



### **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 III**

### **METODOLOGI PENELITIAN**

#### **3.1. Objek Penelitian**

Penelitian dilakukan dengan menggunakan data terkait dengan kemiskinan antara lain Indeks Kemiskinan (Kedalaman & Keparahan), Garis Kemiskinan (Makanan & Non-Makanan), Persentase Penduduk Miskin, dan Jumlah Penduduk Miskin dan juga data pendidikan seperti tingkat pendidikan. Data didapatkan dari publikasi yang dikeluarkan oleh Badan Pusat Statistik Nasional terkait data dan informasi kemiskinan Kabupaten/Kota dari tahun 2013-2015.



***BADAN PUSAT STATISTIK***

**Gambar 3.1 Logo Badan Pusat Statistik**

### 3.2. Karakteristik Data

Data yang digunakan dalam penelitian ini terdiri dari:

#### 1. Data Primer

Data yang didapatkan dari objek penelitian ini yaitu Badan Pusat Statistik berupa publikasi data kemiskinan dan pendidikan berformat PDF. Data kemiskinan digunakan untuk merepresentasikan tingkat kemiskinan pada setiap provinsi di Indonesia, sedangkan data pendidikan digunakan untuk melihat per sebaran tingkatan pendidikan dari setiap kabupaten/kota di Indonesia. Data Kemiskinan yang digunakan antara lain adalah Jumlah Penduduk Miskin, Persentase Penduduk Miskin, Indeks Kedalaman Kemiskinan, Indeks Keparahan Kemiskinan, Garis Kemiskinan, sedangkan data pendidikan merupakan persentase tingkat pendidikan yaitu tidak/belum tamat SD, sudah tamat SD/SMP, sudah tamat SMA ke atas. Total data yang digunakan adalah 3200 baris data.

#### 2. Data Sekunder

Data dan informasi pendukung yang didapatkan dari hasil studi literatur yang mendukung penelitian dan pembuatan visualisasi data antara lain teori mengenai metode *eight-step data visualization and data mining methodology (VDM Project)*, *database, data mining*, dan lain-lain.

### 3.3. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terbagi menjadi 2 kategori yaitu variabel dependen dan independen.

### **3.3.1. Variabel Independen**

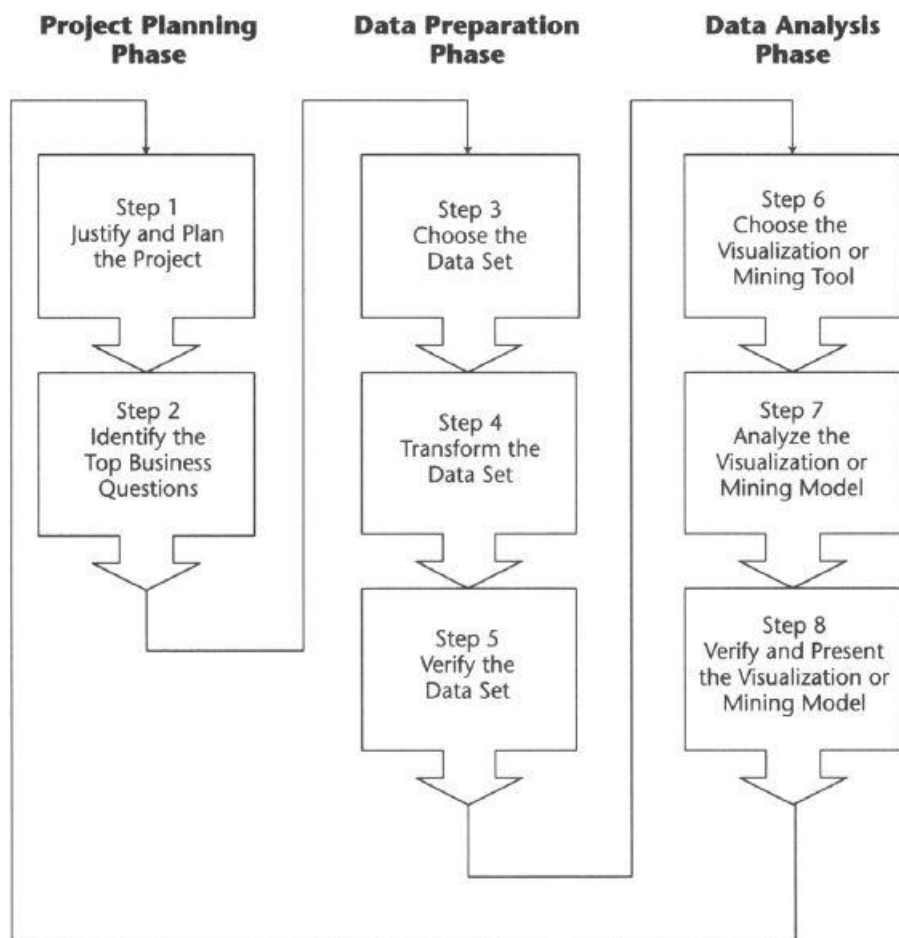
Variabel independen atau variabel bebas merupakan variabel yang mempengaruhi atau merupakan penyebab timbulnya variabel dependen (terikat). Variabel independen pada penelitian ini adalah tingkat pendidikan sebagai salah satu faktor penyebab kemiskinan pada suatu daerah.

### **3.3.2. Variabel Dependen**

Variabel dependen atau variabel terikat merupakan variabel yang dipengaruhi oleh variabel independen atau merupakan akibat dari adanya variabel bebas. Variabel dependen pada penelitian ini adalah tingkat kemiskinan yang ada pada suatu daerah.

## **3.4. Metodologi Penelitian**

Metode yang digunakan dalam penelitian ini adalah *Eight-step data visualization and data mining methodology*, yang merupakan sebuah metode yang digunakan untuk melakukan proses *data mining* dan data visualisasi dalam sebuah proyek yang disebut *Visual Data Mining Project (VDM Project)*. Metode ini dipilih karena sesuai dengan proses pembuatan visualisasi data yang meliputi proses *data mining* dan proses *extract, transform, load (ETL)*.



**Gambar 3.2** *Eight-Step Data Visualization and Data Mining Methodology*

### 3.5. Perbandingan Metodologi

Sebelum memilih metode *eight-step data visualization and data mining methodology* atau *VDM Project*, dilakukan perbandingan metodologi yang digunakan dengan metodologi lainnya yaitu metodologi *User-Centered Design (UCD)* dan metodologi *Neotix*.

#### 1. Metodologi *User-Centered Design (UCD)*

Metodologi ini digunakan untuk perancangan visualisasi data berbasis *user*, dimana dilakukan pengembangan sesuai dengan kebutuhan

organisasi yang dijabarkan melalui wawancara dengan *user*. Tahapan metodologi ini antara lain:

**a. *Plan the Human Centered Process***

Merencanakan proses pengerjaan berbasis *user* (persiapan wawancara).

**b. *Specify the context of use***

Menentukan konteks penggunaan sistem yang akan dibuat.

**c. *Specify user and organizational requirement***

Menentukan *requirements* dari *user* dan organisasi.

**d. *Product design solutions***

Memberikan solusi dengan desain produk (sistem/*dashboard*).

**e. *Evaluate design against user requirements***

Mengevaluasi desain yang dibuat sesuai dengan kebutuhan *user*.

**2. Metodologi Neotix**

Metodologi ini berfokus pada tahapan *system development life cycle* (SDLC) dan lebih berfokus pada kemampuan *developer*. Tahapan yang dilakukan dalam metodologi ini adalah:

**a. Perencanaan**

Melakukan perencanaan terkait apa saja yang akan dikembangkan dalam sistem.

**b. Identifikasi dan perancangan *prototype***

Melakukan identifikasi kebutuhan dan merancang sistem percobaan awal (*prototype*).

**c. Desain**

Melakukan perancangan sistem sesuai identifikasi kebutuhan.

**d. Implementasi dan Validasi**

Melakukan implementasi ke dalam organisasi, dan melakukan validasi terhadap *error*

**e. Deployment**

Melakukan peluncuran sistem untuk digunakan di dalam organisasi.

Alasan pemilihan metode *Eight-step data visualization and data mining methodology (VDM Project)* dibandingkan metode lainnya adalah karena metode ini membahas secara rinci proses *data mining* yang dilakukan termasuk proses *extract, cleanse, transform, load (ECTL)* yang merupakan bagian penting untuk pembuatan visualisasi data, karena merupakan proses pembentukan *data set* yang digunakan dalam visualisasi. Dua metode pembandingan yang dijelaskan tidak memiliki fokus yang serupa terhadap data, melainkan berfokus pada *user* (Metode *User-Centered Design*) dan pada sistem dan *developer* (SDLC pada metode *Neotix*).

**3.6. 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).

### **3.6.1. *Justify and Plan the Project***

Pada tahap ini dilakukan pembentukan tujuan yang ingin dicapai dalam sebuah proyek data visualisasi serta perencanaan proyek tersebut. Pada penelitian ini akan dipilih salah satu dari 3 tipe *scope* data visualisasi yang ada dalam metode ini yaitu *Proof-of-concept*, *Pilot*, dan *Production*.

Setelah ditentukan tipe data visualisasinya, akan dilakukan pembuatan *project justification* untuk menentukan target, dan juga tujuan dari data visualisasi dalam sebuah proyek. Setelah itu harus dilakukan penentuan hal-hal seperti *timeline* dan sumber daya yang digunakan dalam proyek.

### **3.6.2. *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.

## **3.7. *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).



### **3.7.1. 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. Data akan dicari melalui *website* [www.bps.go.id](http://www.bps.go.id) dan apabila diperlukan informasi tambahan dapat dicari melalui sumber lain seperti [www.data.go.id](http://www.data.go.id) atau pun melakukan pengecekan lebih lanjut ke kantor BPS. Setelah mendapatkan data-data yang sesuai dengan pertanyaan tersebut maka dilakukan pemilihan data yang akan dibuat data setnya melalui proses ECTL (*Extract, Cleanse, Transform, Load*) yang kemudian akan di-*import* ke dalam *tool* visualisasi yang akan digunakan.

### **3.7.2. 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 penelitian ini dilakukan perubahan-perubahan yang perlu dilakukan agar data set dapat di-*import* ke dalam *database* visualisasi dan digunakan untuk membuat visualisasi.

### **3.7.3. 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*. Yang akan dilakukan di penelitian ini adalah

memastikan data yang tercatat dalam publikasi valid dan juga data set setelah proses transformasi data tidak mengalami *error type* atau ketidakcocokan format dalam *tool* visualisasi data.

### **3.8. 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).

#### **3.8.1. Choosing the Visualization or Data Mining Tool**

Pada tahap ini, dipilih *tool* yang akan digunakan untuk membuat visualisasi data yang akan menghasilkan *output* yang sesuai dengan pertanyaan bisnis. Pemilihan *tool* dalam penelitian ini akan membandingkan 2 jenis *tool* visualisasi yaitu Tableau dan PowerBI.

#### **3.8.2. 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.

### 3.8.3. *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 dengan cara melakukan presentasi hasil visualisasi kepada *user* yang bersangkutan dan membuat *user acceptance test* dari hasil *feedback user* tersebut. Kemudian setelah selesai diverifikasi, hasil visualisasi dapat di-*upload* ke *repository* di *website* Tableau.

UMN

### 3.9. Teknik Pengumpulan Data

Pada penelitian ini, pencarian data didapatkan dengan cara mengunduh data publikasi yang tersedia untuk umum (*open data*) dari Badan Pusat Statistik melalui website [www.bps.go.id](http://www.bps.go.id). Pencarian informasi lebih lanjut mengenai data dilakukan dengan mendatangi kantor BPS di Jakarta. Data yang digunakan dalam penelitian adalah tingkat kemiskinan dan tingkat pendidikan yang berada pada provinsi di Indonesia.

### 3.10. Teknik Simulasi Data

Simulasi data dilakukan dengan cara membangun *dashboard* visualisasi data menggunakan Tableau sebagai *tool* pembuatan visualisasi data. Proses visualisasi menggunakan metode *eight-step data visualization and data mining methodology* yang di dalam prosesnya membutuhkan proses pengolahan data berbasis ETL (*Extract, Transform, Load*). Data yang berasal dari hasil publikasi PDF diolah menjadi *spreadsheet* Microsoft Excel (XLS) yang kemudian di-*import* ke dalam *database* Tableau untuk dilakukan pembuatan visualisasi data berupa *mapping*, *bar chart*, *pie chart*, dan sebagainya.

UMN

### 3.11. Penelitian Terdahulu

Langkah awal yang dilakukan saat melakukan penelitian ini adalah melakukan studi literatur, dimana pada tahap ini dilakukan pencarian jurnal, buku, artikel sebagai sumber-sumber referensi daripada teori ataupun metode yang dijadikan sebagai acuan dalam melakukan penelitian ini.

Referensi yang dicari pada penelitian ini berkaitan dengan definisi dan juga metode terkait visualisasi data, *database*, *data mining*, *data warehouse*, *data cleaning*, serta penjelasan umum terkait data-data kemiskinan.

Beberapa jurnal yang digunakan sebagai referensi penelitian ini adalah:

1. Jurnal “Visualisasi Data Interaktif Data Terbuka Pemerintah Provinsi DKI Jakarta: Topik Ekonomi dan Keuangan Daerah”. (Syaripul & Bachtiar, 2016). Jurnal ini membahas tentang pembuatan visualisasi data berbasis *mobile* dengan menggunakan *open data* dari Pemerintah Provinsi DKI Jakarta.
2. Jurnal “Teknik Visualisasi Dalam *Data Mining*” (Mulyana & Winarko, 2009). Jurnal ini membahas teknik-teknik visualisasi data dan juga keuntungan dalam menggabungkan beberapa jenis visualisasi yaitu menghasilkan strategi *data mining* yang kuat.