam bentuk persen melainkan dalam satu satuan per periode pengukuran
- SCR: Indikator ini merepresentasikan persentase dari *rate* panggilan ketika panggilan telah berhasil dilakukan sampai panggilan telah selesai dilakukan tanpa diinterupsi oleh sinyal yang terputus, *error* ketika dalam panggilan dan *call drop*
- HOSR (*Handover Success Rate*): Proses aktual dari proses *transfer* suatu panggilan dapat disebut sebagai *handover*. Dalam telekomunikasi ada dua alasan mengapa *handover* harus dilakukan yaitu: (1) Jika pengguna MS telah keluar dari area suatu *base station* dan bisa memperoleh koneksi lebih kuat dari pengirim sinyal, (2) *base station* dalam keadaan koneksi yang sudah memenuhi kapasitas maka bisa di *transfer* ke *base station* terdekat. Jadi definisi dari HOSR adalah jumlah persentase dari internal dan eksternal *handover* keluar yang berhasil dari jumlah total usaha internal dan eksternal *handover* keluar dalam persen
- City: kota dimana elemen jaringan itu berada
- Province: provinsi dimana elemen jaringan itu berada
- Island: pulau dimana elemen jaringan itu berada
- Region: regional dimana elemen jaringan itu berada

Proses berikutnya adalah dari data tersebut dilakukan penyeragaman data baik dari segi penamaan *field*, maupun tipe data yang akan digunakan. Contoh dari ketidakkonsistenan data dapat dilihat pada Tabel 4.5 dan Tabel 4.6.

| Ketidak konsistenan Format Data | | |
|---------------------------------|------------|------------|
| Tabel | Contoh | Format |
| Alcatel | 27-01-2007 | dd-mm-yyyy |
| Ericsson | 12-22-2007 | mm-dd-yyyy |
| Huawei | 08-31-2007 | mm-dd-yyyy |
| Nokia | 08-31-2007 | mm-dd-yyyy |
| Siemen | 08-31-2007 | mm-dd-yyyy |

Tabel 4.5 – Perbedaan Format Data

| Ketidak Konsistenan Penamaan Field | | |
|------------------------------------|----------------|----------------------------------|
| <i>Sheet</i> | <i>field I</i> | <i>field II</i> |
| Huawei | <i>Cellid</i> | <i>Call Drop</i> |
| Nokia | <i>CI</i> | <i>Drop after TCH assignment</i> |

Tabel 4.6 – Perbedaan Penamaan *Field*

Untuk mengatasi permasalahan pada Tabel 4.5 maka dilakukan penyeragaman format data waktu mengikuti format yang banyak digunakan yaitu mm-dd-yyyy, hal ini dilakukan untuk menghindari data yang tidak konsisten yang dapat menimbulkan data yang bias. Di dalam Tabel 4.6 terlihat ada ketidakkonsistenan dalam penamaan *field*, dimana pada *sheet* Huawei pada contoh *field I* memiliki *field cellid*, *field II* memiliki *field Call Drop* sedangkan pada *sheet* Nokia pada contoh *field I* memiliki *field CI*, *field II* memiliki *field Drop after TCH assignment*, hal ini dapat membingungkan bila kedua *field* tersebut dibiarkan terpisah dalam satu tabel yang sebenarnya memiliki makna yang sama. Untuk itu

dilakukan penyeragaman *field* sesuai makna umum yang standar. Contoh penyeragaman penamaan *field* dapat dilihat di Tabel 4.7.

| Penyeragaman Penamaan Field | | |
|-----------------------------|----------------------------------|---------------------|
| <i>Sheet</i> | <i>Field Asal</i> | <i>Field Tujuan</i> |
| Huawei | <i>Cellid</i> | <i>Cellid</i> |
| | <i>Call Drop</i> | <i>Call Drop</i> |
| Nokia | <i>CI</i> | <i>Cellid</i> |
| | <i>Drop after TCH assignment</i> | <i>Call Drop</i> |

Tabel 4.7 – Penyeragaman Penamaan *Field*

4.5.2 Proses *Cleansing*

Proses yang akan dilakukan selanjutnya adalah melakukan pembersihan data (*cleansing*) dari hasil yang diperoleh melalui proses ekstraksi. Proses yang dilakukan pada proses *cleansing* ini adalah proses untuk membersihkan data yang *redundant* dan data yang tidak konsisten satu sama lain. Contoh dari ketidak konsistenan data adalah seperti data yang rusak (*corrupt*) sehingga isi darinya tidak benar, data kosong, data dengan spasi yang berlebihan dan contoh seperti yang terjadi pada data di *field cellname* tabel *cell*. Isi dari *cellname* yang ditampilkan mengalami perubahan sesuai perubahan yang terjadi pada konfigurasi yang ada sehingga ketika data baru yang telah berubah ditambahkan ke *database* maka akan terjadi perbedaan pada *field cellname* tersebut karena itu data *cellname* tersebut harus diubah agar menjadi konsisten.

Contoh kasus dapat dilihat pada tabel 4.8 di bawah ini:

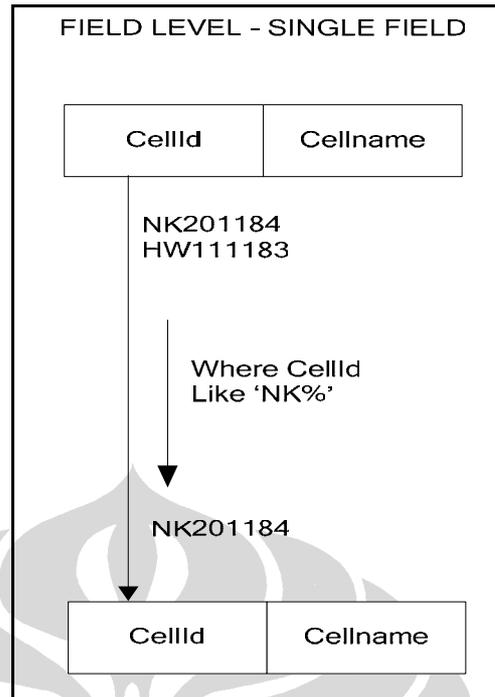
| <i>Cellname</i> | <i>Cellid</i> | <i>Day</i> |
|-----------------|---------------|------------|
| d_bukit_sion1 | 44174 | 02/01/2007 |
| d_bukit_sion2 | 44175 | 02/01/2007 |
| d_bukit_sion3 | 44176 | 02/01/2007 |
| dcs_bukit_sion1 | 44174 | 02/31/2007 |
| dcs_bukit_sion2 | 44175 | 02/31/2007 |
| dcs_bukit_sion3 | 44176 | 02/31/2007 |

Tabel 4.8 – Ketidak-konsistenan Data

4.5.3 Proses Transformasi

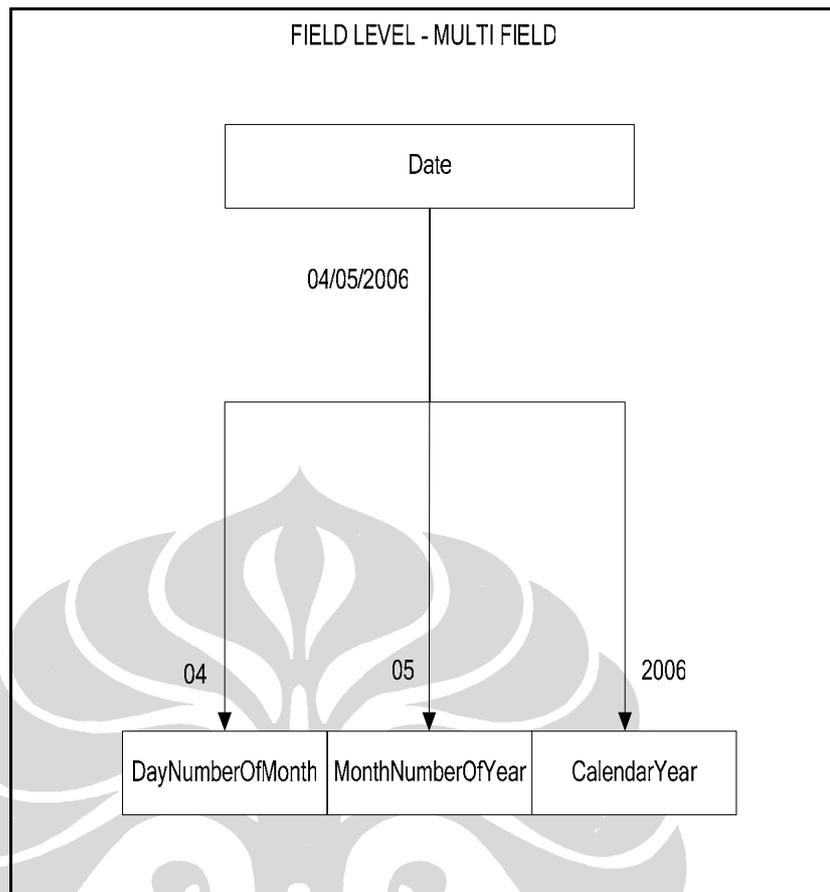
Proses ini dilakukan setelah data yang ada sudah melewati proses ekstraksi dan pembersihan. Proses transformasi yang dilakukan dibagi berdasarkan *level* yaitu *record-level* dan *field-level*, pada proses ini dilakukan proses pemilihan, penggabungan dan agregasi untuk mendapatkan data ringkasan sesuai dengan dimensi yang akan dibuat. Proses transformasi ke *field* yang baru dapat dilakukan dengan menggunakan suatu fungsi tertentu untuk melakukannya.

Contoh dari proses transformasi dengan *record-level* dan *field-level* dapat dilihat pada Gambar 4.6, 4.7 dan 4.8 di bawah ini.



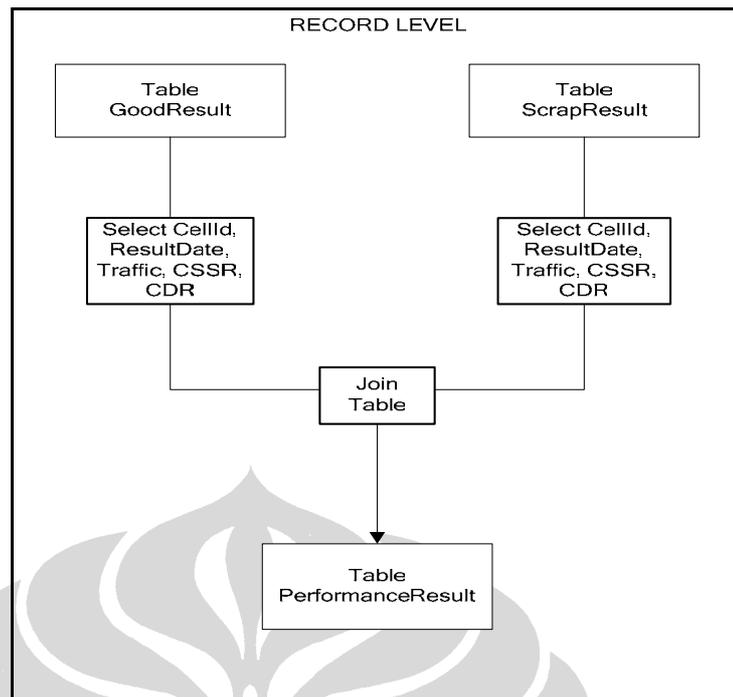
Gambar 4.6 – *Field Level–Single-Field*

Pada Gambar 4.6 menunjukkan proses transformasi *single-field*, yang terjadi pada proses ini adalah data diambil dari *field* sumber (*cellid*) ke *field* tujuan (*cellid*) dengan kriteria bahwa dua huruf pertama dari setiap record harus diawali dengan karakter“NK”, sehingga *cellid* NK201184 akan digunakan sedangkan HW111183 tidak akan digunakan dalam proses tersebut.



Gambar 4.7 – *Field Level – Multiple-Field*

Pada Gambar 4.7 menunjukkan proses transformasi *multiple-field*, proses yang dilakukan adalah mengambil data dari satu *field* sumber (*Date*) kemudian ditransformasikan ke dalam tiga *field* tujuan (*DayNumberOfMonth*, *MonthNumberOfYear*, *CalendarYear*).



Gambar 4.8 – *Record Field*

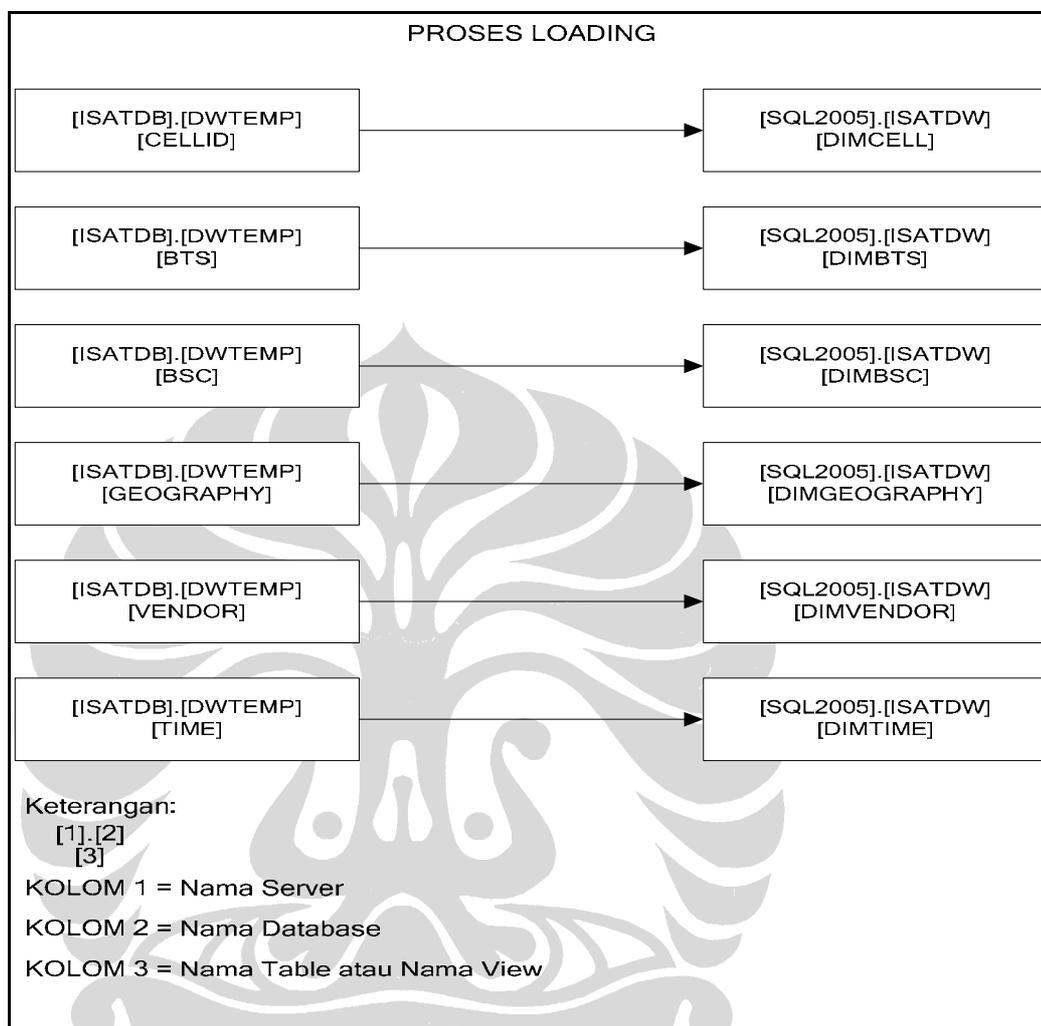
Sedangkan pada Gambar 4.8 ditampilkan proses transformasi *record field*, pada proses diambil beberapa *field* dari tabel sumber (tabel *GoodResult* dan tabel *ScrapResult*) lalu digabungkan ke dalam tabel tujuan (*PerformanceResult*).

4.5.4 Proses Loading

Proses yang dilakukan pada tahap akhir adalah proses pemuatan data (*loading*). Data yang digunakan pada tahap ini merupakan data dari proses-proses yang dilakukan sebelumnya yaitu ekstraksi, pembersihan (*cleaning*) dan transformasi ke dalam *data warehouse*. Cara pemuatan data ke dalam *data*

warehouse adalah dengan menggunakan *script* yang dijalankan secara periodik.

Alur proses *loading* yang dilakukan dapat dilihat pada gambar 4.9 di bawah ini:



Gambar 4.9 – Proses *Loading*

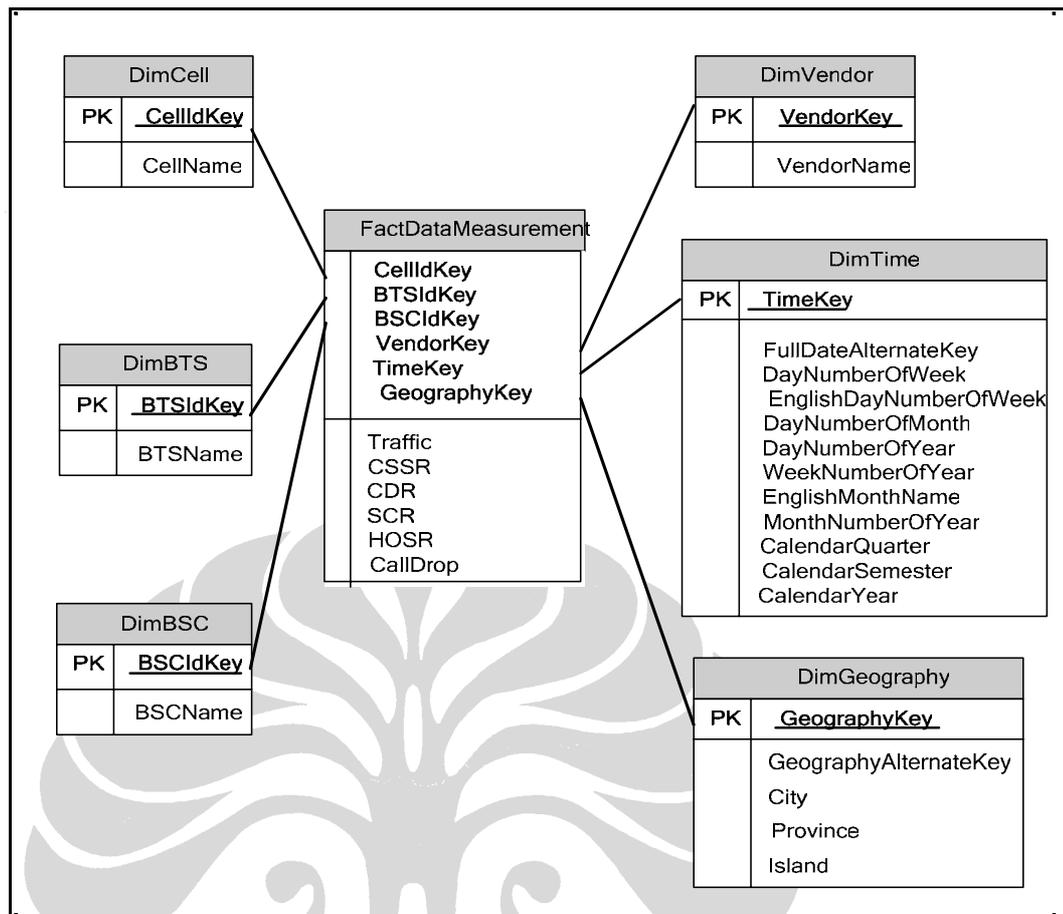
Untuk lebih detilnya mengenai tabel-tabel beserta atribut yang digunakan didalam proses loading dapat dilihat pada lampiran 3 dan lampiran 4

4.5.5 Manajemen Data

Proses manajemen data ini dilakukan oleh *warehouse manager*. Tugas dari manajemen data ini melakukan semua operasi yang berhubungan dengan manajemen data pada *data warehouse*. Operasi yang dilakukan termasuk menganalisis data untuk menjaga konsistensi data (*metadata, detailed data, lightly and highly summarized data*), transformasi dan penggabungan sumber data dari penyimpanan sementara ke dalam tabel *data warehouse* dalam bentuk *lightly and highly summarized data*, membuat index dan view untuk keperluan *data warehouse*, membuat denormalisasi dan agregasi jika diperlukan dalam proses *data warehouse*, dan melakukan *backup data (archive data)*.

4.6 Skema Bintang

Pada sub-bab ini diuraikan komponen *database* dari *data warehouse*. Teknik yang digunakan untuk mendeskripsikannya adalah dengan *dimensional modeling (DM)*, yaitu dengan model dimana di dalamnya terdapat tabel fakta dan tabel dimensi. Model skema yang akan digunakan pada perancangan *data warehouse* adalah skema bintang. Dalam perancangan ini, skema bintang yang digunakan terdiri dari satu tabel fakta dan enam tabel dimensi. Model dari skema bintang ini dapat dilihat pada Gambar 4.10 dibawah ini.



Gambar 4.10 – Star Schema Data Measurement

4.6.1 Tabel Fakta

Tabel fakta yang digunakan dibuat sesuai dengan keperluan perusahaan yaitu hal-hal yang berkaitan dengan proses bisnis yang dijalankan. Dalam penelitian ini yaitu *data measurement* mengenai kinerja jaringan telekomunikasi GSM. Tabel yang akan digunakan yaitu Tabel FactDataMeasurement, tabel ini berisikan data *measurement* yaitu data *traffic*, *call setup success rate* (CSSR), *call drop rate* (CDR), *successful call rate* (SCR), *hand over success rate* (HOSR) dan *call drop*. Kemudian tabel fakta ini juga berisikan mengenai *foreign key* yang

merupakan *primary key* pada tabel dimensi. *Foreign key* tersebut adalah *CellIdKey*, *BTSIdkey*, *BSCIdkey*, *TimeKey*, *VendorKey* dan *GeographyKey*.

4.6.2 Tabel Dimensi

Tabel dimensi adalah tabel yang berisikan data dari berbagai perspektif atau dengan kata lain tabel dimensi adalah tabel yang berisikan *user-defined metada*. Pada gambar 5.1 diatas terdapat enam dimensi, yaitu dimensi *cellid*, dimensi *BTS*, dimensi *BSC*, dimensi *vendor*, dimensi *time*, dimensi *geography*. Dimensi-dimensi itu dipilih dalam upaya untuk meningkatkan kualitas pelaporan yang sudah ada menjadi lebih baik dengan cara memberikan laporan lebih rinci berdasarkan dimensi-dimensi tersebut.

4.6.2.1 Dimensi Cell

Tabel dimensi *cell* adalah tabel yang berisikan data mengenai *cell*. Adapun data yang terdapat dalam tabel dimensi *cell* ini adalah *cellname* yaitu nama dari *cell* yang bersangkutan. Yang menjadi *primary key* pada dimensi *cell* ini adalah *CellIdKey*.

4.6.2.2 Dimensi BTS

Tabel dimensi *BTS* ini berisikan data mengenai *BTS* yaitu *BTSName* dalam hal ini nama dari *BTS* yang bersangkutan. Yang menjadi *primary key* pada dimensi *BTSName* ini adalah *BTSIDkey*.

4.6.2.3 Dimensi BSC

Data yang terdapat pada tabel dimensi *BSC* ini adalah informasi mengenai *BSC* yaitu data *BSCName*, dalam hal ini nama dari *BSC* yang bersangkutan. Yang menjadi *primary key* pada dimensi *BSCName* ini adalah *BSCIDkey*.

4.6.2.4 Dimensi Vendor

Data yang ditampilkan pada dimensi *vendor* ini adalah *VendorKey* yang dijadikan sebagai *primary key* dan *VendorName* yang merupakan nama dari *VendorKey* yang bersangkutan.

4.6.2.5 Dimensi Time

Data yang termasuk dalam tabel dimensi *time* yaitu id tanggal, tanggal secara lengkap (dd-mm-yyyy), periode mingguan (*weekly*), periode bulanan (*monthly*), periode tahunan (*yearly*), periode semester (*semester*), periode kuartal (*quarterly*).

4.6.2.6 Dimensi Geography

Data yang termasuk dalam tabel dimensi *geography* ini adalah nama-nama kota, kabupaten dan provinsi yang berada di wilayah Indonesia beserta dengan ibukota-nya termasuk pulau-pulau yang ada di Indonesia.

BAB V

IMPLEMENTASI DATA WAREHOUSE

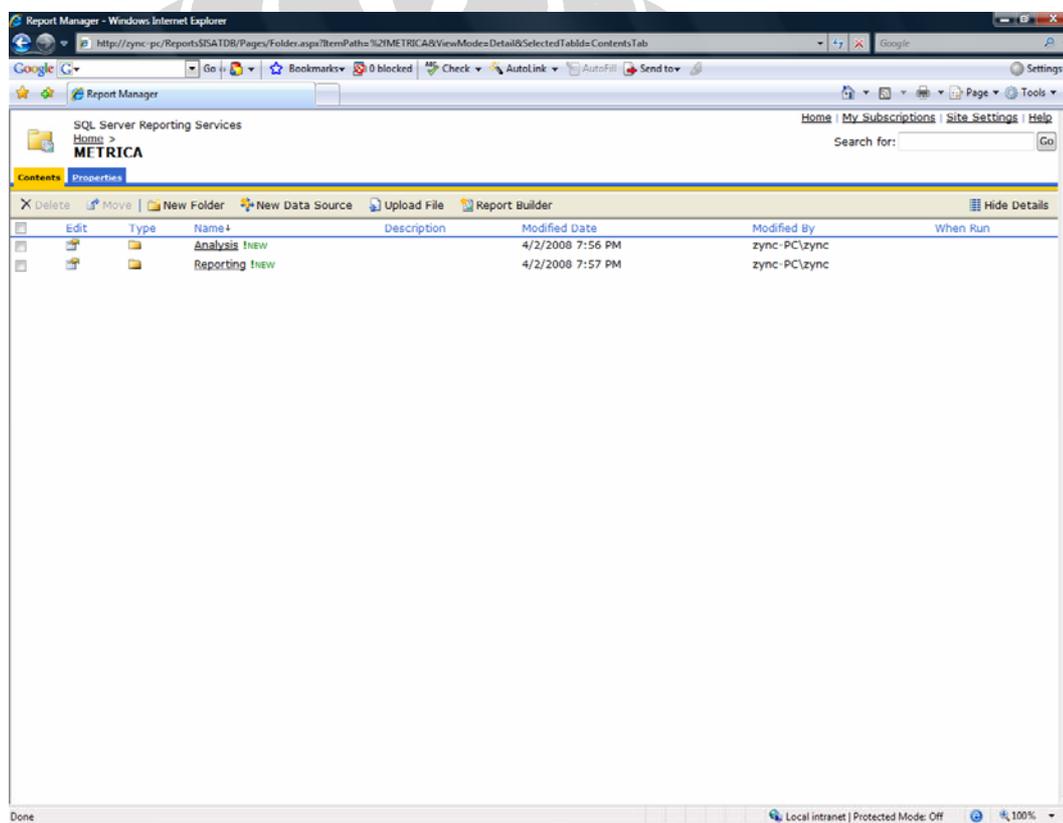
Pada bab ini dijelaskan mengenai kemampuan penyediaan informasi dari *data warehouse* untuk melakukan analisis kualitas jaringan GSM dari berbagai sudut pandang.

5.1 Presentasi *Data Warehouse*

Untuk menunjang proses analisis yang akan dilakukan oleh pengguna maka *data warehouse* yang telah dirancang dengan menggunakan skema bintang perlu dipresentasikan kepada pengguna. *Tools* perangkat lunak yang digunakan untuk menunjang presentasi tersebut adalah *Reporting Service SQL Server 2005* dan *Microsoft IIS 6.0* yang diintegrasikan dengan *platform database* yang digunakan pada *data warehouse*. Sedangkan media yang digunakan untuk menampilkan presentasi tersebut adalah menggunakan media *web*. Dengan menggunakan media *web* diharapkan dapat mempermudah pengguna dalam melihat informasi tersebut, kapan saja dan dimana saja selama terhubung dengan jaringan dan memiliki *internet browser*.

Kelebihan yang dapat diperoleh jika menggunakan *tools Reporting Service SQL Server 2005* adalah terdapat fasilitas untuk *roll-up* maupun *drill-down* dan *measure* yang dapat digunakan untuk menganalisa data dan menampilkannya dalam bentuk diagram (*chart*). *Roll-up* adalah data yang ada ditampilkan dengan

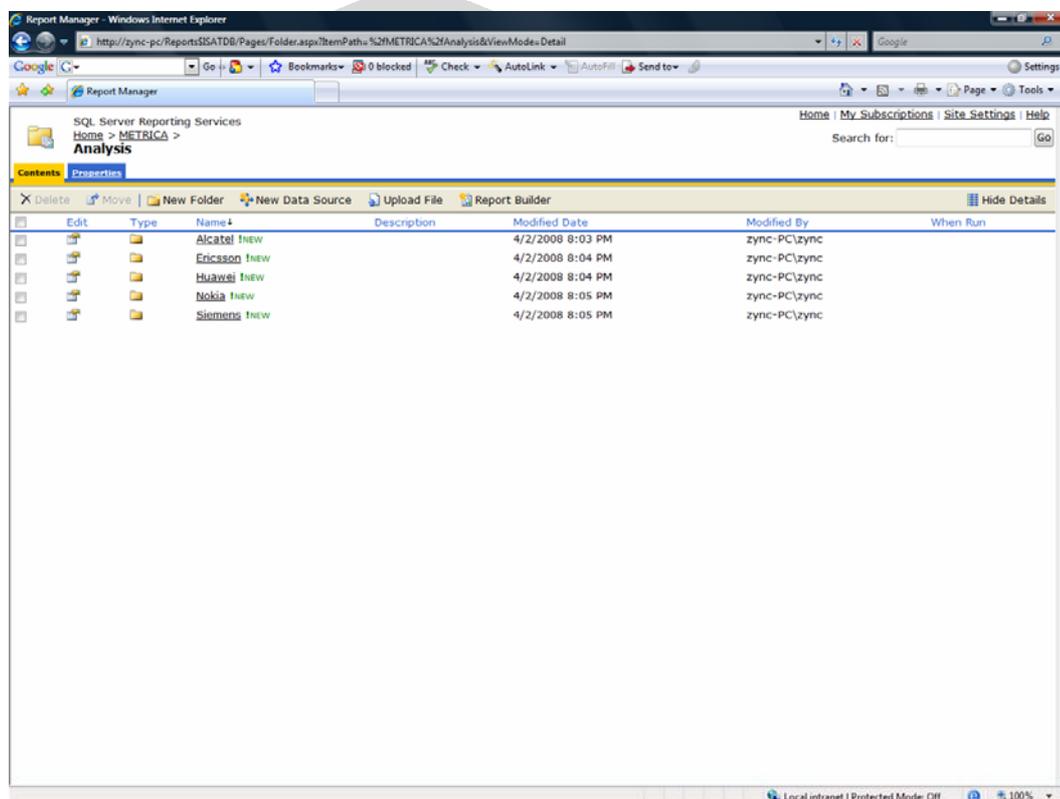
tingkat rincian yang rendah, sedangkan *drill-down* adalah data yang ada ditampilkan dengan tingkat rincian yang tinggi dan *measure* digunakan untuk menampilkan data yang merupakan hasil perhitungan atau kalkulasi. Melalui *tools* tersebut pengguna juga dapat melakukan dan mengeksekusi *query* dengan mudah yang dalam pelaksanaannya dilakukan oleh *query manager*. Berikut ini ditampilkan pada Gambar 5.1 yaitu halaman *web* yang merupakan halaman utama dalam pembuatan *data warehouse* ini. Dalam halaman ini pengguna dapat memilih dari dua sub menu yang tersedia yaitu *analysis* dan *reporting*.



Gambar 5.1 – Halaman Utama *Data Warehouse*

Pada sub menu pertama yaitu menu *analysis* yang dapat dilihat pada pada Gambar 5.2, terdapat pembagian *analysis* menjadi lima *folder* yang dibagi berdasarkan masing-masing *vendor*. Pembagian ini dibagi berdasarkan *vendor* dikarenakan kebutuhan *analysis* terhadap data kinerja jaringan dilakukan sesuai

dengan target *Key Performance Indicator* (KPI) yang berbeda dari masing-masing *vendor*. Dalam masing-masing *folder* terdapat *file* yang menampilkan laporan analisis kinerja jaringan sesuai data yang diperoleh dari masing-masing elemen jaringan per *vendor*, contoh ini dapat dilihat pada Gambar 5.3. Bagian *analysis* ini sangat membantu pengguna, dikarenakan dalam bagian ini pengguna dapat melihat informasi yang ditampilkan secara menyeluruh mengenai data kinerja jaringan sesuai *vendor* yang dipilih.



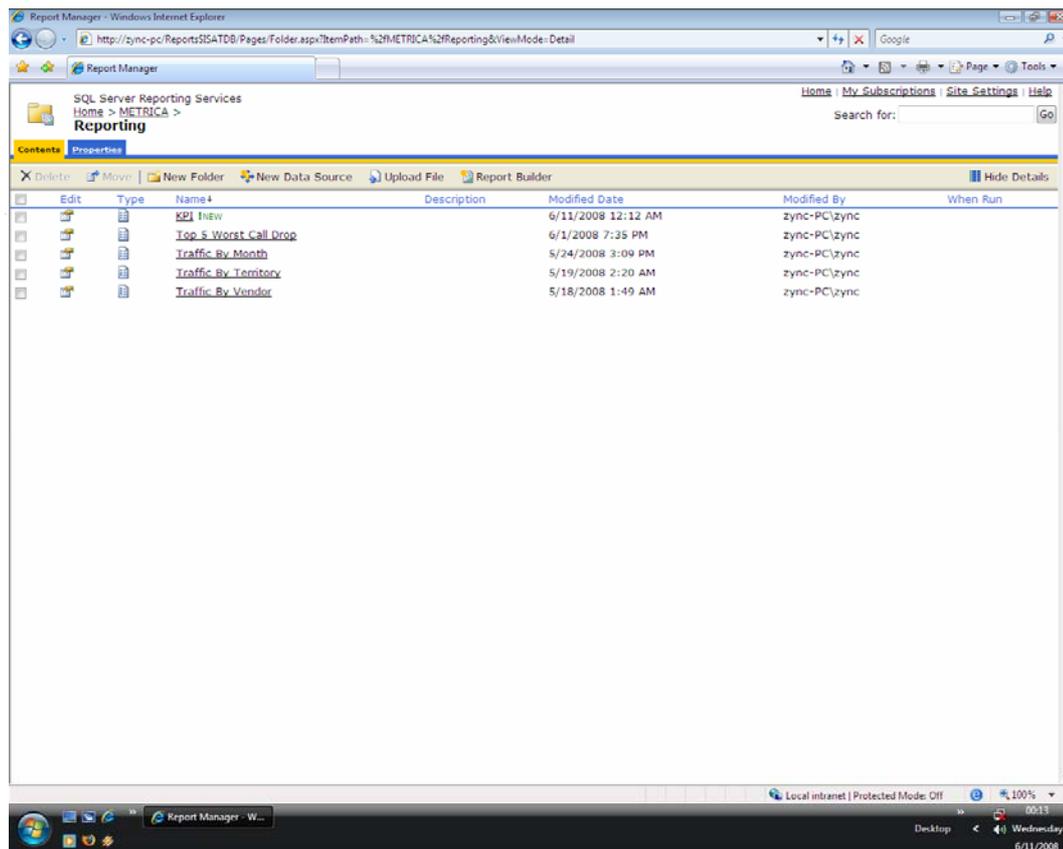
Gambar 5.2 – *Menu Analysis*

Report Manager - Windows Internet Explorer
 http://zync-ipc/Reports/SATDB/Pages/Report.aspx?ItemPath=%2fMETRICA%2fAnalysis%2fNokia%2fAllBranch.Daily_041007
 SQL Server Reporting Services
 Home > METRICA > Analysis > Nokia > AllBranch.Daily_041007
 Performance Analysis NOKIA
 DATE: 4/10/2007

| CellId | Cellname | BSCName | City | Traffic | CSSR | CDR | SCR | HOSR | Calldrop |
|---------|-----------------|---------------|---------|---------|--------|-------|--------|--------|----------|
| NK12011 | CICADAS1 | SADANG_SRG_01 | Bandung | 98.47 | 99.02% | 0.53% | 98.49% | 97.27% | 37 |
| NK12012 | CICADAS2 | SADANG_SRG_01 | Bandung | 123.25 | 98.78% | 0.48% | 98.30% | 95.48% | 48 |
| NK12013 | CICADAS3 | SADANG_SRG_01 | Bandung | 54.47 | 98.71% | 0.75% | 97.98% | 96.38% | 30 |
| NK12014 | DCS_CICADAS1 | SADANG_SRG_01 | Bandung | 17.64 | 98.02% | 0.73% | 97.30% | 94.91% | 10 |
| NK12015 | DCS_CICADAS2 | SADANG_SRG_01 | Bandung | 25.22 | 99.17% | 0.59% | 98.59% | 98.53% | 14 |
| NK12016 | DCS_CICADAS3 | SADANG_SRG_01 | Bandung | 27.42 | 99.33% | 0.31% | 99.02% | 96.28% | 6 |
| NK12021 | CIDAP1 | SADANG_SRG_02 | Bandung | 14.49 | 99.67% | 0.25% | 99.42% | 96.87% | 3 |
| NK12022 | CIDAP2 | SADANG_SRG_02 | Bandung | 37.88 | 93.86% | 2.11% | 91.89% | 93.85% | 65 |
| NK12023 | CIDAP3 | SADANG_SRG_02 | Bandung | 26.56 | 98.00% | 1.52% | 96.51% | 92.88% | 35 |
| NK12031 | AWILIGAR1 | SADANG_SRG_01 | Bandung | 24.17 | 98.61% | 0.35% | 98.26% | 96.59% | 8 |
| NK12032 | AWILIGAR2 | SADANG_SRG_01 | Bandung | 91.68 | 98.88% | 0.57% | 98.31% | 97.28% | 44 |
| NK12033 | AWILIGAR3 | SADANG_SRG_01 | Bandung | 85.32 | 98.63% | 0.57% | 98.06% | 96.19% | 33 |
| NK12034 | DCS_AWILIGAR1 | SADANG_SRG_01 | Bandung | 0 | | | 0.00% | | 0 |
| NK12035 | DCS_AWILIGAR2 | SADANG_SRG_01 | Bandung | 0 | | | 0.00% | | 0 |
| NK12036 | DCS_AWILIGAR3 | SADANG_SRG_01 | Bandung | 0 | | | 0.00% | | 0 |
| NK12041 | CIKAPAYANG1 | SADANG_SRG_01 | Bandung | 35.97 | 98.64% | 0.81% | 97.84% | 96.32% | 17 |
| NK12042 | CIKAPAYANG2 | SADANG_SRG_01 | Bandung | 33.26 | 97.70% | 1.19% | 96.54% | 97.22% | 24 |
| NK12043 | CIKAPAYANG3 | SADANG_SRG_01 | Bandung | 18.78 | 98.73% | 0.81% | 97.93% | 97.67% | 11 |
| NK12044 | DCS_CIKAPAYANG1 | SADANG_SRG_01 | Bandung | 104.68 | 99.14% | 0.21% | 98.93% | 99.29% | 13 |
| NK12045 | DCS_CIKAPAYANG2 | SADANG_SRG_01 | Bandung | 86.8 | 99.28% | 0.27% | 99.02% | 98.83% | 12 |
| NK12046 | DCS_CIKAPAYANG3 | SADANG_SRG_01 | Bandung | 81.42 | 99.52% | 0.25% | 99.27% | 99.31% | 13 |
| NK12051 | CIKUTRA1 | SADANG_SRG_01 | Bandung | 57.81 | 98.90% | 0.59% | 98.32% | 97.07% | 31 |
| NK12052 | CIKUTRA2 | SADANG_SRG_01 | Bandung | 30.76 | 98.11% | 0.88% | 97.25% | 95.14% | 21 |
| NK12053 | CIKUTRA3 | SADANG_SRG_01 | Bandung | 24.97 | 98.87% | 0.85% | 98.03% | 94.94% | 17 |
| NK12054 | DCS_CIKUTRA1 | SADANG_SRG_01 | Bandung | 128.02 | 99.49% | 0.29% | 99.20% | 99.13% | 24 |

Gambar 5.3 – Contoh *File* pada *Menu Analysis*

Pada sub menu kedua yaitu bagian *reporting* merupakan bagian yang berisikan *report* yang lebih statis dan spesifik, seperti yang terlihat pada Gambar 5.4. Pada bagian ini pengguna dapat melihat laporan secara terperinci dan spesifik sesuai dengan kebutuhan laporan spesifik perusahaan. Contohnya adalah *report* mengenai KPI (mengenai status pencapaian KPI), *Top 5 Worst Call Drop* (*cell* yang memiliki *call drop* terbanyak), *traffic by month* (*traffic* per bulan), *traffic by territory* (*traffic* per wilayah), *traffic by vendor* (*traffic* per vendor). Data yang ditampilkan disini merupakan data yang termasuk kategori *old detail data* (data lama) karena data yang ditampilkan merupakan data pada waktu yang lampau. Data dapat juga ditampilkan dalam bentuk *current (new) detail data* (data saat ini atau baru) jika data yang ditampilkan merupakan data pada waktu saat ini.



Gambar 5.4 – *Menu Reporting*

5.1.1 Informasi *Traffic*

Salah satu tolak ukur kinerja jaringan adalah utilisasi yang dapat dilihat dari informasi *traffic* dari tiap elemen jaringan. Berikut ini adalah jenis-jenis laporan yang berkaitan dengan informasi *Traffic*.

- ✚ Laporan *Traffic* per-vendor (Gambar 5.5)
- ✚ Laporan *Traffic* per-wilayah (Gambar 5.6)
- ✚ Laporan *Traffic* keseluruhan (Gambar 5.7)

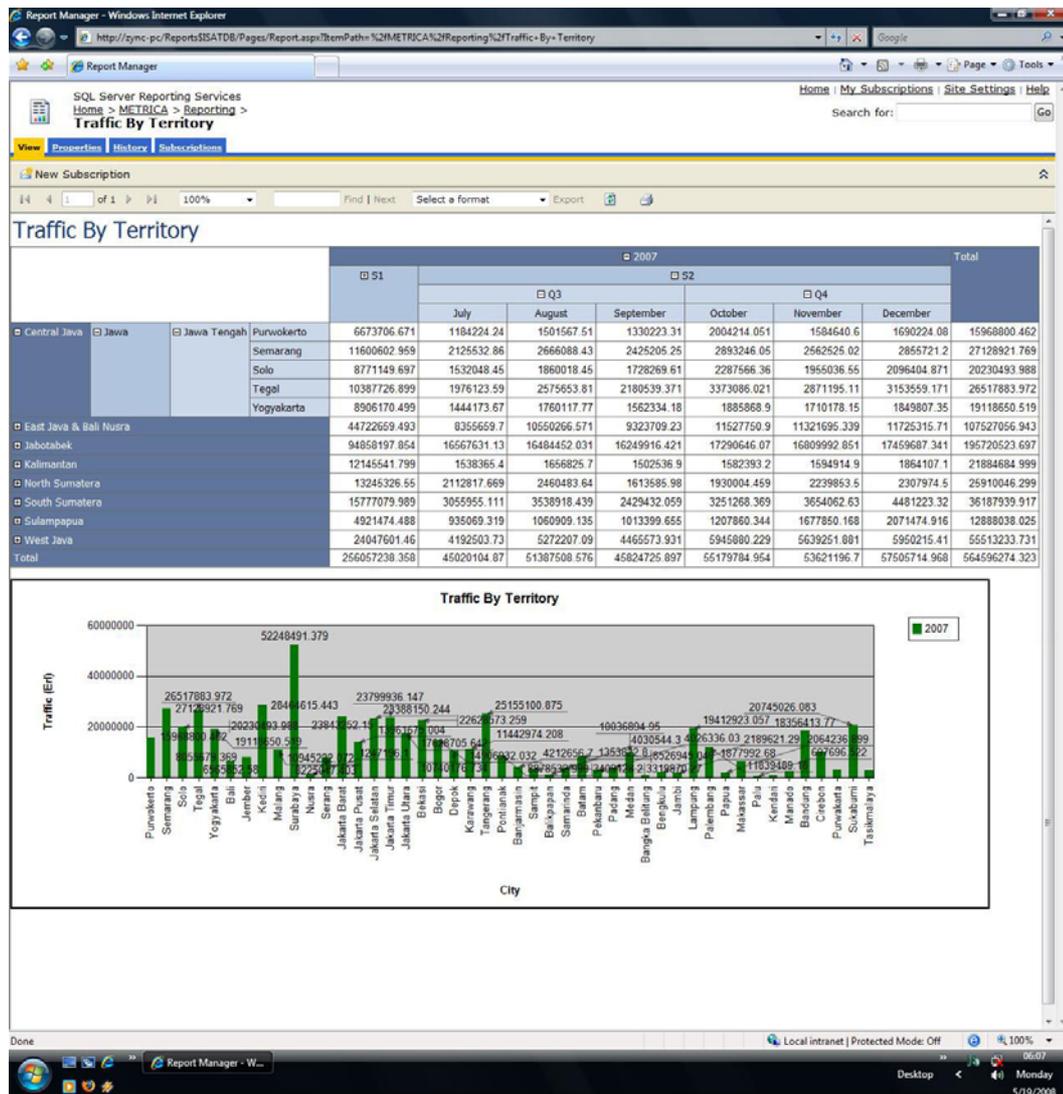
Gambar 5.5 menyajikan informasi *traffic* per-vendor secara terperinci, dimana dengan informasi ini pihak PT. Indosat Tbk dapat mengetahui tren *traffic* dan utilisasi jaringan dari tiap-tiap *network element* sesuai masing-masing *vendor*.

| | S1 | 2007 | | | | | | Total |
|----------|---------------|--------------|--------------|--------------|--------------|--------------|--------------|---------------|
| | | Q3 | | | Q4 | | | |
| | | July | August | September | October | November | December | |
| ALCATEL | 9943066.099 | 1213714.4 | 1279419.7 | 1116239.1 | 1181836.7 | 1200565.4 | 1371237.9 | 17306079.299 |
| ERICSSON | 71614096.183 | 12265397.689 | 11869739.699 | 11869739.699 | 12265397.689 | 11869739.699 | 12265397.689 | 144019508.345 |
| HUAWEI | 4905470.843 | 921180.711 | 1045942.682 | 997541.576 | 1188948.841 | 1656568.748 | 2039815.146 | 12755468.547 |
| NOKIA | 107964035.367 | 19844286.46 | 25081952.401 | 21736976.311 | 28604049.01 | 26332096.981 | 27908495.1 | 257471891.632 |
| SIEMENS | 28067131.028 | 5030854.78 | 5866313.32 | 3960304.649 | 5079046.499 | 5781030.26 | 6654455.05 | 60439135.587 |
| Total | 222493799.521 | 39275434.04 | 45143367.802 | 39680801.335 | 48319278.739 | 46840001.088 | 50239400.886 | 491992083.41 |

Gambar 5.5 – Laporan *Traffic* per-vendor

Data yang ditampilkan disini merupakan *lightly summarized data* (data yang disimpulkan secara ringan) contohnya adalah data *traffic* per bulan yang disimpulkan dari data total *traffic* harian di dalam satu bulan. Kemudian *lightly summarized data* yang disimpulkan menjadi *highly summarized data* (data yang disimpulkan secara berat) contohnya adalah data *traffic* pada tahun 2007 yang merupakan data total *traffic* bulanan yang disimpulkan menjadi total *traffic* pada tahun 2007.

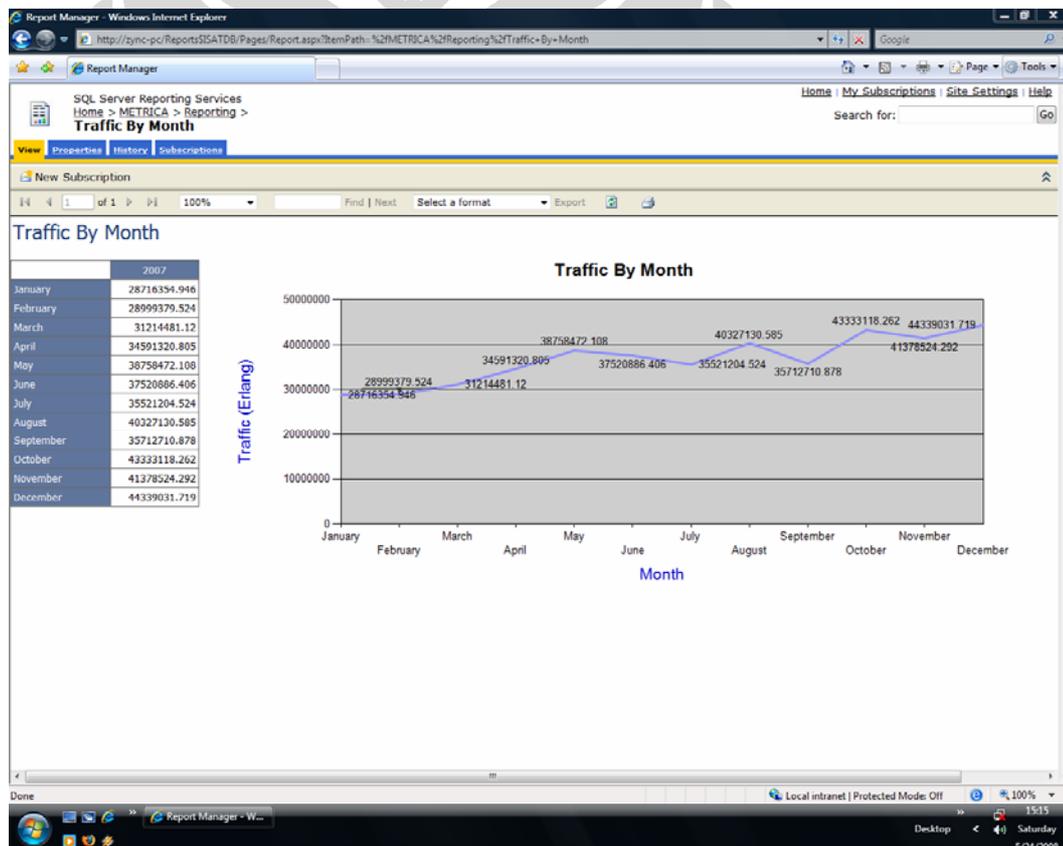
Selain melihat laporan *traffic* per-*vendor*, melalui presentasi *data warehouse*, dalam tampilan *website* ini kita dapat menyajikan data *traffic* kepada pihak PT. Indosat Tbk dalam bentuk grafikal, seperti yang dapat dilihat pada Gambar 5.6 dan Gambar 5.7.



Gambar 5.6 – *Traffic* per-wilayah

Pada Gambar 5.6 informasi yang ditampilkan adalah data *traffic* per wilayah, dengan informasi ini pihak PT. Indosat Tbk dapat melihat wilayah mana saja yang nilai trafiknya tertinggi dan wilayah mana saja yang angka trafiknya

terendah yang dapat di monitor sesuai periode (bulanan, kuartal, semester, tahunan) yang diinginkan, sehingga melalui informasi ini PT. Indosat Tbk dapat melakukan perencanaan strategi investasi seperti keperluan untuk meng-*upgrade* BTS, ekspansi jaringan yang ada agar dapat memenuhi permintaan *traffic* yang lebih tinggi dari sebelumnya. Untuk keperluan melihat kinerja dari jaringan yang dimiliki oleh PT. Indosat Tbk diperlukan suatu laporan yang dapat menampilkan informasi *traffic* dari bulan ke bulan selama satu tahun untuk melihat tren *traffic* yang ada. Hal ini digambarkan pada Gambar 5.7 yaitu laporan mengenai *traffic* berdasarkan waktu.



Gambar 5.7 – *Traffic* per-bulan

Melalui laporan ini terlihat terjadi kenaikan jumlah *traffic* pada bulan januari sampai dengan mei kemudian terjadi penurunan *traffic* dan kenaikan jumlah *traffic* yang terus berulang dari juli hingga desember. Melalui laporan ini pihak perusahaan dapat melihat tren *traffic* yang terjadi per bulan selama satu tahun sehingga dapat dijadikan sebagai bahan analisis untuk rencana strategis tahun berikutnya seperti perencanaan untuk marketing promo.

5.1.2 Informasi Alert

Laporan yang akan ditampilkan dibawah ini adalah laporan yang berkaitan dengan *alert* dari elemen jaringan yang dipantau secara berkala. Berikut ini adalah jenis-jenis laporan yang berkaitan dengan informasi *alert*:

Laporan *Top 5 Worst Call Drop*

Laporan ini menggambarkan informasi mengenai jumlah *call drop* lima teratas yang terburuk dari seluruh elemen jaringan yang ada. Informasi ini diberikan mencakup level regional, BSC dan *cell* yang dimonitor secara harian.

| | Traffic | CSSR | CDR | SCR | HOSR | Call Drop |
|------------------------|---------|--------|-------|--------|---------|-----------|
| Central Java | | | | | | |
| TEGAL_01 | | | | | | |
| DCSKALIGANGSA2 | 121.49 | 98.12% | 9.67% | 88.62% | 98.29% | 1128 |
| DCSDUKUWARU3 | 110.65 | 97.31% | 2.45% | 94.93% | 98.58% | 192 |
| BULAKAMBA2 | 107.56 | 95.12% | 2.25% | 92.98% | 96.79% | 195 |
| KERSAJIA1 | 138.42 | 96.71% | 2.06% | 94.72% | 97.52% | 232 |
| DCSSLAWIZ | 123.84 | 98.28% | 1.87% | 96.44% | 97.54% | 191 |
| PURWODADI | 2254.32 | 86.31% | 8.99% | 78.55% | 94.85% | 3113 |
| CILACAP | 9100.66 | 91.70% | 3.76% | 88.26% | 80.66% | 7101 |
| KUDUS_02 | 472.68 | 96.16% | 3.11% | 93.17% | 96.42% | 677 |
| PEKALONGAN_02 | 9911.56 | 96.90% | 2.71% | 94.35% | 91.46% | 7624 |
| TEGAL_02 | 3070.6 | 96.25% | 2.55% | 94.01% | 94.87% | 3149 |
| PEKALONGAN_01 | 4436.29 | 96.39% | 2.43% | 94.53% | 93.45% | 4035 |
| KUDUS | 3979.6 | 95.85% | 2.19% | 94.01% | 95.00% | 3708 |
| DELANGGU_01 | 5711.29 | 93.69% | 2.02% | 91.87% | 88.37% | 5932 |
| BAHTUL_01 | 645.76 | 97.52% | 1.83% | 95.74% | 96.22% | 633 |
| CEMAHI | 5622.32 | 94.29% | 1.82% | 92.75% | 94.97% | 5382 |
| DELANGGU_02 | 3471.83 | 93.46% | 1.82% | 91.88% | 77.87% | 3527 |
| MUNTILAH_01 | 1125.93 | 96.01% | 1.68% | 94.39% | 96.35% | 1371 |
| MIRICAI_01 | 2566.62 | 91.81% | 1.62% | 90.43% | 94.64% | 2238 |
| BATURADEN | 1718.01 | 96.12% | 1.59% | 94.60% | 87.68% | 1245 |
| PDK_PVG | 2875.98 | 95.43% | 1.56% | 93.95% | 93.34% | 3061 |
| PITPN | 4168.04 | 95.85% | 1.54% | 94.88% | 94.71% | 3493 |
| KEBUMEN | 505.68 | 97.79% | 1.53% | 96.30% | 98.07% | 412 |
| GOMBEL_01 | 2302.17 | 95.50% | 1.25% | 94.62% | 96.49% | 1739 |
| TUGU | 793 | 98.43% | 1.25% | 97.21% | 97.72% | 471 |
| SIDOARUM | 1277.87 | 97.71% | 1.17% | 96.60% | 96.06% | 1040 |
| ASPAC | 1396.86 | 97.29% | 1.00% | 96.43% | 94.22% | 927 |
| GOMBEL_02 | 1385.32 | 97.09% | 0.95% | 96.19% | 97.51% | 752 |
| Kalimantan | | | | | | |
| EVO_G_080S1 | 981.2 | 97.49% | 5.66% | 91.98% | 95.28% | 1472 |
| EVO_BATU_LAYANG1 | 1037.3 | 98.56% | 2.46% | 96.29% | 87.70% | 727 |
| EVO_TERMINIALSINTAING1 | 1333.6 | 97.18% | 2.21% | 96.38% | 97.45% | 933 |
| EVO_KETAPANG1 | 789.8 | 98.25% | 1.92% | 97.04% | 100.00% | 689 |
| EVO_PASAR_BERAU1 | 136.8 | 99.40% | 1.81% | 97.60% | 100.00% | 119 |
| EVO_SIOLATAMA1 | 293.4 | 99.22% | 1.13% | 98.09% | 99.29% | 266 |
| EVO_COHDONG1 | 323.7 | 96.69% | 0.86% | 95.86% | 100.00% | 139 |

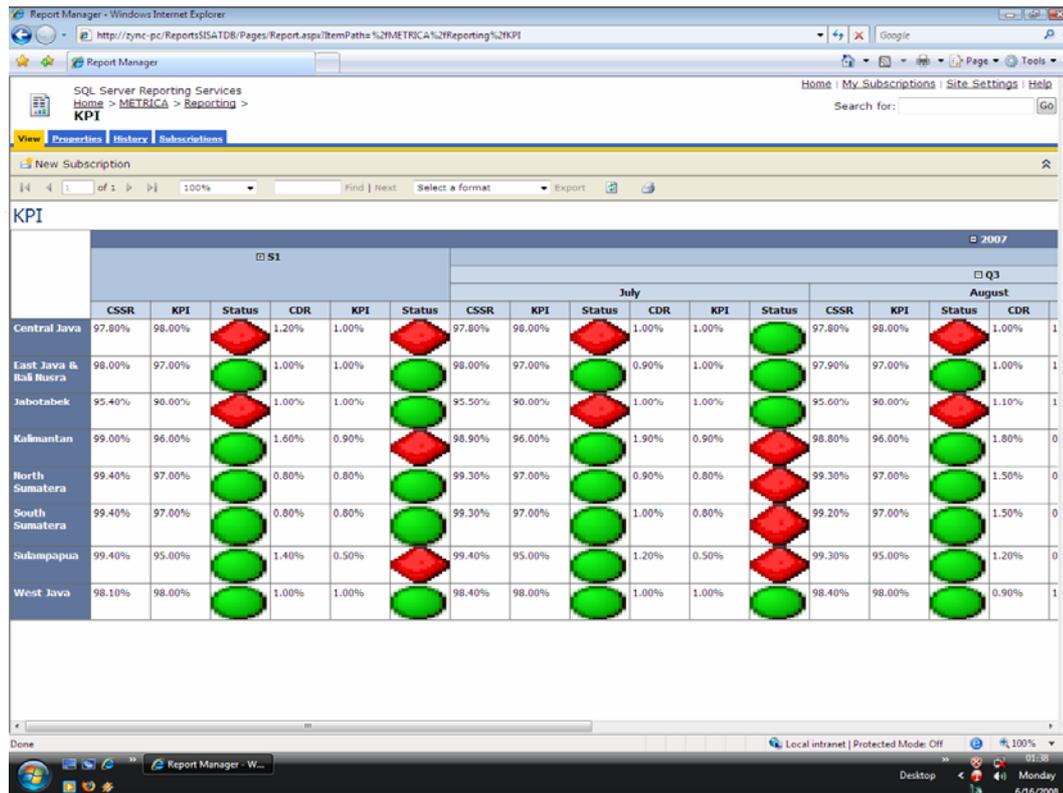
Gambar 5.8 – Top 5 Worst Call Drop

Dari contoh diatas pada posisi *rollup* diperoleh informasi mengenai *call drop* terburuk dari masing-masing *cell* dalam suatu BSC, sedangkan setelah di *drill down* diperoleh informasi *cell* yang memiliki jumlah *call drop* terburuk lima teratas dari masing-masing level BSC. Aturan dalam pemilihan data yang merupakan *call drop* terburuk ini sendiri yaitu dari seluruh data *cell* yang dimiliki dicari data dengan kondisi *traffic* dan *call drop* yang memenuhi (lebih besar) dari suatu batas tertentu baru kemudian data di atur secara menurun dari besar ke terkecil berdasarkan CDR, barulah kemudian diambil data lima *cell* teratas. Sehingga dari informasi *cell* terburuk tersebut dapat dicari lebih dalam

permasalahan yang terjadi yang kemudian dicarikan solusinya untuk meningkatkan kinerja dari *cell* tersebut.

Laporan KPI

Laporan KPI ini menghadirkan informasi mengenai CSSR dan CDR dari masing-masing regional yang dapat dilihat dalam periode tahunan, semester, kuartal dan bulanan. Selain itu pada laporan ini terdapat target KPI dari masing-masing regional sehingga setiap informasi yang diperoleh mengenai CSSR dan CDR dapat dibandingkan langsung dengan target KPI yang ada kemudian hasilnya direpresentasikan dengan gambar pada kolom status. Untuk CSSR Jika angka yang diperoleh lebih besar atau sama dengan target KPI yang ada maka pada status direpresentasikan dengan gambar ● berwarna hijau yang menandakan target tercapai sedangkan jika angka yang diperoleh lebih kecil maka pada status direpresentasikan dengan gambar ◆ berwarna merah yang menandakan target tidak tercapai. Sementara untuk data CDR sebaliknya karena data CDR menunjukkan jumlah rasio panggilan yang terputus sehingga target yang ingin dicapai harus lebih kecil atau sama dengan target KPI yang ditentukan. Jika angka yang diperoleh lebih besar dari target KPI yang ada maka pada status direpresentasikan dengan simbol ◆ berwarna merah yang menandakan target tercapai sedangkan jika angka yang diperoleh lebih kecil atau sama dengan target KPI maka pada status direpresentasikan dengan simbol ● berwarna hijau yang menandakan target tercapai. Untuk lebih jelasnya dapat dilihat pada Gambar 5.9 berikut ini.



Gambar 5.9 – KPI Alert

5.1.3 Benchmark Report

Untuk melihat tingkat efisiensi waktu pembuatan *report* sebelum dan sesudah dilakukannya implementasi *data warehouse*, berikut ini pada Tabel 5.1 dapat dilihat perbandingan keduanya.

| Jenis Report | Sebelum | Sesudah |
|------------------------------|----------|------------|
| <i>Performance Analysis</i> | ± 6 Jam | ± 5 Menit |
| <i>Traffic per-vendor</i> | ± 8 Jam | ± 10 Menit |
| <i>Traffic per-wilayah</i> | ± 12 Jam | ± 10 Menit |
| <i>Traffic keseluruhan</i> | ± 2 Jam | ± 10 Menit |
| <i>Top 5 Worst Call Drop</i> | ± 7 Jam | ± 15 Menit |
| <i>KPI Alert</i> | ± 6 Jam | ± 5 Menit |

Tabel 5.1 – Benchmark Report

BAB VI

IMPLEMENTASI *DATA MINING*

Pada bab ini dijelaskan mengenai implementasi *data mining* untuk mendapatkan dan menggali informasi yang tersembunyi dalam *data warehouse*. Untuk mendapatkan hasil yang baik dan bermanfaat diperlukan suatu proses *data mining* yang terstruktur dengan baik melalui tahapan *data mining*. Karena itu pada bab ini akan dijabarkan bagaimana implementasi *data mining* dijalankan melalui tahapan *data mining* yang telah dijelaskan sebelumnya pada bab II. Untuk lebih jelasnya dapat dilihat pada poin-poin dibawah ini:

✚ *Cleansing Data*

Pada tahap ini akan dilakukan pembersihan data untuk membuang data yang tidak konsisten dan menghilangkan *noise*. Data yang diperoleh melalui proses ekstraksi dari sumber data seperti *operational data store* dan *flat file* (contoh: *sheet file excel*) akan dibersihkan pada tahap ini sebelum memasuki tahap selanjutnya integrasi. Penjelasan mengenai tahap *cleansing data* ini sebelumnya telah dijelaskan pada bagian proses ETL pada *data warehouse* di bab IV

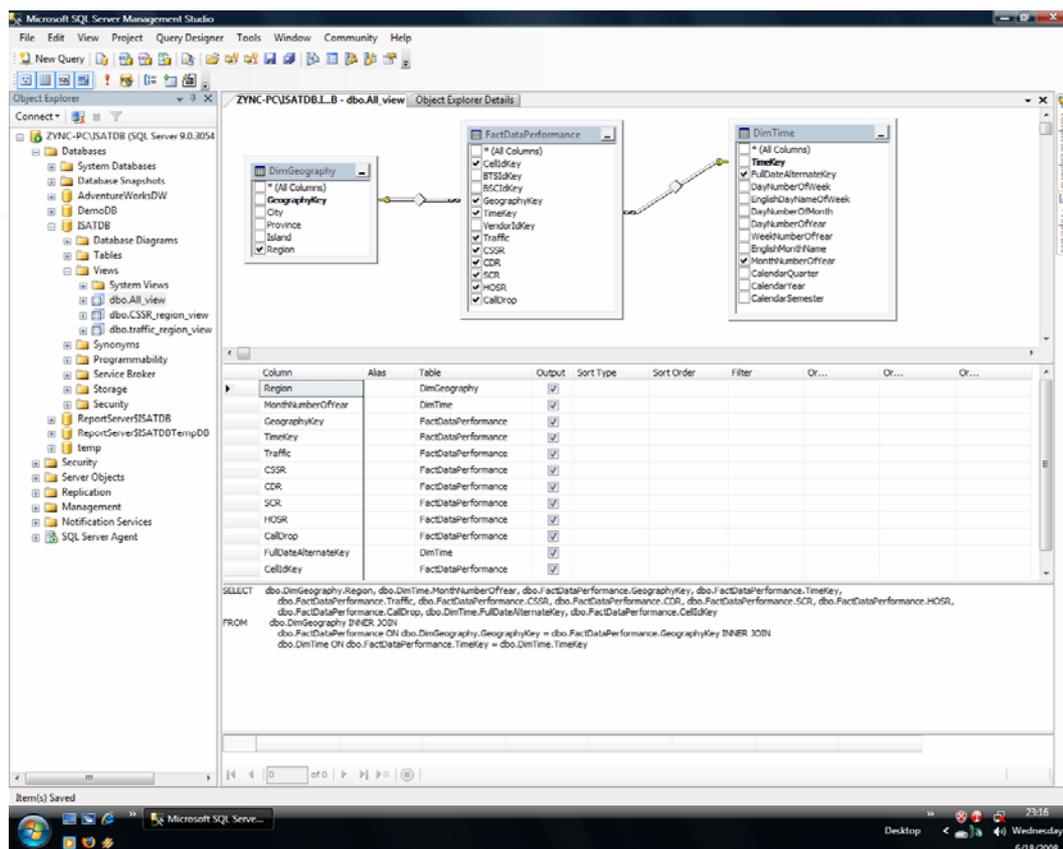
✚ Integrasi Data

Setelah melalui proses *cleansing*, data yang telah bersih disiapkan untuk proses integrasi. Tahap integrasi data ini berkaitan dengan sumber data yang digunakan pada tahapan *data mining*. Sumber data yang digunakan tidak hanya berasal dari satu sumber saja, melainkan dapat diperoleh dari sumber yang beragam. Untuk keperluan penelitian ini sumber data yang

akan digunakan yaitu *database* dan *sheet file excel*. Karena bentuk format struktur data kedua sumber data itu berbeda proses untuk mengintegrasikannya memerlukan pengerjaan yang cermat. Sebagai contoh kasus *fielcd* dari masing-masing sumber data bisa saja dalam penamaan yang berbeda tetapi maksudnya adalah sama. Seperti pada penelitian ini diperoleh *field traffic* dari *database* dan *field call traffic* dari *sheet file excel*, jika keduanya dianggap berbeda maka hal ini akan memberikan pemahaman yang tidak tepat, karena sebenarnya keduanya memiliki arti pemahaman yang sama. Tahap integrasi data ini melibatkan proses transformasi yang kemudian dilanjutkan dengan proses *loading* ke *data warehouse* yang telah dijelaskan lebih rinci pada bab IV sebelumnya

📌 Seleksi Data

Setelah melalui proses *cleansing*, integrasi data dan *loading* berarti sumber data telah berhasil melalui ketiga tahap tersebut dan sudah disimpan di dalam *data warehouse*. Data tersebut kemudian pada tahap seleksi data ini dilakukan pemilihan dan pengambilan terhadap data yang relevan terhadap proses analisis yang akan dilakukan. Pada penelitian ini penerapan teknik *data mining* yang akan digunakan adalah teknik *time series analysis* karena itu pada penelitian ini, dari sejumlah data yang digunakan (lampiran 4) dilakukan pemilihan data dari dalam *data warehouse* (ISATDW) sesuai dengan proses analisis yang diperlukan. Data-data yang akan digunakan adalah *Date*, *CellIdKey*, *Traffic*, *CSSR*, *CDR*, *SCR*, *HOSR*, *Call Drop* yang dapat dilihat pada Gambar 6.1 dibawah ini:



Gambar 6.1 – Seleksi Data

Pada bagian ini dilakukan pemilihan

Transformasi Data

Pada tahap ini data yang sudah di seleksi pada tahap sebelumnya harus di transformasi sebelum data bisa langsung di aplikasikan pada tahap *data mining*. Transformasi data pada penelitian ini melibatkan hal-hal berikut:

- *Aggregation*: dimana *summary* atau *aggregation* di aplikasikan ke data penelitian contohnya data *traffic* harian dan *call drop* harian masing-masing di agregasikan untuk menghitung total *traffic* dan total *call drop* dalam satu bulan. Data CSSR, SCR, CDR dan HOSR masing-masing di rata-ratakan untuk mendapatkan rata-rata di dalam satu bulan

- *Generalization*: generalisasi data yang digunakan adalah atribut kategori *city* yang bisa digeneralisasi ke level lebih tinggi yaitu *province*. *Province* yang bisa digeneralisasi ke level lebih tinggi *region* dan *region* yang bisa digeneralisasi ke level lebih tinggi lagi yaitu *country*
- *Missing value handling*: set data yang hilang dalam penelitian ini dapat disebabkan karena proses integrasi antar dua sumber yang berbeda tidak memiliki definisi atribut yang sama atau memang nilai yang ada kosong karena memang tidak ada data pengukuran yang diterima. Untuk menghadapi permasalahan, penulis mengganti nilai yang kosong dengan mengambil nilai rata-rata sebagai nilai untuk mengganti nilai yang kosong tersebut. Sebagai contoh jika ditemukan nilai *traffic* yang kosong pada suatu *cell* dalam suatu hari dengan periode pengukuran bulanan, maka nilai yang kosong tersebut digantikan dengan nilai rata-rata *traffic cell* tersebut di dalam periode pengukuran satu bulan. Jika dalam suatu *record* terlalu banyak ditemukan data yang hilang maka bisa saja diambil tindakan untuk menghapusnya

Contoh dari transformasi yang dilakukan dapat dilihat pada Tabel 6.1 berikut ini:

| Date | CellIdKey | Traffic | CSSR | CDR | SCR | HOSR | Call Drop |
|----------|-----------|---------|--------|-------|--------|--------|-----------|
| 1-Jan-07 | AL40811 | 85.3 | 99.53% | 0.96% | 98.58% | 98.47% | 30 |
| 1-Jan-07 | AL40812 | 85.7 | 99.60% | 0.80% | 98.81% | 98.50% | 25 |
| 1-Jan-07 | AL40813 | 83.1 | 99.48% | 0.91% | 98.57% | 96.54% | 26 |
| 1-Jan-07 | NK23011 | 39.37 | 96.05% | 1.77% | 94.35% | 94.48% | 61 |
| 1-Jan-07 | NK23012 | 44.59 | 98.49% | 0.79% | 97.72% | 95.43% | 35 |
| 1-Jan-07 | NK23013 | 111.8 | 99.50% | 0.28% | 99.22% | 98.06% | 34 |
| 2-Jan-07 | AL40811 | 71.6 | 99.64% | 0.48% | 99.16% | 97.17% | 14 |
| 2-Jan-07 | AL40812 | 83.4 | 99.69% | 0.73% | 98.96% | 98.38% | 18 |
| 2-Jan-07 | AL40813 | 74.2 | 99.46% | 0.97% | 98.49% | 97.97% | 25 |
| 2-Jan-07 | NK23011 | 22.35 | 95.60% | 1.75% | 93.93% | 94.31% | 36 |
| 2-Jan-07 | NK23012 | 25.72 | 98.78% | 0.43% | 98.35% | 89.16% | 12 |
| 2-Jan-07 | NK23013 | 80.72 | 98.84% | 0.39% | 98.45% | 97.35% | 35 |
| 3-Jan-07 | AL40811 | 81.1 | 99.64% | 0.77% | 98.87% | 97.84% | 21 |
| 3-Jan-07 | AL40812 | 83.3 | 99.55% | 0.81% | 98.74% | 97.69% | 23 |
| 3-Jan-07 | AL40813 | 74.8 | 99.55% | 0.94% | 98.61% | 96.72% | 23 |
| 3-Jan-07 | NK23011 | 21.26 | 96.50% | 1.75% | 94.81% | 96.41% | 28 |
| 3-Jan-07 | NK23012 | 22.87 | 98.60% | 0.61% | 98.00% | 94.95% | 15 |
| 3-Jan-07 | NK23013 | 80.13 | 98.57% | 0.47% | 98.11% | 97.73% | 36 |
| | | | | | | | ... |

Tabel 6.1 – Tabel Awal

Yang ditampilkan di atas adalah data set yang dipakai di dalam pengolahan *data mining* setelah dilakukan proses seleksi data pada Gambar 6.1. Data set ini tidak bisa langsung diterapkan pada algoritma *data mining (time series analysis)* melainkan perlu di transformasi terlebih dahulu disesuaikan dengan algoritma data mining yang digunakan. Hasil dari transformasi dapat dilihat pada Tabel 6.2

| CellIdKey | Traffic (t-2) | Traffic (t-1) | Traffic (t-0) | CSSR (t-2) | CSSR (t-1) | CSSR (t-0) | CDR (t-2) | CDR (t-1) | CDR (t-0) | SCR (t-2) | SCR (t-1) | SCR (t-0) | HOSR (t-2) | HOSR (t-1) | HOSR (t-1) | Call Drop | Call Drop | Call Drop |
|-----------|---------------|---------------|---------------|------------|------------|------------|-----------|-----------|-----------|-----------|-----------|-----------|------------|------------|------------|-----------|-----------|-----------|
| AL40811 | 85.3 | 71.6 | 81.1 | 99.53% | 99.64% | 99.64% | 0.96% | 0.48% | 0.77% | 98.58% | 99.16% | 98.87% | 98.47% | 97.17% | 97.84% | 30 | 14 | 21 |
| AL40812 | 85.7 | 83.4 | 83.3 | 99.60% | 99.69% | 99.55% | 0.80% | 0.73% | 0.81% | 98.81% | 98.96% | 98.74% | 98.50% | 98.38% | 97.69% | 25 | 18 | 23 |
| AL40813 | 83.1 | 74.2 | 74.8 | 99.48% | 99.46% | 99.55% | 0.91% | 0.97% | 0.94% | 98.57% | 98.49% | 98.61% | 96.54% | 97.97% | 96.72% | 26 | 25 | 23 |
| NK23011 | 39.37 | 22.35 | 21.26 | 96.05% | 95.60% | 96.50% | 1.77% | 1.75% | 1.75% | 94.35% | 93.93% | 94.81% | 94.48% | 94.31% | 96.41% | 61 | 36 | 28 |
| NK23012 | 44.59 | 25.72 | 22.87 | 98.49% | 98.78% | 98.60% | 0.79% | 0.43% | 0.61% | 97.72% | 98.35% | 98.00% | 95.43% | 89.16% | 94.95% | 35 | 12 | 15 |
| NK23013 | 111.8 | 80.72 | 80.13 | 99.50% | 98.84% | 98.57% | 0.28% | 0.39% | 0.47% | 99.22% | 98.45% | 98.11% | 98.06% | 97.35% | 97.73% | 34 | 35 | 36 |
| | | | | | | | | | | | | | | | | | | |

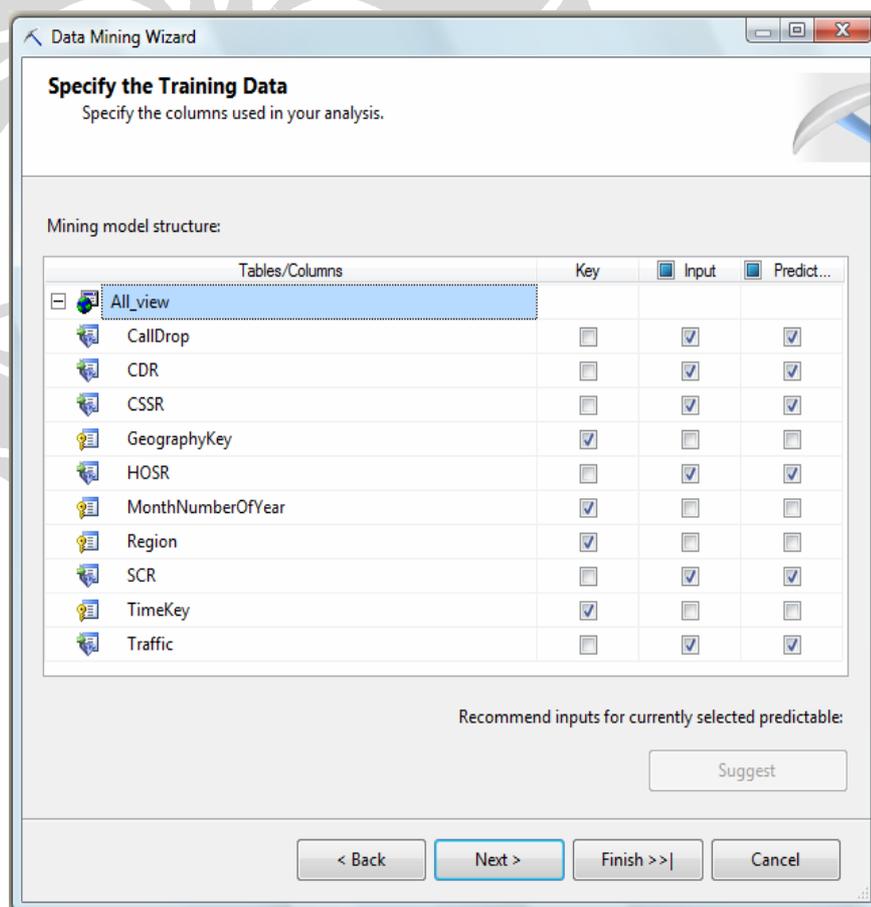
Tabel 6.2 – Tabel Hasil Transformasi



✚ Data Mining

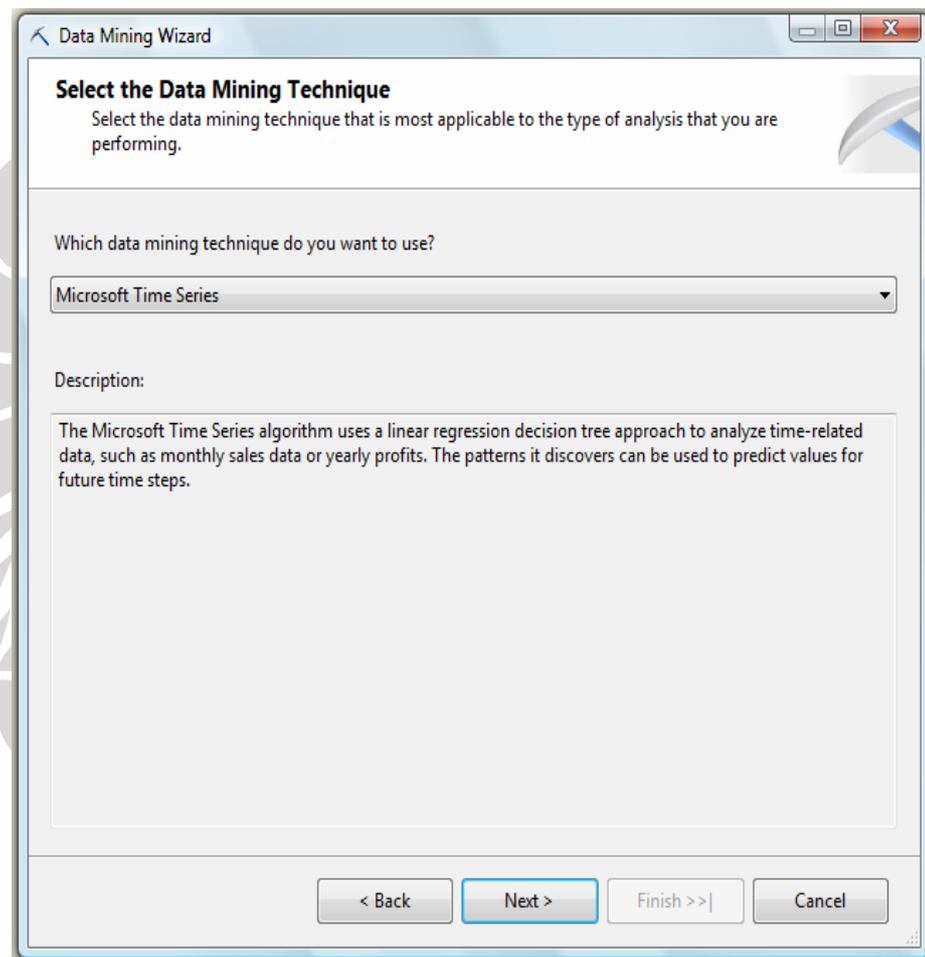
Teknik *data mining* yang digunakan dalam penelitian ini adalah *microsoft time series*. Teknik ini memiliki tiga tahap *data mining* yang implementasinya akan diterapkan pada suatu kasus yaitu: melakukan *time series analysis* terhadap data *traffic* harian untuk memprediksi total *traffic* per *region* periode bulanan untuk tahun berikutnya:

1. Tahap pertama adalah membuat *mining model*. Pada bagian ini ditentukan jumlah kolom yang akan digunakan sebagai key, input, kolom yang di prediksi dan algoritma yang digunakan. Berikut ini pada Gambar 6.2 dapat dilihat *training data* yang akan digunakan:



Gambar 6.2 – Training Data

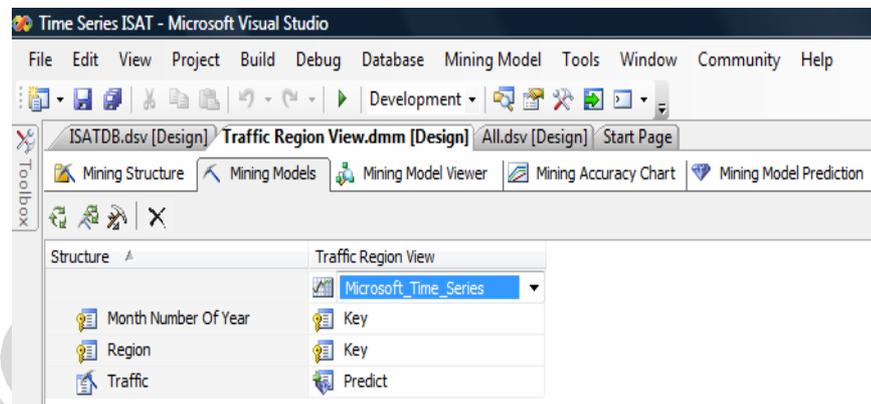
Yang dijadikan *key* pada *training data* ini adalah *geographyKey*, *timeKey* dan *region*. Input yang digunakan adalah *Traffic*, *CSSR*, *CDR*, *SCR*, *HOSR* dan *Call Drop*. Sedangkan yang akan di prediksi sama dengan Input yang digunakan. Langkah selanjutnya dilakukan pemilihan teknik *data mining* yang sesuai dengan analisis yang akan dilakukan, dalam hal ini *microsoft time series* (Gambar 6.3).



Gambar 6.3 – *Select the Data Mining Technique*

2. Tahap yang kedua dari *data mining* adalah *model training* (disebut juga pemrosesan). Pada tahap ini kita memberikan data *historical* ke *data mining engine* yaitu data elemen jaringan selama satu tahun. Dalam tahap *training*, algoritma *microsoft time series*

memulai untuk menganalisa masukan data dan melakukan proses *scanning* set data yang digunakan untuk menemukan korelasi antar nilai atribut yang telah di definisikan (Gambat 6.4)



Gambar 6.4 – *Mining Models*

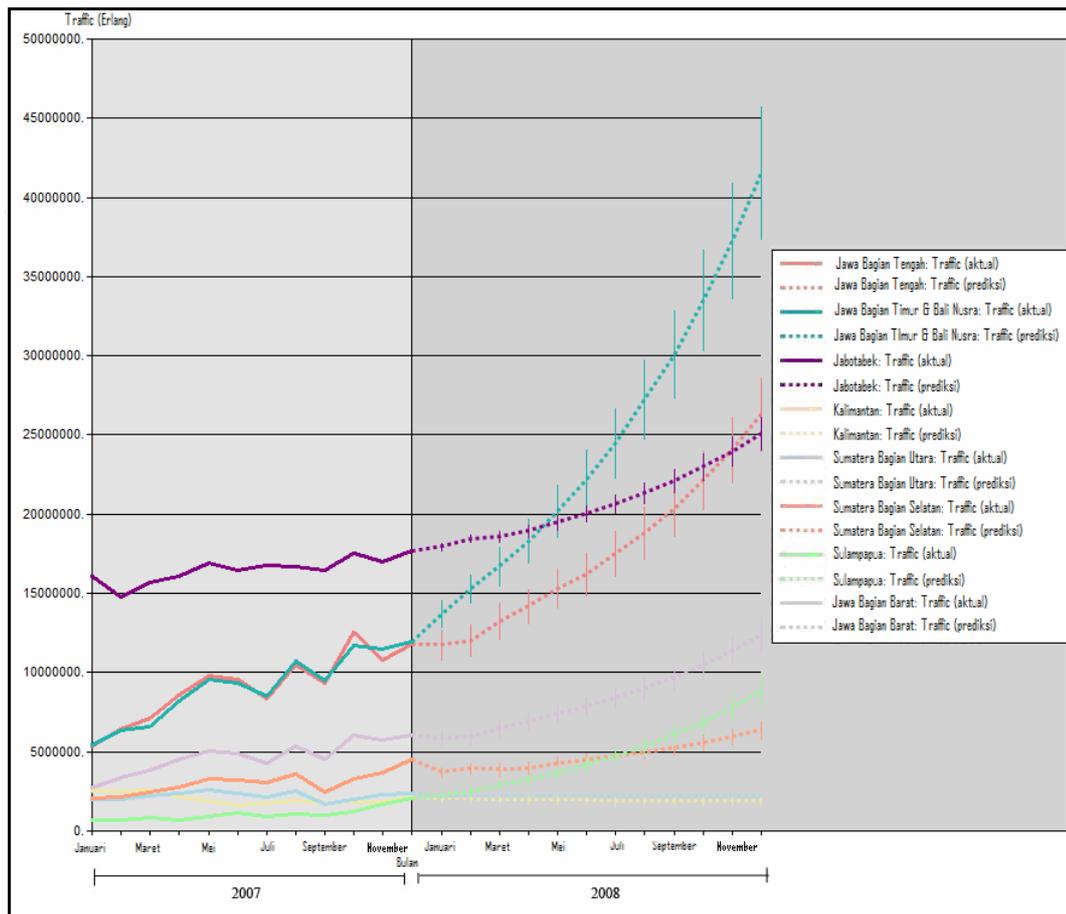
3. Tahap ketiga dari *data mining* adalah *prediction*. Dalam rangka untuk memprediksi nilai *traffic* per bulan tahun berikutnya, diperlukan *trained mining model* dan data set yang baru. Ketika prediksi, *data mining engine* menerapkan aturan yang ditemukan dari tahap *training* ke data set yang baru dan menempatkan hasil prediksi untuk masing-masing kasus inputan. Yang dimaksud kasus inputan disini adalah nilai *traffic* itu sendiri yang akan dijadikan sebagai bahan untuk memprediksi nilai total *traffic* sesuai dengan periode yang ditentukan per masing-masing regional. Pada tahap ini *trained mining model* yang diperoleh dari tahap kedua diterapkan ke data set yang baru untuk mendapatkan prediksi untuk masing-masing total *traffic* per regional. Untuk lebih jelasnya dapat dilihat pada Gambar 6.5 *predicted data*.

✚ Evaluasi Pola

Hasil dari proses data mining yang dilakukan pada tahap sebelumnya kemudian dilakukan proses pengidentifikasian terhadap pola-pola yang ditemukan. Hasil dari data mining ini akan menghasilkan banyak pola, tetapi hanya sebagian kecil dari pola-pola tersebut yang akan menarik perhatian peneliti yaitu pola yang valid (hasilnya sesuai dengan data yang digunakan), secara potensial berguna dan pola tersebut merupakan sesuatu hal yang baru (belum ditemukan sebelumnya). Dalam penelitian ini pola yang digunakan harus memberikan nilai yang valid untuk keperluan *time series analysis* seperti nilai *traffic* yang dihasilkan pada pola baru tidak boleh memiliki yang tidak wajar (contoh perbedaan nilai terlalu besar) dan secara potensial pola yang digunakan harus mendukung keperluan *time series analysis*.

✚ Presentasi

Hasil dari tahap evaluasi pola *data mining* harus ditampilkan menggunakan visualisasi yang baik, sehingga informasi yang ingin diperoleh dapat diterima dengan baik dan merepresentasikan suatu informasi tersembunyi yang sesuai dengan pengetahuan yang ingin digali melalui teknik *data mining*. Contoh tampilan presentasi dari *time series analysis* dapat dilihat pada Gambar 6.5 - *predicted data* dibawah ini:

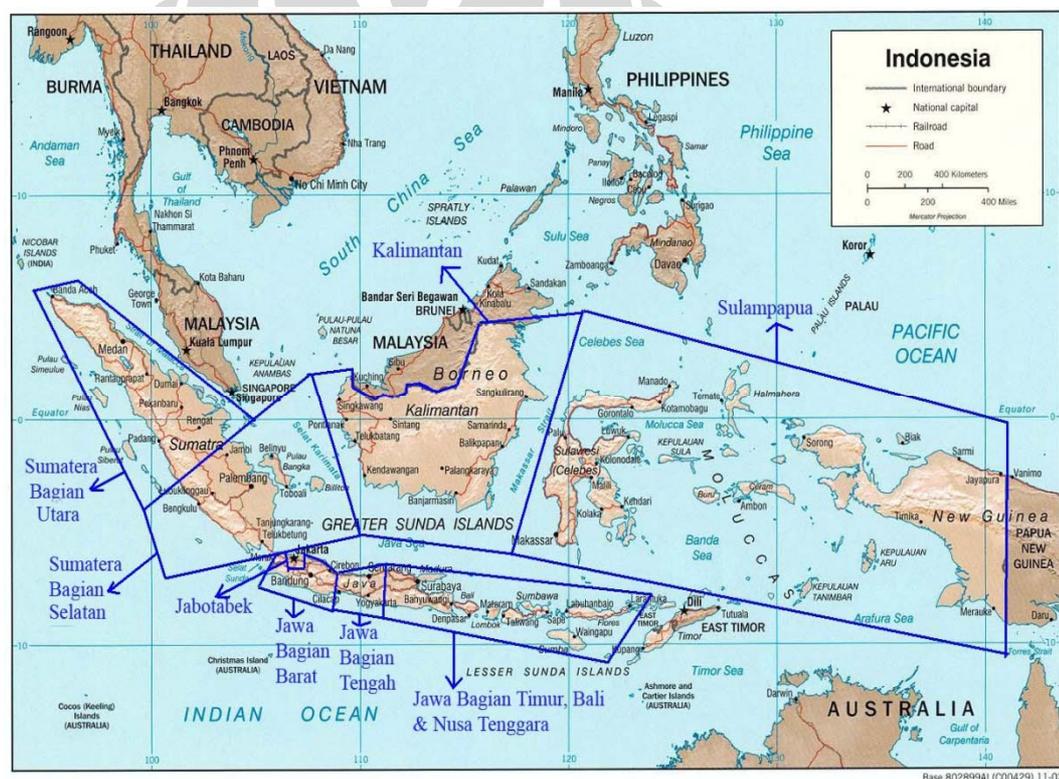


Gambar 6.5 – Actual Data (2007) & Predicted Data (2008)

Yang ditampilkan diatas adalah contoh hasil *time series analysis* dari penelitian ini menggunakan *Microsoft SQL Server 2005*. Dari data *traffic* harian yang dimiliki disimpulkan menjadi data *traffic* bulanan kemudian dilakukan *time series analysis* untuk memprediksi total *traffic* per *region* periode bulanan untuk tahun berikutnya. Grafik garis yang berada pada Januari - Desember 2007 pada sumbu x adalah data *traffic* per bulan pada tahun 2007 sedangkan yang berada pada Januari - Desember 2008 (latar belakang abu-abu) pada sumbu x adalah data *traffic* yang merupakan hasil prediksi *traffic* untuk tahun 2008. Sedangkan garis tegak yang ada diantara garis yang diprediksi merupakan garis deviasi dari masing-masing prediksi per bulan. Kegunaan dari *time series analysis* terhadap data *traffic* ini

adalah PT. Indosat Tbk dapat mengetahui prediksi, melihat tren, pola yang berulang, pola yang sama dan deviasi data *traffic* pada bulan yang akan datang sesuai pembagian regional jaringan PT. Indosat Tbk. Untuk memperjelas pemahaman mengenai pembagian regional jaringan PT. Indosat Tbk pada wilayah Indonesia, pada Gambar 6.6 dibawah ini dapat dilihat peta Indonesia yang telah diberi garis batas untuk membagi wilayah-wilayah Indonesia menjadi regional-regional besar.

🚦 Pembagian Regional Jaringan PT. Indosat Tbk



Gambar 6.6 – Pembagian Regional Jaringan PT. Indosat Tbk

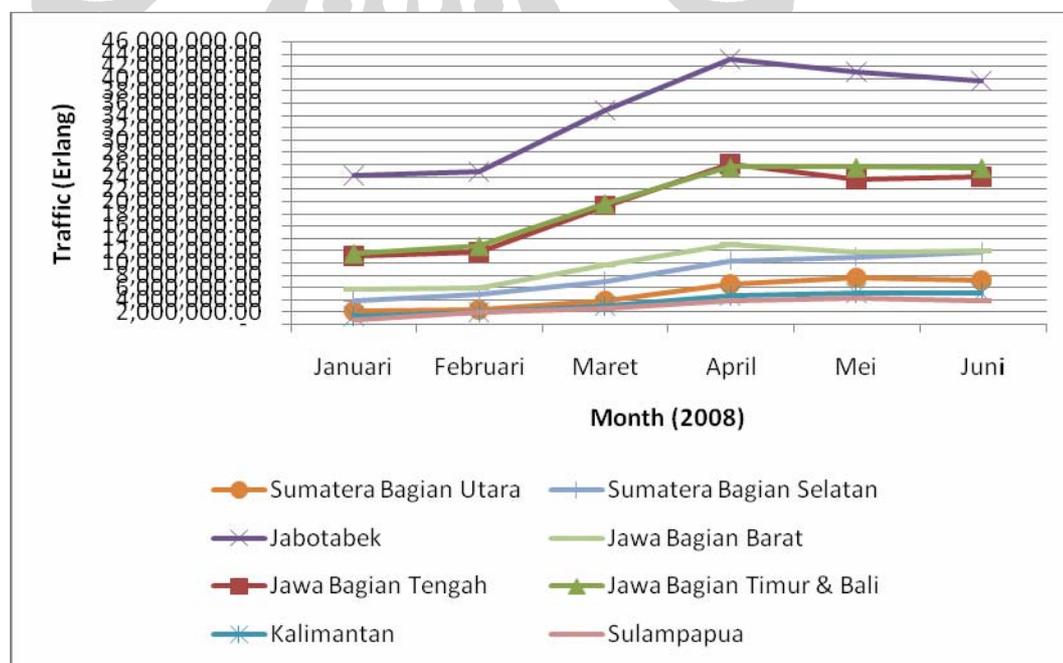
Penjelasan pembagian regional-regional tersebut diantaranya sebagai berikut:

- Sumatera Bagian Utara: Medan, Batam, Pekanbaru, Padang
- Sumatera Bagian Selatan: Palembang, Jambi, Lampung
- Jabotabek: Jakarta, Bogor, Tangerang, Bekasi

- Jawa Bagian Barat: Bandung, Cirebon, Sukabumi, Tasikmalaya, Purwakarta
- Jawa Bagian Tengah: Semarang, Yogyakarta, Solo, Purwokerto, Tegal
- Jawa Bagian Timur, Bali dan Nusa Tenggara: Surabaya, Malang, Madiun
Jember, Denpasar
- Kalimantan: Balikpapan, Samarinda, Pontianak, Sampit, Banjarmasin
- Sulampapua: Makassar, Manado, Palu, kendari, Papua

✚ Data Aktual

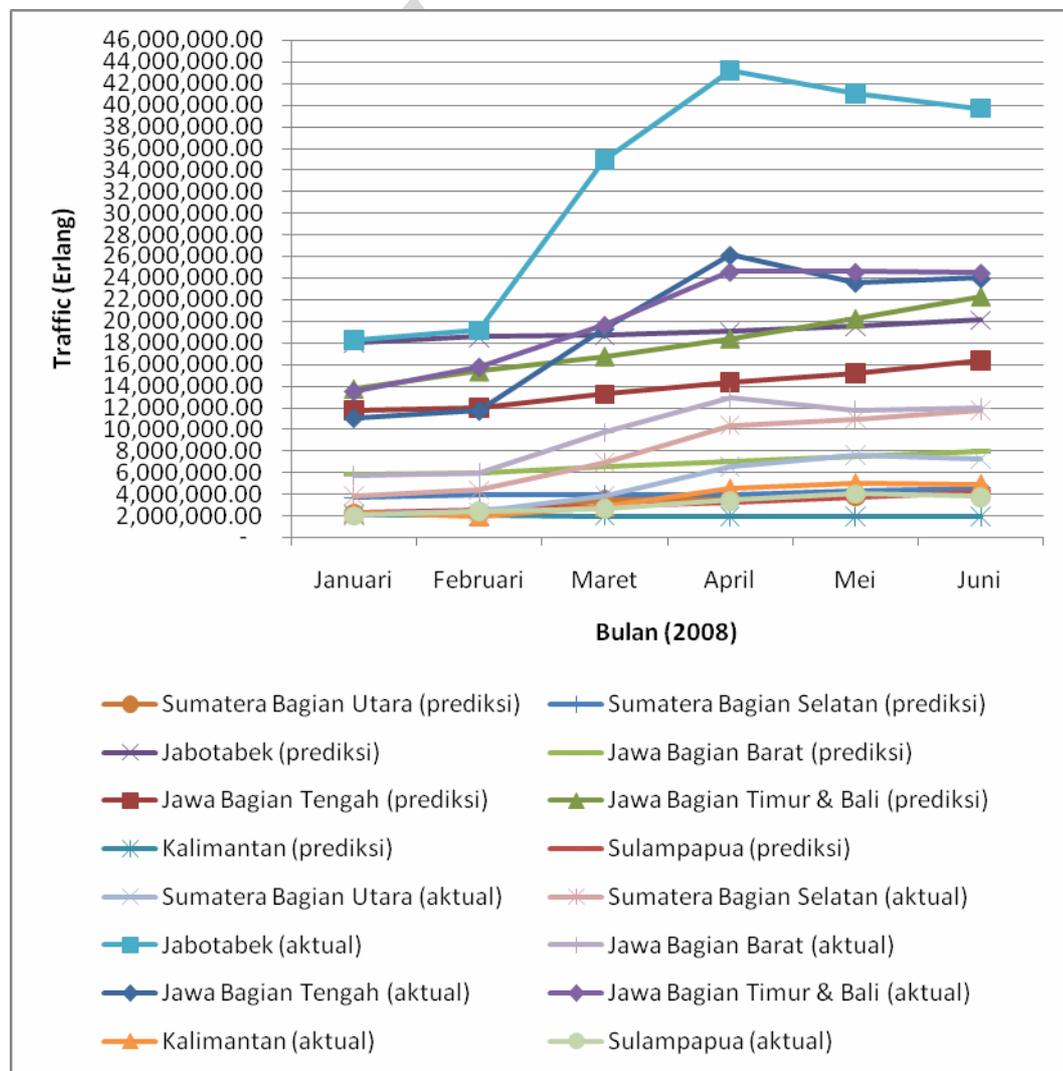
Untuk membandingkan prediksi data *traffic* dengan data aktual *traffic* diperlukan sumber data yang menggambarkan keadaan data aktual *traffic* dari periode yang berjalan. Periode waktu yang dijadikan acuan sebagai pembandingan adalah periode Januari sampai dengan Juni 2008. Berikut gambar data aktual *traffic* untuk periode tersebut:



Gambar 6.7 – Data Aktual Januari – Juni 2008

✚ Contoh perbandingan data hasil prediksi dengan data aktual

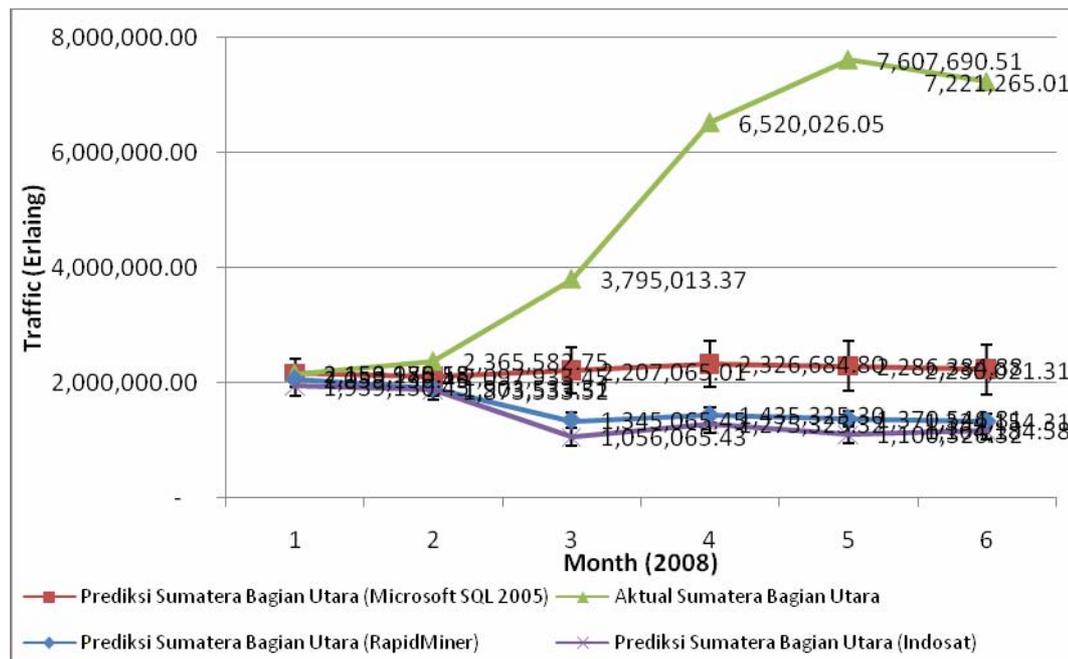
Prediksi yang dihasilkan melalui *time series analysis* kemudian dibandingkan dengan data *traffic* sesungguhnya (aktual) sesuai masing-masing regional pada periode bulan Januari 2008 sampai dengan Juni 2008. Contoh perbandingan data hasil prediksi (*Microsoft Time Series*) dengan data aktual dapat dilihat pada Gambar 6.8 berikut ini:



Gambar 6.8 – Perbandingan Prediksi Data dengan Data Aktual

Untuk melihat lebih detail perbandingan prediksi data *traffic* (Januari-Juni 2008) dengan menggunakan *tools* Microsoft SQL 2005, *tools* RapidMiner dan prediksi Indosat dengan data aktual yang ada dari masing-masing regional dapat dilihat pada Gambar 6.9 – Gambar 6.17 dibawah ini:

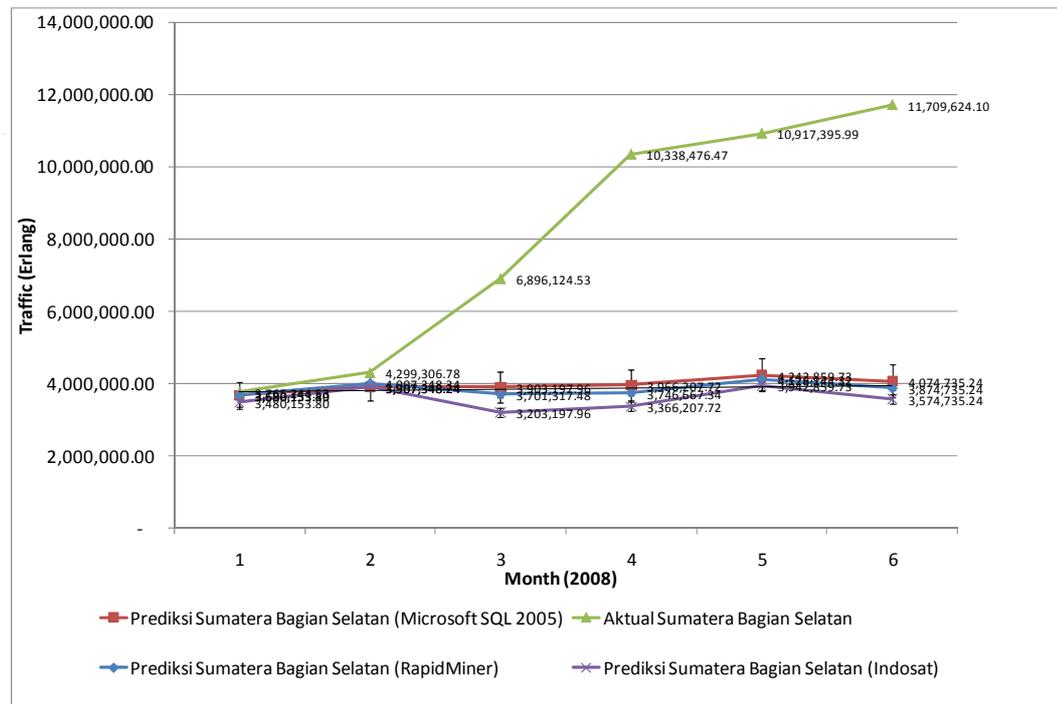
Sumatera Bagian Utara



Gambar 6.9 – Perbandingan Prediksi Data dengan Data Aktual
(Sumatera Bagian Utara)

Hasil dari perbandingan prediksi data *traffic* terhadap data aktual pada regional Sumatera Bagian Utara menggunakan *tools* Microsoft SQL 2005 dan RapidMiner menghasilkan pola prediksi data *traffic* yang mendekati data aktual *traffic* (berada di rentang nilai deviasi prediksi data *traffic*) untuk bulan Januari dan Februari, hal ini sama dengan nilai prediksi yang dilakukan Indosat. *Tools* Microsoft SQL 2005 dan RapidMiner memprediksi nilai dengan lebih akurat dibandingkan prediksi yang dilakukan oleh Indosat.

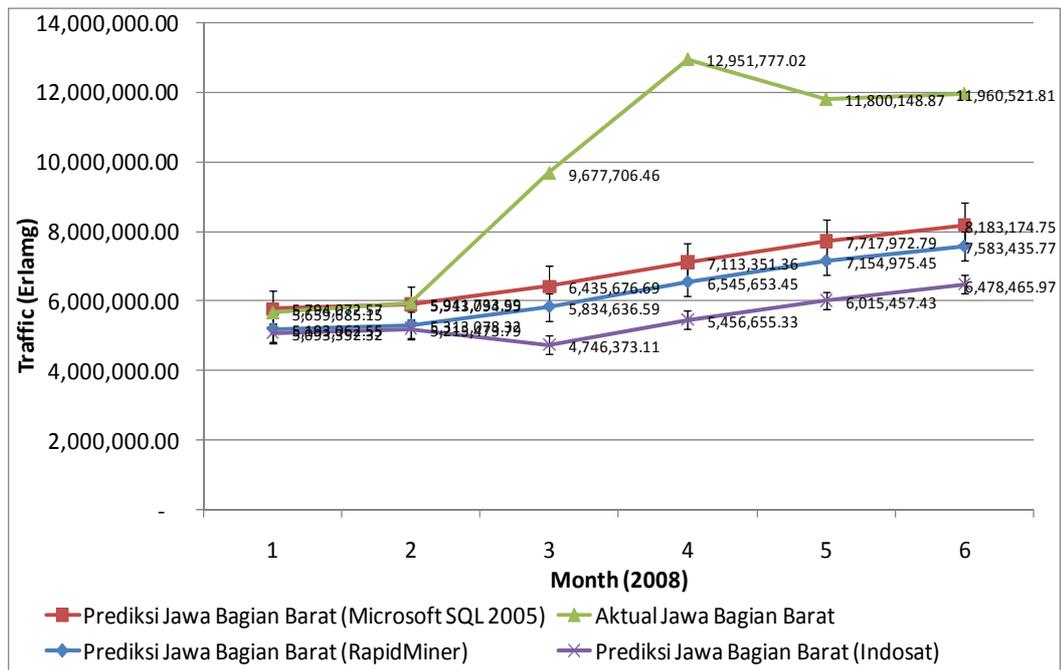
Sumatera Bagian Selatan



Gambar 6.10 – Perbandingan Prediksi Data dengan Data Aktual
(Sumatera Bagian Selatan)

Dari hasil perbandingan prediksi data *traffic* terhadap data aktual pada regional Sumatera Bagian Selatan diperoleh nilai prediksi *traffic* yang kesemuanya (*Microsoft SQL 2005*, *RapidMiner* dan *Indosat*) berada dibawah dari data aktual. Untuk bulan Januari dan Februari menghasilkan pola prediksi data *traffic* yang mendekati data aktual *traffic* untuk kesemuanya. Sedangkan bulan Maret sampai dengan Juni memiliki perbedaan nilai yang besar dibandingkan dengan dua bulan sebelumnya dimana prediksi *tools Microsoft SQL 2005* dan *RapidMiner* lebih akurat dibandingkan prediksi *Indosat* . Perbedaan terbesar yang terjadi pada data *traffic* pada bulan Juni dimana nilai yang terjadi mencapai 2 ½ kali nilai *traffic* yang diprediksi sehingga data aktual berada jauh dari rentang deviasi prediksi data.

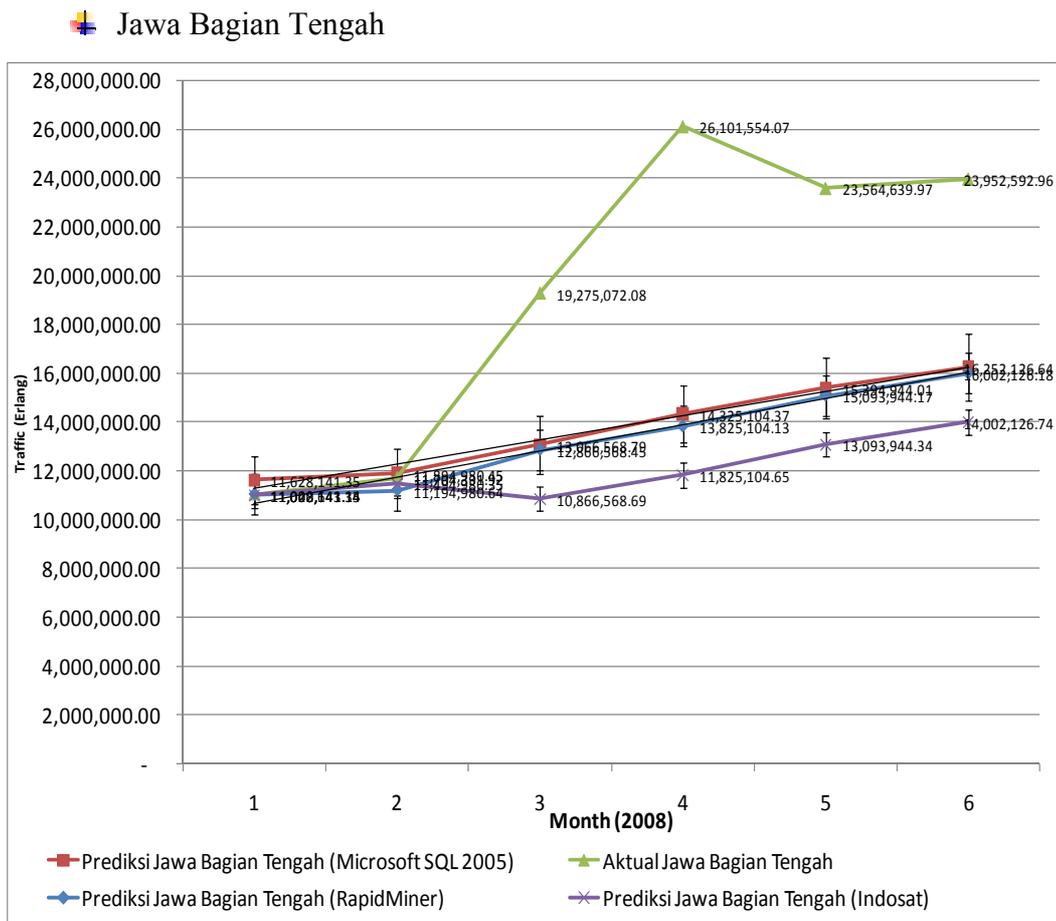
✚ Jawa Bagian Barat



Gambar 6.11 – Perbandingan Prediksi Data dengan Data Aktual

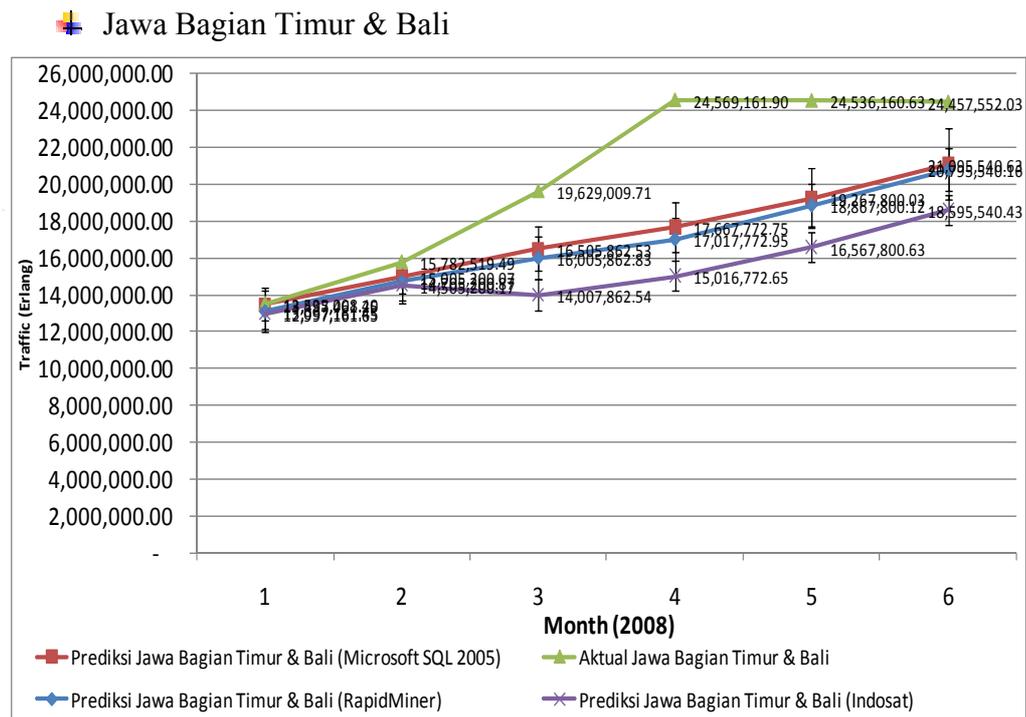
(Jawa Bagian Barat)

Perbandingan prediksi data *traffic* terhadap data aktual pada regional Jawa Bagian Barat memiliki ketepatan prediksi untuk bulan Januari dan Februari yang berada di rentang nilai deviasi prediksi data *traffic* untuk *tools* Microsoft SQL 2005, *tools* RapidMiner dan prediksi Indosat. Kedua bulan tersebut memiliki nilai perbedaan yang lebih kecil dibanding dengan nilai perbedaan untuk Bulan Maret – Juni sehingga nilai aktual semakin mendekati prediksi data. Perbedaan terbesar yang terjadi pada data *traffic* pada bulan April dimana perbedaan nilai yang terjadi sebesar 5.927.817,12 erlang. Prediksi *tools* Microsoft SQL 2005 dan *tools* RapidMiner terlihat lebih akurat dibandingkan prediksi Indosat



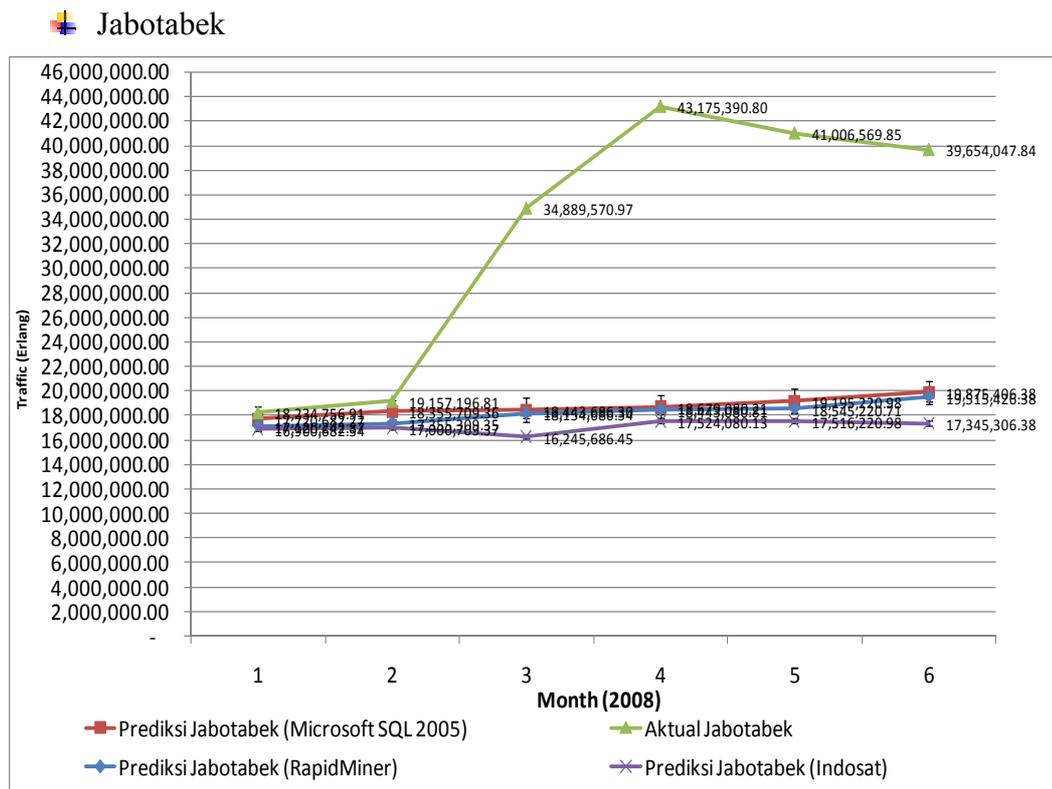
Gambar 6.12 – Perbandingan Prediksi Data dengan Data Aktual
(Jawa Bagian Tengah)

Perbandingan prediksi data *traffic* terhadap data aktual pada regional Jawa Bagian Tengah memiliki ketepatan prediksi untuk bulan Januari dan Februari untuk *tools* Microsoft SQL 2005, *tools* RapidMiner dan prediksi Indosat. Hasil prediksi data untuk kedua bulan tersebut memiliki nilai prediksi yang melebihi data aktual. Tetapi meskipun kedua prediksi data tersebut tidak tepat sama dengan data aktual kedua bulan dapat dikatakan memiliki prediksi yang tepat. Hal ini dikarenakan kedua bulan tersebut memiliki nilai deviasi yang berada pada rentang deviasi dari prediksi data. Hal ini jauh berbeda dengan bulan Maret - Juni yang memiliki perbedaan yang besar antara prediksi data dengan ktual data. Perbedaan terbesar terjadi pada bulan April dengan nilai sebesar 11.781.571,28 erlang.



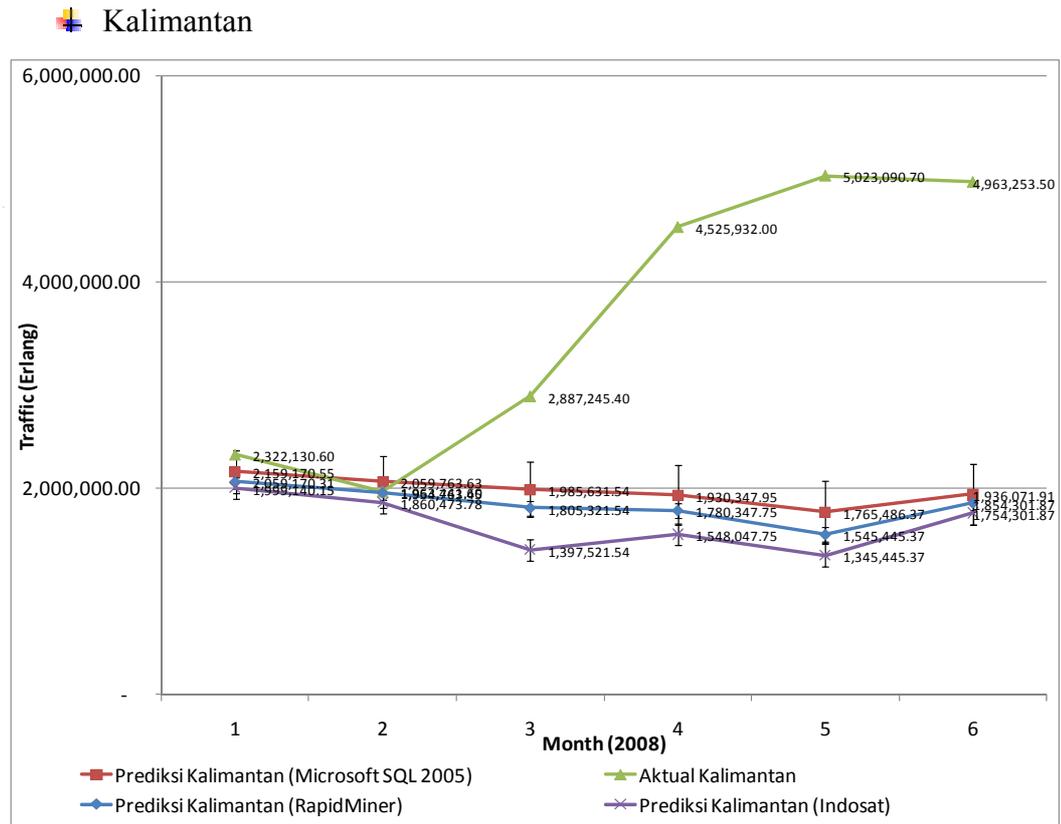
Gambar 6.13 – Perbandingan Prediksi Data dengan Data Aktual
(Jawa Bagian Timur, Bali & Nusa Tenggara)

Untuk regional Jawa Bagian Timur, Bali & Nusa Tenggara terdapat 2 bulan dimana data aktual berada pada rentang deviasi dari prediksi data yaitu bulan Januari dan Februari untuk *tools* Microsoft SQL 2005, *tools* RapidMiner dan prediksi Indosat. Dari kedua bulan tersebut perbedaan prediksi dengan aktual data terendah terdapat pada bulan Januari yaitu sekitar -1.5% dari prediksi data. Untuk bulan Maret sampai dengan bulan Mei memiliki nilai perbandingan yang cukup besar. Hal ini juga terjadi pada regional Jawa lainnya dimana pada bulan-bulan tersebut perbandingan antara prediksi data dengan data aktual memiliki perbedaan yang signifikan. Prediksi dengan menggunakan *tools* Microsoft SQL 2005 dan *RapidMiner* lebih baik dibandingkan prediksi Indosat, karena prediksi Indosat tidak melakukan pergantian nilai yang kosong pada data dengan suatu nilai yang ditentukan seperti contoh digantikan dengan nilai rata-rata *traffic*.



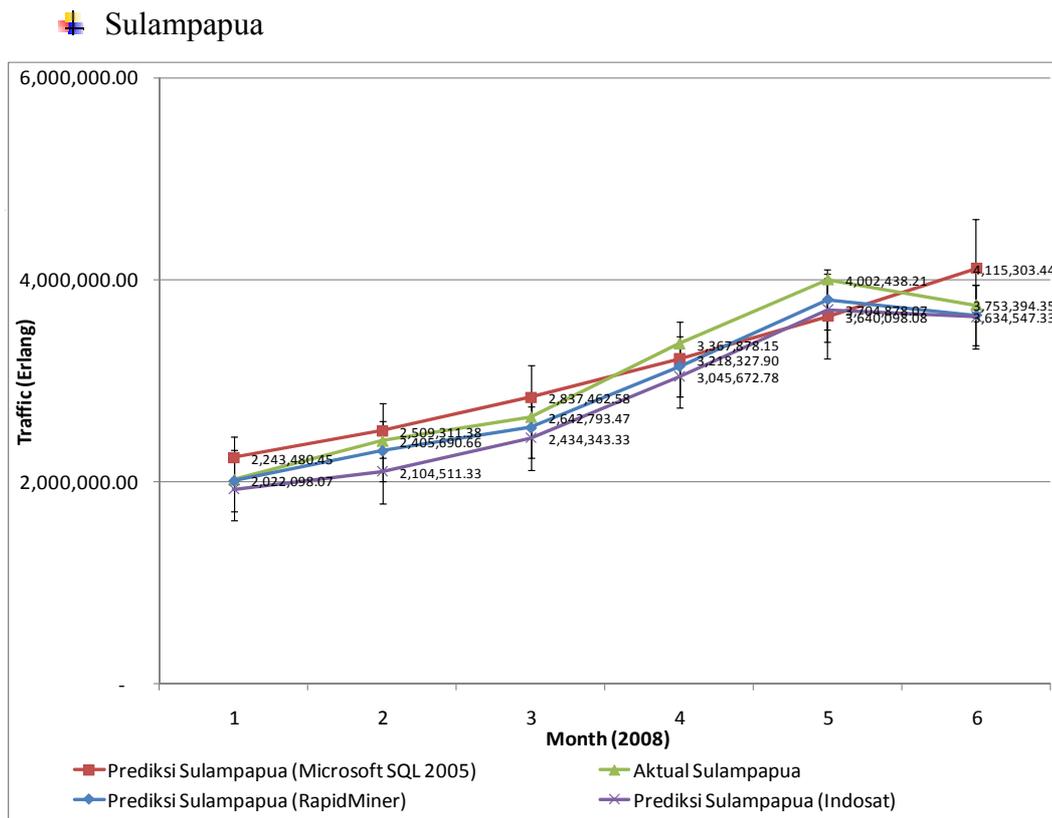
Gambar 6.14 – Perbandingan Prediksi Data dengan Data Aktual
(Jabotabek)

Regional Jabotabek dibandingkan dengan regional lainnya memiliki nilai aktual *traffic* yang sangat besar hingga mencapai 43 juta erlang pada bulan April. Data *traffic* aktual ini jauh berbeda dengan prediksi data yang diperkirakan pada bulan April jumlah *traffic* yang akan terjadi berkisar 18,6 Juta erlang. Jumlah nilai prediksi data ini sebenarnya cukup wajar terjadi melihat pola *traffic* yang terjadi di regional jabotabek pada bulan dan tahun sebelumnya (Gambar 6.5) kenaikan dan penurunan nilai *traffic* cukup stabil. Tetapi pada kenyataannya data *traffic* pada tahun 2008 melonjak tinggi secara drastis sejak akhir bulan Februari – April baru kemudian data *traffic* turun perlahan-lahan. Pada regional Jabotabek prediksi data *traffic* yang mendekati data aktual hanya terjadi pada bulan Januari dan Februari untuk *tools* Microsoft SQL 2005, *tools* RapidMiner dan prediksi Indosat.



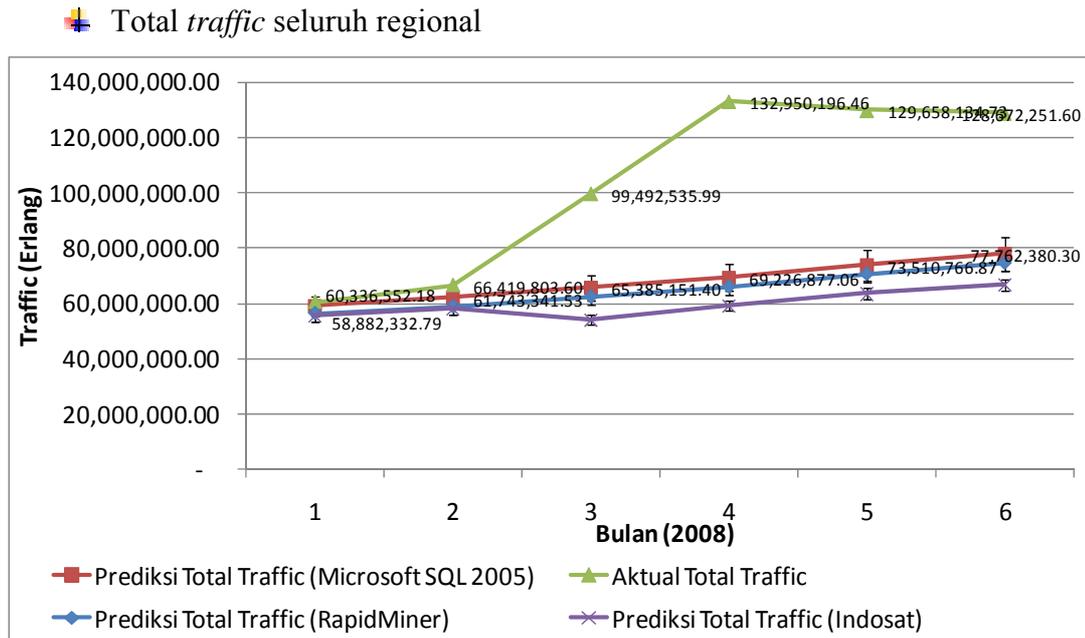
Gambar 6.15 – Perbandingan Prediksi Data dengan Data Aktual
(Kalimantan)

Regional Kalimantan prediksi data yang mendekati akurat terjadi pada bulan Januari dan Februari untuk *tools* Microsoft SQL 2005, *tools* RapidMiner dan prediksi Indosat. Kedua bulan ini memiliki selisih antara prediksi data dengan data aktual yang lebih kecil dibandingkan dengan bulan-bulan berikutnya. Prediksi data yang paling akurat dari prediksi data yang dilakukan untuk bulan Januari – Juni terjadi pada bulan Februari untuk *tools* Microsoft SQL 2005, RapidMiner dan Indosat. Perbedaan antara prediksi data dengan data aktual yang ada sebesar 47,563.44 erlang sehingga semakin dekat dengan data aktual. Berbeda dengan bulan lainnya yang jauh berbeda nilai prediksinya dengan data aktual untuk *tools* Microsoft SQL 2005, *tools* RapidMiner dan prediksi Indosat.



Gambar 6.16 – Perbandingan Prediksi Data dengan Data Aktual
(Sulampapua)

Untuk regional sulampapua, pola data aktual *traffic* 2008 terlihat kenaikan dan penurunan *traffic* per bulan yang relatif stabil. Seluruh prediksi data yang dilakukan pada regional ini memiliki tingkat akurasi yang cukup baik untuk *tools Microsoft SQL 2005, tools RapidMiner* dan prediksi Indosat. Hal ini terlihat dari masing-masing bulan data aktual memiliki perbedaan yang kecil dan yang terpenting data aktual pada masing-masing bulan berada pada rentang deviasi dari prediksi data sehingga prediksi data dapat dikatakan terpenuhi. Dari perbandingan prediksi data dengan data aktual yang dilakukan sesuai masing-masing regional yang ada, regional sulampapua memberikan hasil yang paling baik, dimana tingkat keakurasian prediksi data dari masing-masing bulan cenderung tinggi.



Gambar 6.17 – Perbandingan Prediksi Data Total *traffic* dengan Data Aktual Total *Traffic*

Perbandingan prediksi data total *traffic* dengan data aktual total *traffic* memperlihatkan pola dimana pada bulan Januari dan Februari memiliki prediksi data total *traffic* yang hampir mendekati data aktual total *traffic* untuk tools *Microsoft SQL 2005*, tools *RapidMiner* dan prediksi Indosat, dengan nilai total *traffic* aktual yang berada pada rentang deviasi prediksi data total *traffic*. Kedua bulan ini memiliki perbedaan data yang terendah jika dibandingkan dengan bulan Maret sampai dengan Juni. Perbedaan dengan selisih terbesar terjadi pada bulan April dimana perbedaan total *traffic* aktual dengan total mencapai 61,389,439.06 erlang. Hasil prediksi data *traffic* yang dapat dilihat pada Gambar 6.9 sampai dengan Gambar 6.16 menggambarkan perbandingan antara prediksi data *traffic* dengan data aktual *traffic* untuk regional Sumatera Bagian Utara, Sumatera Bagian Selatan, Jawa Bagian Barat, Jawa Bagian Tengah, Jawa Bagian Timur (termasuk Bali dan Nusa Tenggara), Jabotabek, Kalimantan dan Sulampapua. Ke

delapan regional tersebut memiliki pola prediksi data *traffic* yang mendekati data aktual *traffic* (berada di rentang nilai deviasi prediksi data *traffic*) untuk bulan Januari dan Februari untuk *tools Microsoft SQL 2005*, *tools RapidMiner* dan prediksi Indosat. Sedangkan untuk bulan Maret - Juni prediksi data *traffic* untuk ke delapan regional memiliki perbedaan yang cukup signifikan kecuali untuk regional sulampapua. Hasil prediksi data yang mendekati data aktual atau sebaliknya memiliki perbedaan yang cukup signifikan sangat dipengaruhi oleh faktor eksternal. Pada kenyataannya data aktual *traffic* pada penelitian ini dipengaruhi oleh promo yang dilakukan oleh pihak Marketing. Prediksi data *traffic* data mendekati aktual *traffic* pada bulan Januari dan Februari 2008 untuk *tools Microsoft SQL 2005*, *tools RapidMiner* dan prediksi Indosat lebih disebabkan oleh tidak adanya promo marketing yang terjadi pada bulan-bulan tersebut, hal ini berbeda dengan bulan-bulan berikutnya. Pada bulan Maret-Juni pihak marketing melakukan promo besar-besaran seperti paket promo IM3 Rp40,- sepuasnya, Mentari *Freetalk* (00.00-05.00), IM3 Rp 0,00..1 dan lainnya sehingga terjadi *traffic* yang tinggi pada bulan-bulan tersebut menyebabkan prediksi data *traffic* memiliki perbedaan yang besar dengan data aktual *traffic*. Untuk regional sulampapua, pola data aktual *traffic* 2008 (Gambar 6.16) terlihat kenaikan dan penurunan *traffic* per bulan yang relatif stabil. Hal ini menggambarkan pada regional tersebut nilai *traffic* tidak terlalu terpengaruh oleh promo yang dilakukan oleh pihak marketing, sehingga prediksi data *traffic* pada regional ini cukup akurat. Dari ketiga prediksi (*tools Microsoft SQL 2005*, *tools RapidMiner* dan Indosat) diperoleh hasil prediksi yang lebih baik untuk *tools Microsoft SQL 2005* dan *RapidMiner* dibandingkan prediksi yang dilakukan oleh Indosat.

✚ Presentasi hasil prediksi yang ditemukan untuk menghasilkan aksi

Dengan mengetahui hasil prediksi nilai *traffic* per regional (Gambar 6.5) maka akan memberikan beberapa manfaat bagi perusahaan antara lain:

1. Melakukan perencanaan strategis investasi untuk mengatasi jumlah lonjakan *traffic* yang akan terjadi pada bulan yang akan datang seperti melakukan upgrade atau penambahan BTS baru dan elemen jaringan lainnya agar dapat mengatasi manajemen *load traffic* secara optimal
2. Dengan melihat informasi *traffic* serta pola yang dibentuk dari masing-masing *traffic* per regional dapat dimanfaatkan untuk melakukan *resource planning* dan *operational maintenance*.
3. Membantu pihak manajemen untuk menentukan apakah keputusan yang diambil akan memberikan keuntungan atau kerugian. Contoh ketika diprediksi akan terjadi lonjakan *traffic* pada bulan desember tetapi pada bulan selanjutnya diprediksi akan mengalami penurunan maka pihak perencanaan dapat mengambil keputusan apakah akan menambah kapasitas atau mengubah konfigurasi elemen jaringan yang ada untuk mengatasi permasalahan tersebut
4. Membantu pihak marketing dalam menentukan saat yang tepat untuk melakukan promo dengan cara melihat pola *traffic* dari masing-masing regional dalam suatu waktu. Contoh ketika *traffic* di prediksi rendah januari di regional kalimantan maka pihak marketing bisa melakukan promo di regional kalimantan pada bulan tersebut.

Semua manfaat tersebut dapat diperoleh dengan lebih baik menggunakan *tools Microsoft SQL 2005* atau pun *RapidMiner* dibandingkan *tools* prediksi Indosat.