



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB II

LANDASAN TEORI

2.1 Data *Outlier*

Outlier biasa juga disebut dengan data yang tidak biasa (*anomaly*), sekumpulan data yang telah di deteksi terjadi penyimpangan atau data tersebut menyimpang jauh dari data yang ada dalam rangkaian data. Definisi kedua dikemukakan (Chandola, 2007) bahwa Sebuah *outlier* adalah pengamatan yang menyimpang jauh dari pengamatan lain untuk membangkitkan kecurigaan bahwa itu dihasilkan oleh mekanisme yang berbeda. Data *outlier* harus dikaji terlebih dahulu apa penyebab dan pengaruhnya sehingga dapat diputuskan apakah data tersebut harus dipertahankan atau dikeluarkan.

Deteksi *outlier* bisa juga disebut dengan *anomaly detection*, *event detection*, *novelty detection*, deteksi kesalahan dan penyalahgunaan. Teknik untuk mendeteksi data *outlier* bisa dengan beberapa cara yaitu klasifikasi, *clustering*, visualisasi, *statistic*, teori informasi, dekomposisi *spectral*, *signal processing*.

Menurut (Budiarto, Tarno, Warsito, 2013) ada 2 jenis *outlier* yang akan muncul pada data *time series*, yaitu *additive outlier* dan *innovation outlier*. *Additive outlier* adalah *outlier* yang hanya mempengaruhi satu observasi saja, yaitu observasi yang terjadi pada waktu terjadinya *outlier*, misalnya dikarenakan ada kesalahan pencatatan dan kesalahan input ke data komputer. Sedangkan jenis *Innovation outlier*, tidak hanya observasi tertentu atau satu observasi saja tetapi

akan berpengaruh pada observasi-observasi selanjutnya. Biasanya jenis ini terjadi karena adanya gangguan yang tidak biasa atau disebut *unusual innovation outlier*, seperti ketidakstabilan politik, pemogokan, krisis moneter, bencana alam, perang, dan lain-lain.

2.2 Data Mining

Data mining adalah proses mengelola data dalam jumlah yang besar, data yang sudah ada dari tahun ke tahun, biasanya *data mining* terjadi karena data yang sudah menumpuk.

Proses ini bertujuan untuk mendapatkan informasi yang diharapkan dapat menjadi nilai tambah untuk organisasi. Dalam proses menggali informasi dalam data mining terdapat enam fungsi menurut (Kusrini, Taufiq, 2009) yaitu:

1) Fungsi Deskripsi

Untuk memberikan gambaran secara keseluruhan dan ringkas, gambaran tersebut diambil dari kumpulan data yang besar serta jenis-jenis yang berbeda. Metode yang digunakan antara lain, *Neural Network*, *Decision Tree*, *Exploratory Data Analysis*.

2) Fungsi Estimasi

Dapat menebak sebuah nilai yang sebelumnya tidak diketahui. Metode yang dapat digunakan *Point Estimation dan Confidence Interval Estimations*, *Simple Linear Regression dan Correlation*, dan *Multiple Regression*.

3) Fungsi Prediksi

Dapat memprediksi nilai di masa datang, seperti memprediksi stok *material* di masa yang akan datang. Termasuk dalam fungsi ini metode *Neural Network, Decision Tree, dan k-Nearest Neighbor*.

4) Fungsi Klasifikasi

Berfungsi menemukan model atau fungsi yang membedakan serta menjelaskan berbagai konsep atau kelas data. Bertujuan untuk menemukan jika terdapat suatu objek dalam kelas yang Label nya tidak diketahui. Metode yang dapat digunakan, yaitu *Neural Network, Decision Tree, k-Nearest Neighbor, dan Naive Bayes*.

5) Fungsi Pengelompokan (*Clustering*)

Pada fungsi ini data yang memiliki karakteristik yang berbeda akan dikelompokkan dan diidentifikasi. Dapat menggunakan metode *Hierarchical Clustering, metode K-Means, dan Self Organizing Map (SOM)*.

6) Fungsi Asosiasi

Mengidentifikasi kombinasi yang terjadi antar *item* pada waktu yang sama. Metode atau algoritma dalam fungsi ini adalah *Apriori, Generalized Sequential Pattern (GSP), FP-Growth dan GRI algorithm*.

2.3 Data Warehouse

Definisi *data warehouse* oleh (Inmon, 2015) yaitu kumpulan data yang berorientasi-subjek, terintegrasi, *nonvolatile*, dan varian-waktu untuk mendukung

keputusan manajerial. Dengan kata lain data warehouse tidak hanya data yang ada di masa sekarang tetapi sekumpulan data yang besar berasal dari beberapa sumber yang berbeda, disimpan dalam periode yang panjang, diakumulasikan dari tahun ke tahun yang jumlahnya besar dan tersimpan dalam *database* operasional. Penyimpanan *data warehouse* ini berfungsi untuk mendukung pengambilan keputusan. Data-data tersebut dapat di akses oleh alat-alat analisis seperti OLAP, visualisasi data, sistem pendukung keputusan, *data mining* dan aplikasi lain yang mengatur proses pengumpulan data dan mengirimkan ke *business user*.

2.4 Visualisasi Data

Visualisasi data merupakan komunikasi visual *modern* dengan menggunakan grafik informasi, seperti tabel atau grafik. Bertujuan mengkomunikasikan informasi secara jelas kepada pengguna dan membantu pengguna dalam menganalisa data yang kompleks.

Menurut Friedman (2008) tujuan utama dari visualisasi data adalah untuk mengkomunikasikan informasi secara jelas dan efektif dengan cara grafis. Di samping itu visualisasi data berguna untuk memaparkan ide secara efektif, bentuk estetik dan fungsionalitas. Perancangan visualisasi harus mencapai keseimbangan antara bentuk dan fungsi sehingga dapat menciptakan visualisasi data yang menawan dan mencapai tujuan utamanya yaitu mengkomunikasikan informasi.

2.5 Business Intelligence

Business intelligence (BI) merupakan suatu konsep dan metode untuk meningkatkan kualitas pengambilan keputusan bisnis yang berbasis data,

dengan kata lain BI ini dapat membantu sebuah organisasi dengan mengelola data menjadi informasi (Powers, 2007). Konsep BI berfokus pada 5 dayanya:

1. *Data Sourcing*

Data sourcing dimana dalam konsep BI ini mampu untuk mengakses berbagai data dan berbagai informasi dari berbagai sumber meskipun data atau informasi tersimpan dalam format penyimpanan apa saja, misalnya *excel*, *csv*, dan lain-lain.

2. *Data Analysis*

Dalam hal ini BI (*Business Intelligence*) memiliki kemampuan dalam menganalisis data dengan jumlah yang banyak sekalipun, data dan informasi yang didapatkan dari aktivitas organisasi ini dapat di analisis dan nantinya dapat dijadikan pengetahuan untuk meningkatkan proses bisnis suatu organisasi.

3. *Situation Awareness*

Situation awareness maksudnya adalah dimana BI sebagai sebuah jalan keluar yang dapat menyediakan sebuah sistem untuk dapat digunakan dalam keadaan yang darurat terdesak. Misalnya dalam mengambil atau mencari sebuah informasi yang dibutuhkan oleh organisasi.

4. *Risk Analysis*

Memberikan perhitungan untuk risiko yang akan dihadapi organisasi, BI dapat memperhitungkan akibat-akibat yang akan terjadi dari setiap pilihan yang diambil perusahaan.

5. *Decision support*

Memberikan pertimbangan-pertimbangan sebelum mengambil keputusan. Informasi dapat didapat dari internal dan eksternal organisasi berdasarkan berbagai macam perhitungan dan pengolahan data.

2.6 *Extraction Transformation Loading (ETL)*

Proses mempersiapkan data yang dari sumbernya untuk masuk ke *data warehouse*, secara ringkas ETL bertujuan untuk mengumpulkan, menyaring dan mengelola data, di sisi lain ETL juga berfungsi untuk mengintegrasikan data. Di bawah ini tahapan ETL:

a. *Extrac*

Mengekstrak atau melakukan pembersihan terhadap data yang telah diambil melalui sumbernya. Agar mendapatkan pola struktur data yang diinginkan.

b. *Transformation*

Memilih kolom tertentu untuk dimasukkan ke dalam data warehouse, menerjemahkan nilai berupa kode, mengkodekan nilai-nilai, melakukan perhitungan nilai-nilai baru, menggabungkan data dari berbagai sumber dan membuat ringkasannya.

c. *Load*

Memasukkan data yang sudah siap ke dalam data warehouse, berkontribusi dalam tampilan dan kualitas data dari proses ETL.

2.7 Online Analytical Processing

Sebuah metode yang dapat memproses, mengoleksi, menyimpan dan memanipulasi data dalam database dengan struktur multidimensi. Bertujuan untuk menganalisis data dengan cepat. Menurut Scheps (2008) *Online Analytical Processing* (OLAP) adalah sebuah konsep data multidimensional dengan konsep sentralisasi data transaksional perusahaan. Bukan hanya mengagregasi data, OLAP memberikan kemampuan pada sistem BI (*Business Intelligence*) untuk melihat data dengan cara baru. Menurut Turban, Sharda, Delen, dan King struktur operasional utama dalam OLAP didasarkan pada konsep yang disebut kubus (*cube*). Kubus (*cube*) di dalam OLAP adalah struktur data multidimensional (*actual* atau *virtual*) yang memungkinkan analisis data yang cepat. di bawah ini adalah tiga operasi dasar dalam OLAP:

1) Konsolidasi (*roll-up*)

Data yang disajikan pada operasi ini bersifat global, misalnya penjualan dari suatu perusahaan besar dikelompokkan berdasarkan penjualannya di berbagai negara.

2) *Drill-down*

Drill-down merupakan kebalikan dari *roll-up*, ialah sebuah bentuk data global yang ditampilkan lebih rinci dan *detail*.

3) *Slicing and Dicing*

Dikenal dengan istilah *pivoting*, kemampuannya adalah menampilkan data dengan berbagai sudut pandang atau dengan kata lain pengguna dapat

melihat data dari berbagai perspektif. Operasi ini juga digunakan untuk melihat data sebagai visualisasi dari kubus.

2.8 Tableau

Tableau *software* merupakan *tools business intelligence* yang membantu dalam proses visualisasi data. Pengguna dapat melakukan koneksi data secara mudah dan data yang diproses ke dalam tableau bisa dalam berbagai macam format (Tableau, t.thn). Secara garis besar kelebihan tableau:

a. *Fast Analytics*

Menghubungkan dan memvisualisasikan data dalam hitungan menit.

b. *Ease of Use*

Kemudahan dalam penggunaannya, pengguna dapat menganalisis data dengan teknik drag & drop, tidak perlu menggunakan pemrograman yang kompleks.

c. *Big Data, Any Data*

Dapat digunakan untuk data yang berkapasitas besar maupun data dengan jenis yang berbeda.

d. *Smart Dashboards*

Menggabungkan beberapa tampilan data agar informasi yang didapat lebih banyak.

e. *Update Automatically*

Data yang dikoneksikan langsung dapat di-*update* secara otomatis sesuai dengan jadwal yang ditentukan.

f. *Share in Seconds*

Dashboard yang telah selesai dapat langsung dipublikasikan lewat *web* dan perangkat *mobile*.

2.9 Wawancara

Dalam pengumpulan data teknik *interview* adalah metode yang menghubungkan pencari informasi dengan narasumber secara langsung. Pencari informasi yang biasa disebut pewawancara akan bertatap muka dengan narasumber. Menurut (Sugiono, 2009) wawancara adalah proses memperoleh informasi (tanya jawab) untuk tujuan penelitian, dengan cara bertatap muka antara peneliti dan narasumber. Bentuk-bentuk wawancara antara lain:

1. Wawancara berita diambil untuk keperluan bahan berita.
2. Wawancara yang disiapkan lebih dulu.
3. Wawancara telepon yaitu wawancara yang dilakukan lewat telepon.
4. Wawancara pribadi.
5. Wawancara dengan banyak orang.
6. Wawancara dadakan atau mendesak.
7. Wawancara kelompok, dimana seorang wartawan mewawancarai seorang pejabat, seniman, dan sebagainya.