



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. Visualisasi Data

Visualisasi data adalah sebuah teknik untuk memindahkan data atau informasi ke dalam bentuk visual. (Murray, 2013). Penggunaan visualisasi data dapat memperjelas data atau informasi yang kita miliki agar dapat dikomunikasikan kepada orang lain. (Few & EDGE, 2007)

Tujuan dari visualisasi data adalah mendapatkan *insight* atau gambaran dengan medium gambar/grafis yang interaktif, ke dalam berbagai aspek yang berhubungan dengan sebagian proses yang kita minati. (Telea, 2014)

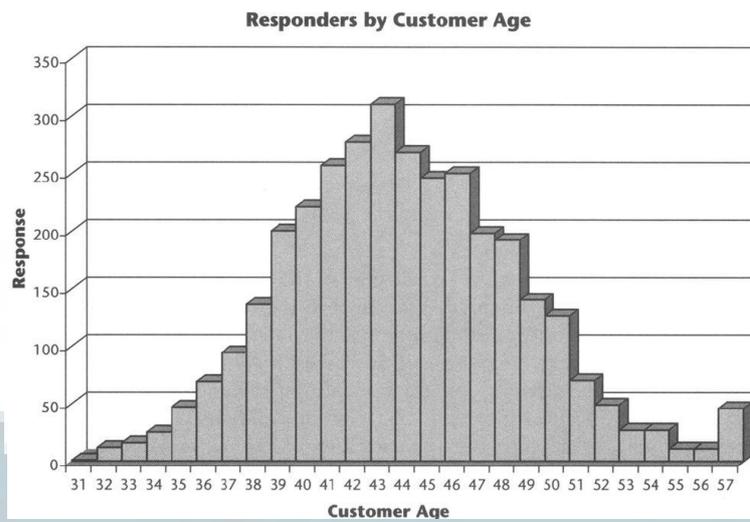
Terdapat dua jenis *tool* data visualisasi yang dapat digunakan yaitu:

- a. ***Multidimensional visualization tools: Scatter, line, column, bar, pie chart.***
- b. ***Specialized landscape and hierarchical visualization tools: map, tree graphs.***

Dalam melakukan analisa dan evaluasi set data, ada beberapa contoh visualisasi yang biasa digunakan untuk menjawab pertanyaan-pertanyaan bisnis, antara lain:

- a. ***Frequency Graphs***

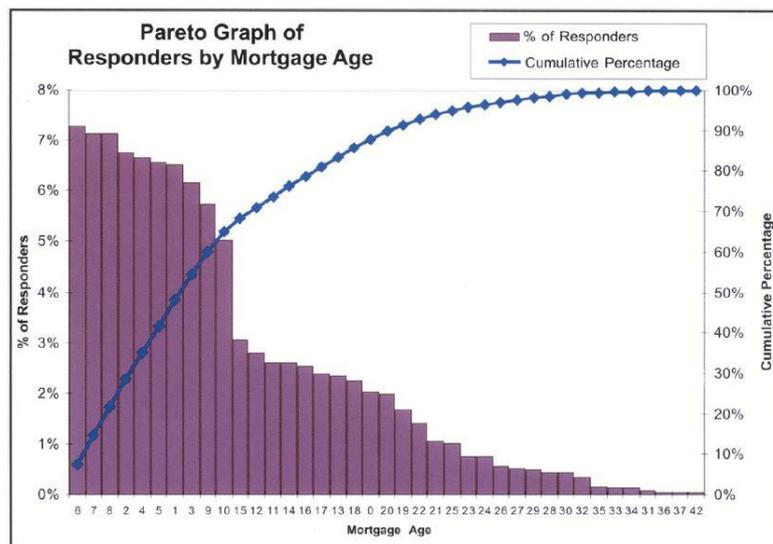
Frequency graphs digunakan untuk mengevaluasi indikator bisnis utama seperti tingkat respon terhadap suatu program marketing dengan membandingkan mereka dengan kolom lain dalam data set.



Gambar 2.1 Frequency Graph

b. Pareto Graphs

Pareto graphs mengevaluasi pentingnya kolom dalam sebuah data set dengan merujuk kepada indikator bisnis seperti tingkat respon, jumlah kegagalan, atau klaim penipuan.

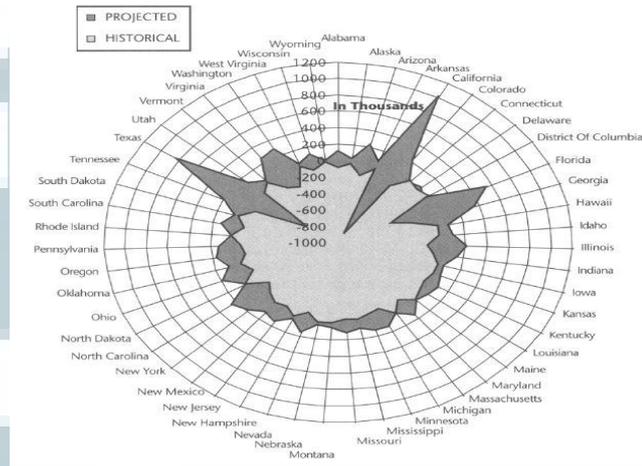


Gambar 2.2 Pareto Graph

c. **Radar Graphs**

Radar graphs mencari dan mengevaluasi tren musiman atau tren lain dalam set data secara keseluruhan.

Projected versus Historical Female Population Changes by State

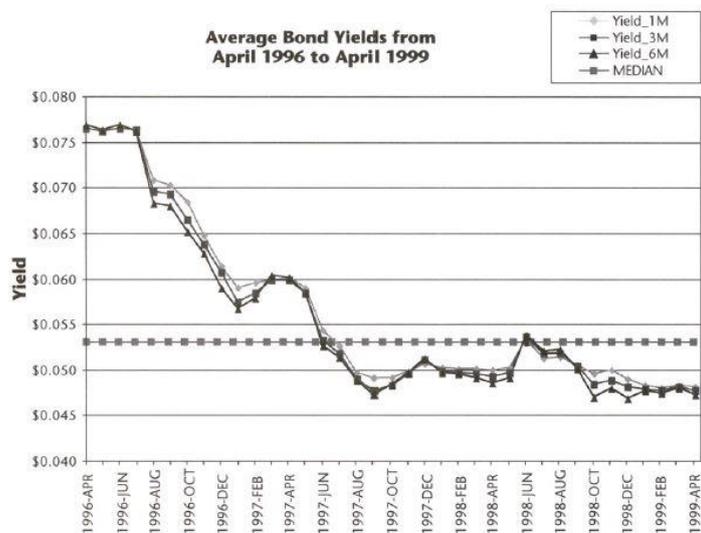


Gambar 2.3 Radar Graph

d. **Line Graphs**

Line graphs menganalisa dan mencari tren berdasarkan waktu (*time-based trends*) dan pola (*patterns*).

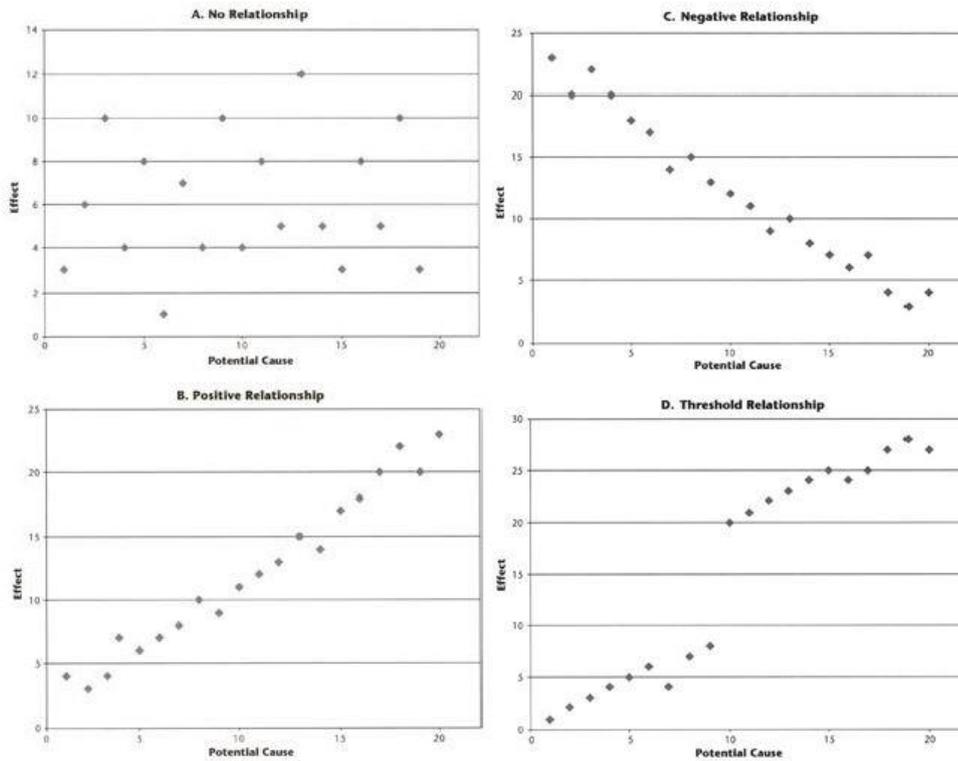
Average Bond Yields from April 1996 to April 1999



Gambar 2.4 Line Graph

e. Scatter Graphs

Scatter graphs berfungsi untuk mengevaluasi hubungan sebab-akibat (cause-and-effect relationships).



Gambar 2.5 Scatter Graph

2.2. Dashboard

Dashboard adalah tampilan visual dari informasi penting yang dibutuhkan untuk mencapai sebuah tujuan, tersusun dalam sebuah layar sehingga informasi tersebut dapat dilihat hanya dalam sekilas pandang. (Few, 2013)

2.3. *Database*

Database adalah kumpulan dari data yang berhubungan. *Database* dirancang, dibangun, dan diisi dengan data untuk tujuan yang spesifik untuk kumpulan pengguna yang spesifik pula. (Elmasri & Navathe, 2007)

Database adalah struktur komputer yang terbagi dan terintegrasi yang menyimpan kumpulan data dari pengguna (*end-user*) yang berisi fakta mentah (*raw facts*) dari *end-user* tersebut dan *metadata* yaitu data mengenai data, dimana data *end-user* terintegrasi dan disusun. (Rob & Coronel, 2009)

2.4. *Database Management System (DBMS)*

Database Management System (DBMS) adalah kumpulan program yang memungkinkan pengguna untuk membuat dan memelihara sebuah *database*. (Elmasri & Navathe, 2007)

Berikut adalah keuntungan penggunaan DBMS (Rob & Coronel, 2009):

- Meningkatkan pembagian data
- Meningkatkan keamanan data
- Integrasi data yang lebih baik
- Menguras inkonsistensi data
- Akses data yang lebih baik
- Pengambilan keputusan yang lebih baik
- Meningkatkan produktivitas pengguna

2.5. *Data Warehouse*

Data Warehouse adalah kumpulan data yang berorientasi subjek, terintegrasi, berubah-ubah setiap waktunya, dan *non-volatile* (data tidak berubah/diganti tetapi selalu diberikan penambahan). Tujuan dari *Data Warehouse* adalah untuk mengintegrasikan data korporat yang berukuran besar ke dalam sebuah penyimpanan dimana pengguna dapat dengan mudah menjalankan *query*, menghasilkan laporan, dan melakukan analisis. (Connolly & Begg, 2010)

Keuntungan menggunakan *Data Warehouse* adalah (Connolly & Begg, 2010):

- *Return of Investment* (ROI) yang tinggi.
- Keunggulan kompetitif (*Competitive advantages*).
- Meningkatkan produktivitas dari para pengambil keputusan.

2.6. *Data Mart*

Data Mart adalah bagian dari *Data Warehouse* yang mendukung kebutuhan dari departemen atau divisi bisnis tertentu. (Connolly & Begg, 2010)

2.7. *Data Mining*

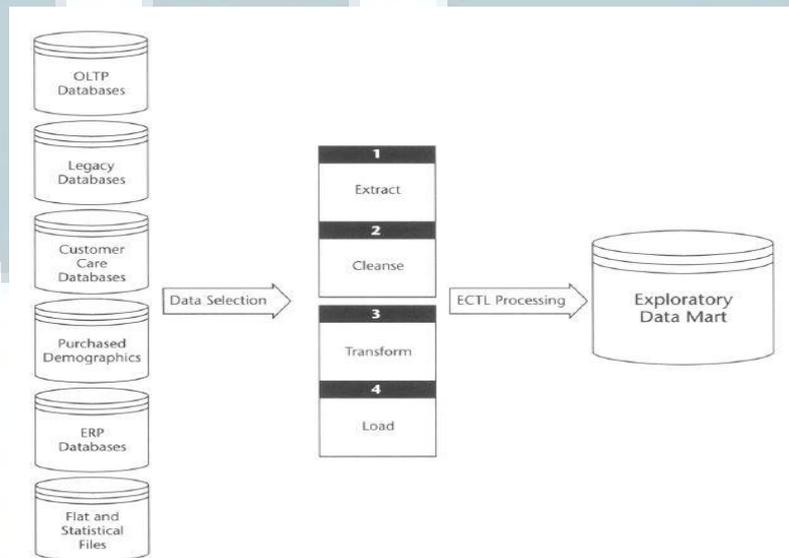
Data Mining adalah proses mengambil informasi yang valid dan dapat dipahami dari database yang besar dan menggunakannya untuk mengambil keputusan bisnis yang krusial. (Connolly & Begg, 2010)

2.8. *Data Cleaning*

Data cleaning, disebut juga *data cleansing* atau *data scrubbing*, adalah proses mendeteksi dan menghapus error dan inkonsistensi dari data yang bertujuan untuk meningkatkan kualitas data yang akan dimasukkan ke dalam *database* ataupun *data warehouse*. (Li, 2014)

2.9. *Extract, Transform, Load (ETL)*

Proses ETL (*Extract, Transform, Load*) adalah proses yang penting dalam pengembangan data warehouse.



Gambar 2.6 Proses *Extract, Cleanse, Transform, Load*

Sesuai dengan namanya, proses ini terbagi dalam 3 bagian yaitu:

a. *Extract*

Pada tahap ini dilakukan pengambilan data dari berbagai sumber yang bervariasi. Tahapan ini merupakan aspek yang cukup menantang karena diperlukan ketepatan dalam pengambilan data. Tujuan dari tahap ini adalah untuk mendapatkan data-data yang telah diambil dari berbagai sumber untuk

digabungkan menjadi satu format yang dapat digunakan untuk transformasi data.

(Ali, 2014)

b. Transform

Pada tahapan ini, dilakukan transformasi atau perubahan data dari format awal dari sumbernya menjadi format yang akan digunakan dalam *data warehouse* melalui beberapa cara yang serupa (Ali, 2014), yaitu:

- Mengambil beberapa kolom tertentu dalam data
- Menerjemahkan nilai yang tersandi
- Memberikan kode pada data yang bebas
- Mendapatkan perhitungan nilai baru
- Pensortiran
- Penggabungan data dari banyak sumber
- Membagi sebuah kolom menjadi beberapa kolom baru
- Agregasi
- Menerapkan beberapa validasi data yang sederhana ataupun kompleks

Pada *eight-step data visualization and data mining methodology*, terdapat dua jenis *logical transformations* yang diperlukan sebelum melakukan visualisasi data:

a. Table-Level Logical Transformations

Dalam jenis transformasi ini, perubahan pada struktur dan jumlah baris dalam keseluruhan set data. Perubahan-perubahannya antara lain:

1. ***Transforming Weighted Data Sets:*** Merubah kolom dari set data dan mengubahnya dengan baris tambahan yang menyerupai bobot aslinya.

2. ***Transforming Time Series Data Sets:*** Melakukan perubahan struktur pada *time series*. Banyak data visualisasi menggunakan *time series* berupa jumlah observasi pada waktu tertentu.

3. ***Aggregating Data Sets:*** Melakukan penggabungan data untuk melakukan eksperimen, misalnya dalam visualisasi peta, tingkat data yang diperlukan adalah provinsi. Maka data set harus digabungkan agar bisa dapat mencapai level provinsi.

4. ***Filtering Data Sets:*** Melakukan penyaringan data untuk melakukan pemilihan data yang lebih spesifik guna menelusuri aspek berbeda dari pertanyaan bisnis.

b. Column-Level Logical Transformations

Dalam jenis transformasi ini, Perubahan hanya terjadi pada struktur kolom dari keseluruhan set data. Perubahan-perubahannya antara lain:

1. Simple Column Transformations: Perubahan sederhana dalam kolom yaitu menghapus kolom, menambah kolom, dan merubah tipe data kolom.

2. Column Grouping Transformations: Membuat kolom baru berdasarkan *range criteria* yang dipilih pada kolom awal.

c. Load

Memuat data ke dalam tujuan akhir (biasanya adalah *data warehouse*). Data dapat dimuat dalam beberapa cara seperti memuat data sekaligus (*bulk loading*) melalui *software* DBMS ataupun melalui proses *insert data* dalam bentuk baris (*row*) secara bertahap (*sequence*). Biasanya *bulk loading* lebih disarankan karena memasukkan data secara bertahap (*sequence*) memakan waktu yang lama dalam proses *parsing* (penguraian) *statement* yang dimasukkan. (Anitha & Babu, 2014)

2.10. Indeks Kedalaman Kemiskinan

Indeks Kedalaman Kemiskinan (Poverty Gap Index-P1), merupakan ukuran rata-rata kesenjangan pengeluaran masing-masing penduduk miskin terhadap garis kemiskinan. Semakin tinggi nilai indeks, semakin jauh rata-rata pengeluaran penduduk dari garis kemiskinan. (Badan Pusat Statistik, 2017)

2.11. Indeks Keparahan Kemiskinan

Indeks Keparahan Kemiskinan (Poverty Severity Index-P2) memberikan gambaran mengenai penyebaran pengeluaran diantara penduduk miskin. Semakin tinggi nilai indeks, semakin tinggi ketimpangan pengeluaran diantara penduduk miskin. (Badan Pusat Statistik, 2017)

2.12. Garis Kemiskinan (GK)

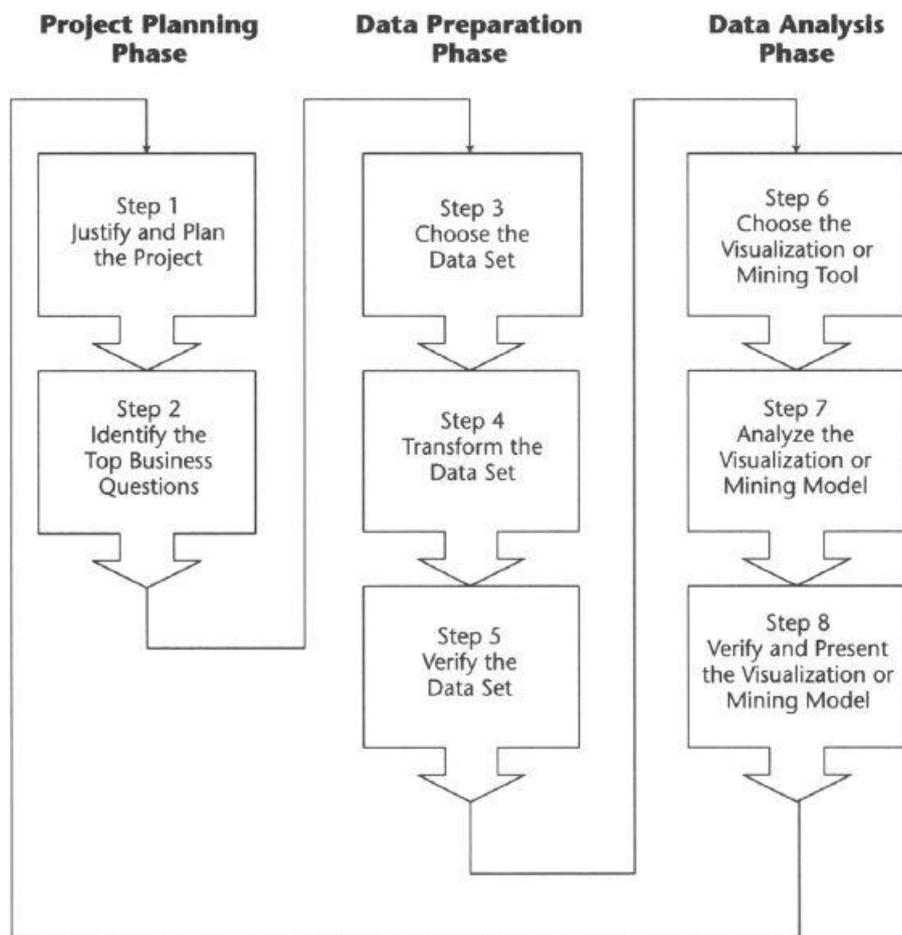
Garis Kemiskinan (GK) merupakan penjumlahan dari Garis Kemiskinan Makanan (GKM) dan Garis Kemiskinan Non Makanan (GKNM). Penduduk yang memiliki rata-rata pengeluaran per kapita per bulan di bawah Garis Kemiskinan dikategorikan sebagai penduduk miskin. (Badan Pusat Statistik, 2017)

Garis Kemiskinan Makanan (GKM) merupakan nilai pengeluaran kebutuhan minimum makanan yang disetarakan dengan 2100 kilokalori per kapita per hari. Paket komoditi kebutuhan dasar makanan diwakili oleh 52 jenis komoditi (padi-padian, umbi-umbian, ikan, daging, telur dan susu, sayuran, kacang-kacangan, buah-buahan, minyak dan lemak, dll). (Badan Pusat Statistik, 2017)

Garis Kemiskinan Non Makanan (GKNM) adalah kebutuhan minimum untuk perumahan, sandang, pendidikan dan kesehatan. Paket komoditi kebutuhan dasar non makanan diwakili oleh 51 jenis komoditi di perkotaan dan 47 jenis komoditi di pedesaan. (Badan Pusat Statistik, 2017)

2.13. *Eight-step Data Visualization and Data Mining Methodology*

Eight-step data visualization and data mining methodology adalah sebuah metode yang digunakan untuk melakukan sebuah proses *data mining* dan data visualisasi dalam sebuah proyek yang disebut *Visual Data Mining Project (VDM Project)*.



Gambar 2.7 Eight-step Data Visualization and Data Mining Methodology

Metode ini terbagi dalam 3 fase yaitu:

1. Project Planning Phase:

- a. Justify and Plan the Project*
- b. Identify the Top Business Questions*

2. Data Preparation Phase:

- a. Choose the Data Set*
- b. Transform the Data Set*
- c. Verify the Data Set*

3. Data Analysis Phase:

- a. Choose the Visualization or Mining Tool*
- b. Analyze the Visualization or Mining Model*
- c. Verify and Present the Visualization of Mining Model*

A. Project Planning Phase

Pada fase ini dilakukan proses perencanaan sebuah proyek visualisasi data, terdapat 2 tahap dalam fase perencanaan proyek, yaitu *justify and plan the project* (pembentukan tujuan dan perencanaan proyek) dan *identify the top business questions* (mengidentifikasi pertanyaan bisnis).

1. Justify and Plan the Project

Pada tahap ini dilakukan pembentukan tujuan yang ingin dicapai dalam sebuah proyek data visualisasi serta perencanaan proyek tersebut.

Hal yang pertama kali dilakukan dalam tahap ini adalah menentukan tipe proyek data visualisasi yang terdiri dari 3 jenis yaitu:

a. *Proof-of-concept*

Scope proyek ini adalah menentukan apakah visualisasi dan *data mining* akan bermanfaat bagi bisnis, membuktikan nilai guna dari visualisasi dan *data mining* kepada para pengambil keputusan, serta memberikan pengalaman dan konsep mengenai visualisasi dan data mining ke dalam organisasi.

b. *Pilot*

Scope proyek ini adalah untuk mengetahui, menganalisa, dan menjawab satu atau lebih pertanyaan bisnis untuk menentukan apakah ROI (Return of Investment) dari hasil penemuan atau jawaban tersebut menjamin proyek dapat berlanjut menjadi *Production project*.

c. *Production*

Scope proyek ini mirip dengan *pilot project*, yaitu menganalisa dan menjawab pertanyaan bisnis, akan tetapi hasil dari proyek tersebut kemudian langsung diimplementasikan ke dalam produksi.

Setelah menentukan jenis proyek, yang selanjutnya dilakukan adalah membuat *project justification*. *Project Justification* adalah tahap dimana kita menentukan target, dan juga tujuan dari data visualisasi dalam sebuah proyek.

Setelah *project justification*, yang selanjutnya dilakukan adalah melakukan perencanaan proyek dengan menentukan hal-hal seperti *timeline*, sumber daya yang digunakan dalam proyek, dan membentuk tim-tim untuk mengerjakan proyek.

2. Identify the Top Business Questions

Pada tahap ini dilakukan penentuan pertanyaan bisnis mana yang harus ditelusuri, kemudian memetakan pertanyaan bisnis tersebut ke dalam rumusan masalah yang dapat dijawab dengan data visualisasi dan *data mining*. Tahap ini juga mengharuskan kita untuk menemukan dan memperjelas apa hasil akhir yang sebenarnya diharapkan dalam proyek ini.

B. Data Preparation Phase

Pada fase ini dilakukan proses persiapan data yang akan digunakan dalam proses visualisasi data. Selain melakukan pengumpulan data, ada 3 tahap yang akan dilakukan dalam fase ini yaitu *choose the data set* (memilih data), *transform the data set* (transformasi data), dan *verify the data set* (verifikasi data).

1. Choose the Data Set

Setelah sebelumnya telah dilakukan identifikasi pertanyaan bisnis mana yang akan dijawab, maka diperlukan pencarian data operasional yang relevan dengan pertanyaan bisnis tersebut. Setelah mendapatkan data-data yang sesuai dengan pertanyaan tersebut maka dilakukan pemilihan data yang akan masuk ke dalam *exploratory data mart* melalui proses ECTL (*Extract, Cleanse,*

Transform, Load). *Exploratory data mart* menyimpan data dan informasi yang telah dikumpulkan dari berbagai sumber untuk menjawab pertanyaan bisnis.

2. *Transform the Data Set*

Set data yang telah dipilih biasanya masih membutuhkan beberapa perubahan atau transformasi sebelum dilakukan visualisasi agar representasi data menjadi lebih akurat dan sesuai dengan *value* yang dimiliki. Dalam metode ini

3. *Verify the Data Set*

Tahap terakhir setelah melakukan pemilihan dan transformasi data adalah melakukan verifikasi set data yang akan divisualisasikan. Proses verifikasi dilakukan guna mengkonfirmasi data yang digunakan adalah data yang *valid* dan bebas dari *error*. Apabila dalam prosesnya ditemukan adanya ketidakbenaran, ketidakurutan ataupun kesalahan pada logika data, maka proses ECTL (*Extract, Cleanse, Transform, Load*) dan *logical transformation* harus diperbaharui kembali.

C. *Data Analysis Phase*

Pada fase terakhir ini, akan dilakukan analisis pada set data yang dimiliki untuk menjawab pertanyaan bisnis yang ditanyakan di awal dengan cara melakukan visualisasi pada set data tersebut. Pada fase ini terdapat 3 tahap yang akan dilakukan yaitu *choosing the visualization or data mining tool* (pemilihan *tools* untuk visualisasi), *analyze the visualization or data mining model*

(menganalisa model visualisasi), dan *verify and present the visualization or mining model* (melakukan verifikasi dan menampilkan visualisasi).

1. *Choosing the Visualization or Data Mining Tool*

Pada tahap ini, akan dipilih *tool* yang akan digunakan untuk membuat visualisasi data yang akan menghasilkan *output* yang sesuai dengan pertanyaan bisnis.

2. *Analyze the Visualization or Data Mining Models*

Setelah menentukan *tool* yang akan digunakan untuk melakukan visualisasi data, maka langkah selanjutnya yang akan dilakukan adalah menganalisa output yang dihasilkan oleh visualisasi data untuk mendapatkan gambaran dan pemahaman dari set data untuk menjawab pertanyaan bisnis.

3. *Verifying and Presenting the Visualizations or Mining Models*

Tahap terakhir dalam metodologi *Visual Data Mining project* ini terdiri dari 3 bagian. Yang pertama adalah melakukan verifikasi bahwa hasil dari visualisasi data telah memenuhi tujuan bisnis. Langkah berikutnya adalah membuat presentasi untuk menjelaskan hasil visualisasi pada para pengambil keputusan. Terakhir, sesuai dengan tipe *VDM project* yang dibuat, ada kemungkinan untuk menyebarkan model visualisasi tersebut ke dalam lingkungan produksi.

Dalam mempresentasikan hasil visualisasi, ada 3 bagian utama yang harus diperhatikan yaitu:

- a. Penjelasan tujuan dari *VDM project*.
- b. Menampilkan poin penting dari hasil visualisasi untuk menjawab pertanyaan bisnis yang ada.
- c. Mengambil kesimpulan dan mendiskusikan langkah selanjutnya.

